IT'S BOTH WHO YOU ARE AND WHERE YOU'RE FROM:
RELATING VOCATIONAL INTERESTS AND SOCIOECONOMIC STATUS TO BIAS IN
BIODATA AND SJTS

By

Joshua Prasad

A THESIS

Submitted to
Michigan State University
In partial fulfillment of the requirements
for the degree of

Psychology-Master of Arts

2017

ABSTRACT

IT'S BOTH WHO YOU ARE AND WHERE YOU'RE FROM:
RELATING VOCATIONAL INTERESTS AND SOCIOECONOMIC STATUS TO BIAS IN
BIODATA AND SJTS

By

Joshua Prasad

Differences in responding to biodata and situational judgement tests (SJTs) based on gender and racial minority group status were evaluated. It was hypothesized that vocational interests and socioeconomic status (SES) could be used to help characterize the differences in experience between groups (e.g. Cottrell, Newman, Roisman, 2015; Nye, Su, Rounds, & Drasgow, 2012). As a result, interests and SES may help explain differences in both the constructs assessed by biodata and SJTs as well as differences in item functioning (DIF; Drasgow, 1987). Hypotheses were evaluated using multiple-indicator multiple-cause models to simultaneously model latent constructs and item responses (MIMIC; Muthén, 1989). Findings indicate that interests helped explain differences across gender in both the constructs assessed as well as DIF. Interests explained few differences based on minority group status and SES did not seem to meaningfully explain differences in either of the demographic group comparisons. Many items still exhibited DIF as a function of gender or minority group status after accounting for vocational interests and SES, suggesting that further work is needed to identify additional substantive explanations of DIF. Overall, the present work constitutes a thorough examination of differential functioning in noncognitive assessments and establishes a meaningful relationship between the noncognitive constructs assessed here and vocational interests.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

As Industrial/Organizational (IO) psychology grapples with issues related to adverse impact and cognitive testing, organizations have been increasingly reliant on biodata and situational judgment tests (SJTs), due to the fact that they show less potential for adverse impact while maintaining reasonable validities for predicting important outcomes like training or job performance (Robertson & Smith, 2008; Ployhart, 2006; Schmitt, Keeney, Oswald, Pleskac, Billington, Sinha, & Zorzie, 2009). Though the potential for adverse impact may be lessened, recent work has demonstrated that there may still be room for concern with biodata (Imus, Schmitt, Kim, Oswald, Merritt, & Westring, 2010) and SJTs due to measurement bias (Kim, Schmitt, Friede, Oswald, Ramsay, et al., 2004), which can contribute to adverse impact (Nye & Drasgow, 2011). Measurement bias occurs when individuals with the same standing on the latent trait assessed by the test, but sampled from different subgroups, have unequal observed scores on the scale (Drasgow, 1987). In other words, bias represents differences in the way that individuals respond to a measure rather than actual differences in the latent trait. In addition to its potential effects on adverse impact, measurement bias can also influence comparisons across groups. Therefore, bias in the measure is an important concern that needs to be addressed.

The explanations for the bias on biodata measures and SJTs provided by both Imus et al. (2010) and Kim et al. (2004) relied solely on group differences in access to experiences relevant to these measures. Although these studies identified group differences due to measurement bias, simply identifying these differences does little to explain the psychological mechanisms underlying them. There are likely to be several factors that differ across groups and that may cause bias in the measurement of psychological constructs and understanding these factors will provide additional information about how issues of measurement bias in biodata and SJTs can be

addressed in future research. Therefore, the present work attempts to broaden the investigation of how and why bias might occur in biodata and SJTs to address this gap in the literature.

Biodata and SJTs can be designed to measure social and motivational attributes (e.g. perseverance, adaptability, leadership) that may be relevant to a particular position, but are not typically captured by assessments focused on cognitive abilities. In other words, biodata and SJTs provide the methodology for assessing a broad range of attributes in a systematic manner. Biodata assessments involve asking respondents the frequency with which they engage in behaviors that are thought to be job relevant, with the thought that respondents may likely continue these behaviors while on the job (Hough & Oswald, 2000; Whitney & Schmitt, 1997). SJTs, on the other hand, consist of presenting the respondent with job-relevant dilemmas and asking him or her to evaluate the appropriateness of a set of potential responses (McDaniel, Morgeson, Finnegan, Campion, and Braverman, 2001).

The utility of these assessments has been established by demonstrating the unique validity these assessments hold alongside other predictors, such as cognitive assessments, when predicting criteria like job performance (McDaniel et al., 2001), training (Robertson & Smith, 2008), or early college success (Schmitt et al., 2009). Given the flexibility of these assessments, their potential relationships with key workplace outcomes, and their low to moderate subgroup differences, biodata and SJTs present a powerful pair of assessment methods. Further, the ability to create items that are clearly connected to the workplace tend to generate favorable reactions from both job applicants and HR professionals (Ployhart, 2006). However, these benefits can only be realized if the latent constructs assessed by these methods are measured appropriately. Additionally, due to the differences in how these measures are constructed and the way they measure latent constructs, examining bias in each of these techniques will provide a more general understanding of bias, rather than the current norm of merely demonstrating that bias

occurs without explaining why it occurs (Gierl, 2005). As a result, biodata and SJTs are the focus of this investigation, and will be introduced in turn.

Biodata. The key principle of biodata assessments is that past behavior is predictive of future behavior. By assessing the frequency or quality of behaviors that may be job relevant, the continuation of these behaviors on the job may prove beneficial to performance (Hough & Oswald, 2000; Whitney & Schmitt, 1997). In addition to their job relevance, the behaviors targeted in a biodata assessment are selected based on whether they are thought to causally influence the construct of interest (Dean, 2013). A strength of biodata assessments is that they can be tailored to a specific job in order to increase their validity with important outcomes (Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004). This strength comes with a drawback, however, because the term biodata is often used in an unsystematic way. Mael (1991) attempted to address this criticism by establishing a framework to clarify what constitutes a biodata assessment. This framework describes biodata assessments broadly as tapping into behaviors that an individual has enacted in order to adapt to his or her environment, as well as behaviors that are consistent with his or her personal and social identity.

Despite the variety of biodata assessments, reasonable validities have been observed when using biodata as a predictor across a number of jobs (Hunter & Hunter, 1984) and criteria, such as performance, salary progress, and person-organization fit (Mael & Ashforth, 1995; Schneider & Schmitt, 1986). These benefits also come with frequent observations of low adverse impact (Bliesener, 1996; Shackleton & Newell, 1997). It is common practice to screen biodata items for differential functioning across groups. However, little work has been produced that can inform actual evidence based principles in item screening (Whitney & Schmitt, 1997). In applied settings, biodata assessments are most commonly used as a selection tool (Dean, 2013; Whitney & Schmitt, 1997). They have been used successfully for a wide range of jobs such as managers,

hotel staff, and equipment distributors, among many others, though they are still less common than conventional selection methods such as interviews (Robertson & Smith, 2008).

Biodata assessments are typically scored using either rational scoring or empirical keying. Rational scoring involves subject matter experts weighting each item based on how job relevant they view the item to be (Hough & Paullin, 1994). Empirical keying, on the other hand, refers to collecting data on strong and weak employees and using their responses to weight items based on how well the item discriminates between those employees (Mumford & Owens, 1987).

Situational Judgment Tests. Rather than assess attributes by having individuals endorse statements about themselves and their experiences, SJTs present a series of job-relevant dilemmas, as well as several potential responses, and participants must evaluate the appropriateness of the responses (McDaniel et al., 2001). Contemporary versions of these tests are developed using subject matter experts, who generate the dilemmas, provide realistic responses, and identify the best and worst options (McDaniel et al., 2001). Participants are given scores based on whether or not their answers align with what subject matter experts (SMEs) deemed to be the best and worst responses to each dilemma (Schmitt et al., 2009). As is the case with biodata, SJTs are typically used in selection contexts and their popularity has been increasing, particularly in the United States and Europe (McDaniel, Hartman, Whetzel, & Grubb, 2007). SJTs can also be helpful as a training and development tool, given the ability to practice analyzing situations and choosing a response (Ployhart, 2006).

SJTs were originally criticized for being a reflection of intelligence rather than a unique ability, which appeared to be the case given initial correlations with tests of mental ability (McDaniel et al., 2001). Further refinement of these tests, however, began to differentiate them from cognitive ability measures. A key advance in the development of SJTs involved asking respondents to rate both the best and worst responses to a situation. This modification resulted in

4

moderate validities with job performance criteria and weak relationships with mental ability (Motowidlo, Dunnette, & Carter, 1990). However, McDaniel et al. (2001) caution that this was often found to be the case in samples where range restriction was a possibility.

An additional point of investigation regarding the construction of SJTs has been related to the instructions provided for these measures. The two broad categories of instructions are referred to as behavioral tendency and knowledge instruction. Behavioral tendency instructions ask the participant to rate which actions they would be most and least likely to take, whereas knowledge instructions have participants report which actions should be taken (McDaniel et al., 2007). Though a seemingly subtle difference, meta-analytic work has suggested that SJT instruction types are related to important outcomes of the assessment process. Whetzel, McDaniel, and Nguyen (2008) showed that SJTs administered with knowledge instructions produced the greatest racial subgroup differences, which was explained by the fact that SJTs with this type of instruction correlated higher with measures of cognitive ability. McDaniel et al. (2007) point out that the weaker relationships between the behavioral tendency instructions and cognitive ability may be due to the production of scores that reflect typical performance, whereas cognitive ability describes maximal performance. Behavioral tendency instructions come with their own drawbacks, however, as responses may be more prone to distortion. Respondents engaging in impression management or faking may select responses that are socially desirable rather than those that reflect typical performance (Ployhart, 2006).

Measurement Bias

Measurement bias is problematic for selection settings because it can result in the disproportionate selection of one group over another. This can increase errors in the selection process and result in either selecting unqualified individuals or rejecting those who are qualified (Aguinis & Smith, 2007). Further, use of a biased test can contribute to adverse impact, or the

selection of minority group members at a significantly lower rate than those of a majority group (Zedeck, 2010). An important contributor to bias at the test level is bias at the item level (i.e. differential item functioning or DIF; Nye & Drasgow, 2011).

DIF is generally characterized as either uniform or nonuniform. Uniform DIF refers to differences across groups that are consistent throughout the entire range of the latent trait. Nonuniform DIF, on the other hand, refers to differences that are not consistent across range of the latent trait, such that group differences vary in size depending on standing on the latent trait (Mellenbergh, 1989). Identifying either form of DIF and its contribution to scale scores may help to decompose why score differences may exist between groups. It should be noted, though, that removal of items flagged for DIF may only result in a modest reduction in group differences (e.g. reduction in standardized group difference of roughly .05; Schmitt & Quinn, 2010).

Figure 1 illustrates how items relate to a latent factor in a measurement model and serves as the starting point for further testing of DIF. In the figure, a latent factor represents the trait assessed by the measure (i.e., either biodata or SJT). Each item within the measure has a linear relationship with the latent factor that is defined by an intercept and a factor loading. The intercept describes the predicted value of item responses when the latent trait is at a value of zero. In the analyses conducted here, zero represents an average standing on the latent trait, so an item's intercept describes how an individual with an average standing on the latent trait would respond to that item. A uniform DIF effect can be thought of as differences in an item intercept across groups because this difference would be constant across the entire range of the latent trait (Woods & Grimm, 2011).

The factor loading of an item describes how a difference in standing on the latent trait corresponds to a difference in the predicted item response. For example, if an individual has a latent score of one (latent standing is one standard deviation above the mean), the factor loading

Figure 1.

Example measurement model of a latent factor with scale items serving as observed indicators
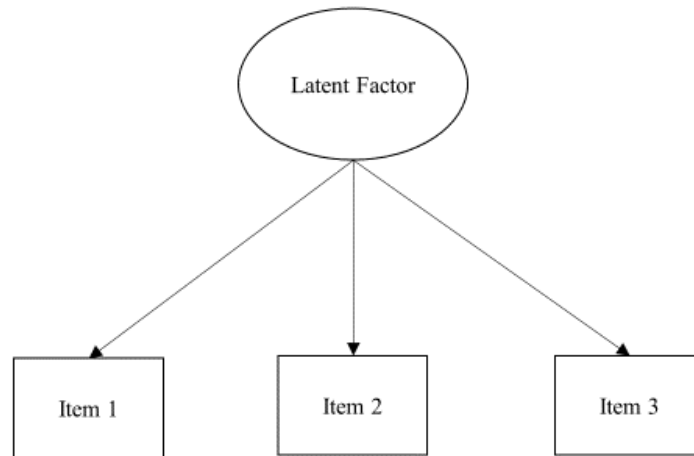
Note. Items 1 through 3 represent the items of a particular noncognitive scale, measuring a Latent Factor. Paths originating from the latent factor leading to each item represents that item's factor loading. Item intercepts are not depicted graphically but are estimated in the measurement models analyzed in this study.

would describe how much higher that individual's predicted item response would be when compared to an individual with an average standing on the latent trait. A nonuniform DIF effect can be thought of differences in the factor loading of an item across groups. A significantly different factor loading across groups would result in predicted item response differences across groups as well. In an ideal situation when no DIF exists, neither the intercept nor the factor loading of an item would differ as a function of group membership. In contrast, DIF occurs when the factor loadings and/or intercepts are influenced by non-random factors that differ across groups.

Potential for Bias in Biodata and SJTs

As mentioned above, perhaps the chief appeal of using biodata and SJTs to measure constructs of interest is their purported ability to avoid adverse impact while being of use in

predicting performance (Robertson & Smith, 2008; Ployhart, 2006; Schmitt et al., 2009). Tests that avoid adverse impact are important because of their ability to reduce legal liability and result in a more ethical employee selection process (Aguinis & Smith, 2007). The claim that biodata and SJTs demonstrate little risk for adverse impact is not without contest. Bobko and Roth (2013) reviewed meta-analytic evidence of black-white subgroup differences and found differences of $d = .38$ for SJTs (Whetzel, McDaniel, & Nguyen, 2008) and $d = .33$ for biodata (DeCorte, Lievens, & Sackett, 2007), favoring whites in both cases. Characterizing these effect sizes as small, or as minimal risk, is somewhat misleading as Bobko and Roth (2013) point out, but also potentially underestimates the effects in actual selection contexts. This is due to the fact that a majority of the studies contributing to these meta-analytic effect size estimates are from incumbent, rather than applicant samples. Weekeley, Ployhart, and Harold (2004) provide one of the few studies where applicant samples were available, and conversion of their results into effect sizes produced differences in SJT scores of $d = .39$ for incumbents and $d = .79$ for applicants when comparing whites to non-whites (Bobko & Roth, 2013). Whetzel, McDaniel, and Nguyen (2008) found similar results for Hispanic ($d = .24$) and Asian ($d = .29$) respondents, with the extent to which the test was associated with cognitive ability being the strongest explanatory variable. Here, cognitive ability appears to be a moderator of these subgroup differences (Whetzel et al., 2008). In addition, black-white differences in biodata scores also appear to be explained in part by cognitive ability (DeCorte, Lievens, & Sackett, 2007). It should be noted that with regard to gender, overall SJT differences favored females ($d = -.11$), with conscientiousness and agreeableness, rather than cognitive ability, serving as meaningful meta-analytic moderators (Whetzel et al., 2008).

Additional work examining subgroup differences in biodata and SJTs has explored differences across cultural groups. For example, recent work by Prasad, Schmitt, Ryan, Showler,

Nye (2016) examined differences in the operation of biodata and SJTs when comparing American and Chinese student samples. Though not wholly supported, hypotheses were based on differences in experience based on differing educational systems, as well as differences in cultural values as measured by the GLOBE study (House & Hanges, 2004). Prasad et al. (2016) found that score differences on biodata scales assessing leadership, knowledge, adaptability, and perseverance aligned with differences in the educational systems and cultural values of the two groups. On the other hand, scales assessing continuous learning, social responsibility, and academic values (i.e. behaving in accordance with a well-developed set of values), did not align with hypothesized differences, revealing the possibility of other influences on group mean differences in biodata assessments. Additionally, Prasad et al. (2016) conducted MACS analyses, as outlined by Nye and Drasgow (2011), to try to separate latent trait differences from bias. The role of bias in influencing observed mean differences became quite apparent as differences between observed and corrected effect sizes ranged from $d = .05$ to $d = .51$ across biodata scales and the SJT. Further, any observed differences between groups on the SJT were eliminated after accounting for bias, suggesting that raw score differences, at least in the comparison of American and Chinese students, reflect bias in measurement more so than meaningful differences in situational judgment. At the scale level, attempts to understand why group differences exist have been limited to the prediction of group differences or the use of moderators in meta-analysis. Other work examines item-level responses in efforts to better understand how different groups use biodata and SJTs.

Multiple studies have shown that bias at the item-level is an important concern with both biodata and SJTs (Imus et al., 2010; Kim et al., 2004; Whitney & Schmitt, 1997). Whitney and Schmitt (1997) explored the potential influence of differences in cultural values on item use across black and white respondents. Though they found that culture was a general predictor of

9

response selection, they did not find that culture explained DIF. Imus et al. (2010) argued that the cause of DIF was due to members of certain demographic groups having reduced access to the experiences probed in some of the biodata items. Upon examination, Imus et al. (2010) found that the degree to which an item operated differently across sexes was negatively correlated with how much more accessible the item was judged by females (when compared to male judgments of accessibility; $r = -.51$, $p < .05$). The researchers described accessibility as the degree to which an individual felt they had ample opportunity to experience the event or situation described in the item stem. Though DIF was also observed across races, accessibility was not shown to explain these differences.

Based on the work at both the item- and scale- levels, it is clear that bias can play a role in certain biodata assessments and SJTs. The explanation of bias, however, must be further developed. Substantive explanations of bias are important for understanding why bias occurs, but have progressed slower than our ability to statistically model bias (Gierl, 2005). In other words, a substantial amount of work has been conducted to identify bias but much less work has examined potential explanations for the bias that is identified. As Whitney and Schmitt (1997) point out, test developers have little evidence based guidance about how to write items in a way that avoids bias. Consequently, DIF can only be detected post-hoc and may require researchers to drop items after the data have already been collected (Imus et al., 2010). Dean (2013) demonstrated that the identification and removal of biased items substantially improved the measurement qualities of a biodata measure. Specifically, a substantial reduction of subgroup differences was observed with little influence on the predictive validity of the assessment. Further understanding of why items function differently may help to engender the improvements observed by Dean (2013) in other biodata measures and SJT prior to conducting a costly test validation study. Therefore, the goal of the present study was to propose a conceptual model

(depicted in Figure 2) that will help to build upon past developments to understand the sources of

bias in biodata measures and SJTs. Two potential mechanisms behind these sources of bias are

explained next.

Figure 2.
Proposed analytic approach of testing how interests and SES influence individual responses to biodata and SJT items
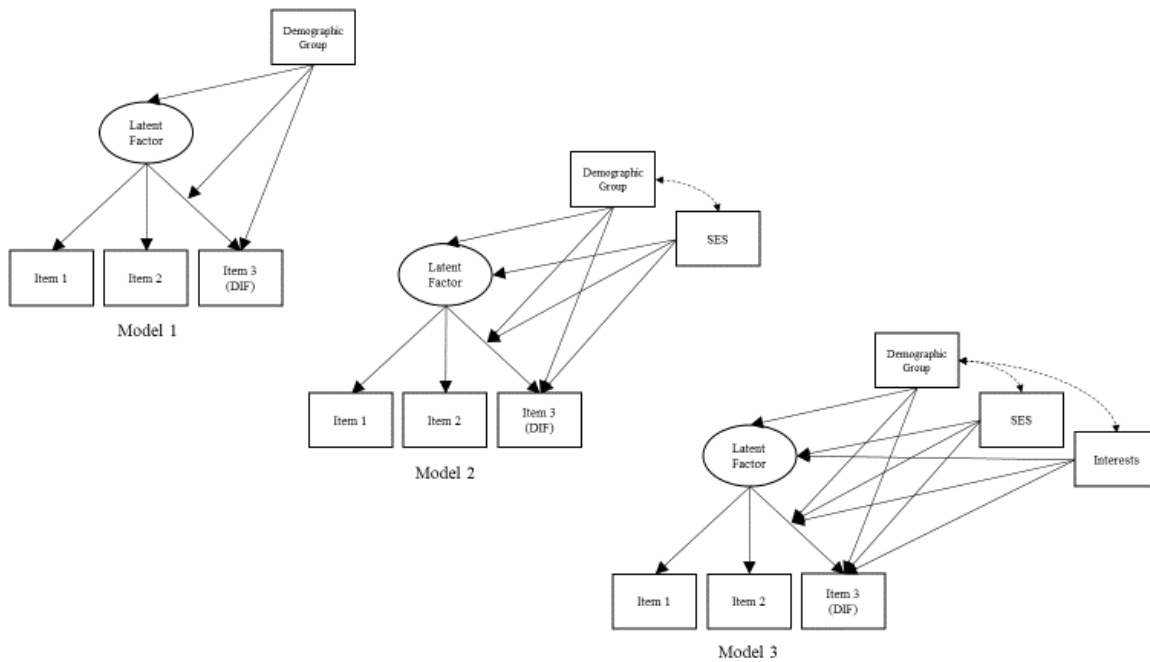
Note. Models 1 through 3 denote sequential MIMIC models incorporating additional explanatory variables (SES and Interests). Items 1 through 3 represent the items of a particular noncognitive scale, measuring a Latent Factor. Item 3 represents an item that has been flagged for DIF. Paths originating from the latent factor leading to each item represents that item's factor loading. Paths originating from Demographic Group, SES, and Interests leading to the Latent Factor represent the regression of the Latent Factor onto each variable. Paths originating from Demographic Group, SES, and Interests leading to the factor loading of Item 3 and Item 3 itself represent tests of nonuniform and uniform DIF, respectively. Curved, dotted lines between Demographic Group and both SES and Interests represent the observed biserial correlations between those variables.

Frame of Reference. When responding to biodata or SJT items, participants must often

make an evaluation in reference to some other person or group. This can occur explicitly in

either the question stem (e.g. "How often do others tend to compliment you on your

determination to continue with a project under difficult circumstances?") or the response options

(e.g. "Much more than most people"). Even if not done explicitly, items may ask respondents to evaluate an abstract amount, whereby making some sort of social comparison may help participants respond to the item. This can be problematic, as argued by Robert, Lee, and Chan (2006), since an assumption that many test creators make is that participants are sampled from the same population and make evaluations against that population. This may not be the case as individuals of different backgrounds may evaluate against a reference group closer to themselves and not against the population intended by the researcher. Robert, Lee, and Chan (2006) refer to this as the frame of reference effect, whereby the participant evaluates against a local comparison group rather than a global one (i.e. population of interest).

The frame of reference effect is thought to exert its influence by producing non-equivalent intercepts on items when comparing different groups. Robert, Lee, and Chan (2006) describe this as the product of individuals responding to an assessment based on the perceived differences between themselves and their comparison group. This problem has been represented graphically using Figure 3. Two individuals ("A" and "B") are shown to have equivalent standings on some latent trait. However, both individuals are making comparisons against local comparison groups with different standings on that same trait. When responding to an item related to this latent trait, individual A's response may be inflated due to the relatively large perceived difference between individual A and his/her local comparison group. Other participants with a similar background to individual A (i.e. use a similar local comparison group) will respond systematically higher to items reflecting this latent trait than would participants similar to individual B. As a result, responses would not be comparable between the groups that individuals A and B come from.

Figure 3.
Depiction of individuals with equivalent standings on a latent trait but nonequivalent comparison groups
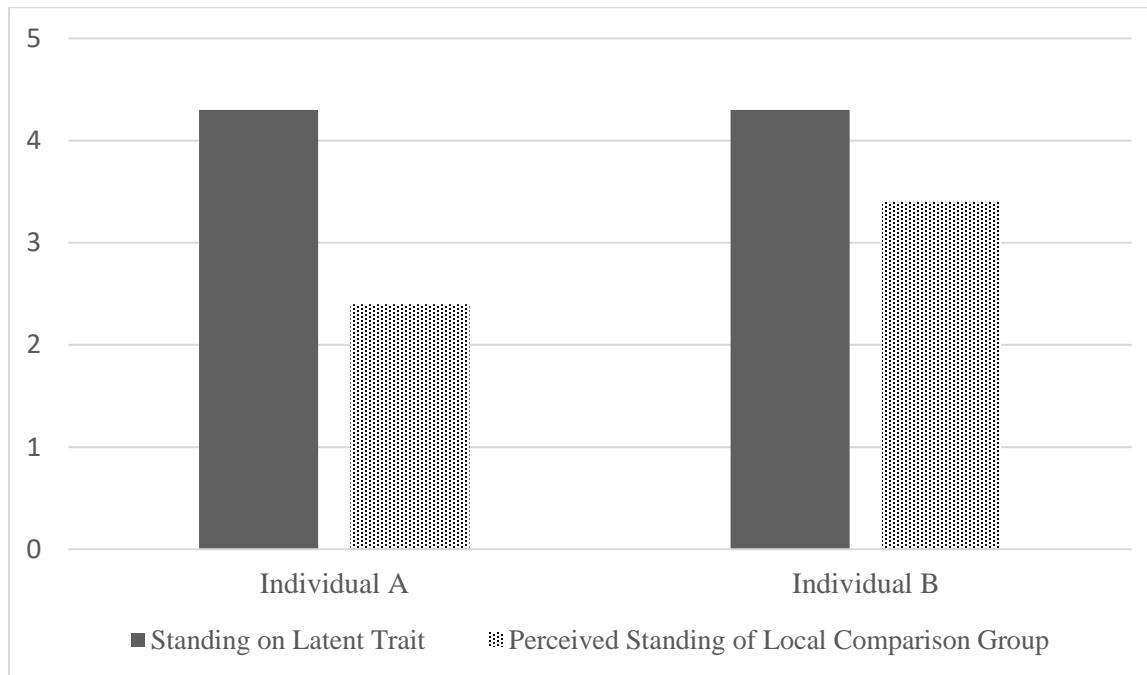
Item Accessibility. In addition to the frame of reference effect, Robert, Lee, and Chan (2006) suggest that the relevance of an item to the construct being assessed by a scale may differ as a function of group membership, which may bias responding. In other words, the degree to which item content accurately reflects a construct may vary between groups. When responding to biodata and SJT items, individuals are often presented with a specific behavior or situation meant to serve as an example of a broader construct. Problems may arise when the item content is more construct-relevant for one group than another. For example, consider the following biodata item stem: "In the past six months, how often did you read a book just to learn something?" This item uses the specific behavior of voluntarily reading a book as an indicator of the broader construct of continuous learning. Should an evaluator use this item to compare continuous learning between younger and older adults, the underlying assumption would be that books are an equally applicable means of seeking new information for both groups of respondents. If this is not the

case, the item will have a weaker relationship with the construct of interest for the group who are less likely to read books irrespective of continuous learning. Put more generally, if item content differs in relevance between groups, the item will have a weaker loading on the latent construct for the group who finds the content less relevant.

Item accessibility, as investigated by Imus et al. (2010), can be thought of as a specific case of why items may vary in construct-relevance between groups. Imus et al. (2010) define item accessibility as differences between groups in the opportunity to have specific experiences due to social barriers resulting in those experiences being differentially construct relevant. For example, Imus et al. (2010) found that Black respondents felt as though they had less opportunity to take, "a leadership role in High School and/or organized activity," than White respondents. Further, Imus et al. (2010) go on to describe that items that are differentially accessible would also differ in how informative they are as indicators of the underlying construct.

The proposed study seeks to apply the frame of reference effect and item accessibility to potential DIF in biodata and SJT items. Based on previous research, the current work proposes that SES and vocational interests will influence an individual's frame of reference when responding to biodata and SJT. Further, the accessibility of the item content in biodata and SJT items may vary across groups due to SES (i.e. restricted access). The conceptual model that is proposed here is illustrated in Figure 2 and the rationale behind these proposed mechanisms is described below.

Socioeconomic Status

The socioeconomic status (SES) of the community an individual comes from may strongly influence the experiences her or she has had and the factors that come to mind when evaluating items related to academic pursuits. The broad argument presented here is that the content assessed by biodata and situational judgment items may be influenced by SES (e.g. Kim

et al., 2004, Imus et al. 2010). As reviewed by Cottrell, Newman, and Roisman (2015), low SES communities can suffer a number of setbacks. These researchers also connect SES differences to race via census data. Updated 2014 estimates of household income indicate that Black families across the United States have a median income of $35,398 and Hispanic families have a median income of $42,491. Both figures are substantially lower than the median incomes of White ($60,256) and Asian ($74,297) families (United States Census Bureau, 2015). Cottrell et al. (2015) argue that these race differences in SES can provide a partial explanation for subgroup differences on some psychological characteristics. As such, differences in SES are also likely to be a potential source of both DIF and true score differences in biodata and SJTs.

It is important to note that observed score differences on a latent construct are a function of both bias and true score differences across groups. These true score differences are commonly referred to as impact, and despite being a valid representation of a particular difference between groups, can still contribute to adverse impact in selection contexts (Cottrell, Newman, & Roisman, 2015). Using an appropriate methodology, it is possible to differentiate bias and impact in group differences. As described below, SES is likely to be one source of bias on biodata and SJT items. However, SES may also produce score differences on these types of measures by influencing the latent constructs that are assessed. Because biodata and SJT items assess past experiences and the procedural knowledge developed as a result (Lievens & Motowidlo, 2016; Mael, 1991), individuals from low SES communities may have had fewer relevant experiences resulting in lower levels of the latent trait being assessed after accounting for bias in the measure. Should this be the case, it is likely that economically disadvantaged minorities have been denied access to the opportunities necessary to develop the latent abilities measured by biodata and SJTs.

H1: The effects of minority status on the standings of the latent traits measured by biodata and SJTs will be partially explained by SES such that minority status will initially predict lower standings on the latent traits measured by biodata and SJTs and this effect will be weakened upon inclusion of measures of SES.

Though the effects of SES are likely broad, SES may specifically influence the characterization of a local comparison group. SES may be particularly characteristic of the comparison group for those in the current study given that they are just transitioning out of high school and that the conditions of their school are likely linked to the SES of the surrounding area. Poor communities may have reduced access to educational resources, providing fewer opportunities for students to engage in academic pursuits than in more affluent areas (Duncan & Magnuson, 2005). Additionally, parents in these communities may have less leisure time to spend with their children, resulting in fewer occasions to convey educational aspirations (Cottrell et al., 2015). Consequently, when individuals from low SES communities consider how often they engage in the academic activities assessed in the biodata scales of knowledge, continuous learning, and perseverance, the norm for their local comparison group may be relatively lower due to the reduced academic resources of those around them. This norm may lead them to use a different frame of reference than other individuals from more affluent areas. In turn, this may lead individuals from low SES communities to overestimate their actual standing on these items.

Further, Cottrell et al. (2015) also point out that poorer communities are also relatively more dangerous, serving as a less stable environment for individuals residing in these areas. Again, when evaluating questions related to academic values and social responsibility, individuals from low SES communities may have a different frame of reference. This can result in DIF across groups in the form of different item intercepts, due to the potentially lower standing of their local comparison group on these dimensions (e.g. Robert, Lee, & Chan, 2006),

16

such that individuals from low SES communities are more likely to endorse higher response

options than individuals from more affluent communities. However, it should be highlighted that

though SES will likely lead to intercept differences as a function of race, factor loading

differences will be unlikely among biodata items. This is due to the relatively general nature of

biodata item content facilitating construct relevance regardless of SES. Such a prediction aligns

with the findings of Imus et al. (2010) whereby perceptual differences in accessibility between

Black and White participants did not correlate with biodata item slope parameters.

H2: The effects of minority status on DIF in biodata items will be partially explained by

SES such that minority status will initially predict higher item intercepts and this effect will be

weakened upon inclusion of SES.

Bias may also be observed in SJTs when they assess content that may have a relationship

with SES. However, this may not be related to the frame of reference effect since responding to

SJT items requires the identification of a response to a specific situation, rather than self-

evaluation against a reference group. This feature of SJT items makes the likelihood of item

intercept differences as a result of the frame of reference effect relatively low. Instead, bias in

SJTs may be more related to item accessibility. Kim et al. (2004) related differences in SES to

the differential opportunities hypothesis, which describes how disadvantaged minority group

members may not have access to specific opportunities required to demonstrate their standing on

a particular ability (Deutsch & Brown, 1964; Jachuck & Mohanty, 1974). This lack of access

may hinder the ability of a minority group member to respond to items related to a specific

context, in this case academic situations, when in reality the majority and minority group

members have similar standings on the latent trait. For example, if an SJT item asks about

activities that are associated with SES, that item may not be as relevant to disadvantaged

minority individuals. Due to this difference in item accessibility, the item in question may load

poorly onto the latent construct of situational judgment (e.g. Imus et al., 2010; Robert, Lee, & Chan, 2006). This should result in smaller item factor loadings among members of groups that are of lower SES.

H3: The effects of minority status on DIF in SJT items will be partially explained by SES such that minority status will initially predict smaller item factor loadings and this effect will be weakened upon inclusion of SES.

Vocational Interests

As mentioned by Imus et al., (2010), interests may be relevant to biodata assessments given the role interests play in shaping the experiences of an individual. It is possible that this logic may also be extended to SJTs given the relationship between experience and performance on such measures (Lievens & Motowidlo, 2016; McDaniel et al., 2001). Interests, as theorized by Holland (1997), can be described using a six-dimensional structure, where each dimension represents a distinct domain of behaviors an individual may be interested in. The domains are as follows: 1) Realistic – working with objects, also related to working outdoors, 2) Investigative – working with ideas, particularly in the sciences, 3) Artistic – following creative pursuits, such as writing and visual arts, 4) Social – working with and helping others, 5) Enterprising – taking on leadership or persuasive positions, often associated with pursuits related to economic growth, and 6) Conventional – preferring well-structured or traditional roles or environments. Nye, Su, Rounds, and Drasgow (2012) demonstrated that having a strong interest in a particular domain serves as a precursor to being motivated to engage in behaviors relevant to that domain. Interests help to motivate behavior by directing individuals toward particular goals, influencing the amount of effort expended on certain activities, and promoting perseverance on these activities over time. The direction of behavior has been shown in the past through studies demonstrating that interests can predict choice of an academic major or occupation (Eccles-Parsons, 1983;

Fouad, 1999; Holland, Fritzsche, & Powell, 1994). In terms of effort and persistence, Van Iddekinge, Putka, and Campbell (2011) found that interests moderately predicted effort and intentions to continue. Given the link between interests and motivation, interests should play a determining role in the experiences an individual pursues. Further, experiences resulting from interests may lead to the development of the constructs assessed by biodata and SJTs.

It stands to reason that if an individual consistently directs his or her attention towards behaviors and experiences as a function of interests, their responses to biodata and SJT items may reflect their interests to some extent. Longitudinal meta-analytic work by Low, Yoon, Roberts, and Rounds (2005) shows that vocational interests are quite stable during adolescence and early adulthood. Thus, the role of interests as a precursor to motivation of specific behaviors (e.g. Nye et al., 2012) should be stable during the period leading up to the assessment of biodata and SJT responses in the current study. If individuals are relying on vocational interests to direct behaviors during the period of time that biodata and SJTs focus on, then it is plausible that these noncognitive assessments may overlap with vocational interests. Further, Su and Nye (in press) describe that declarative and procedural knowledge can be developed through engagement in activities that align with an individual's vocational interests. This relationship between interests and knowledge bears some similarity to the theoretical account of SJT responding provided by Lievens and Motowidlo (2016). These researchers argue that selecting an appropriate behavior to a particular situation presented in an SJT item is partially determined by procedural knowledge gained through experience. Regarding biodata, Mael (1991) argues that responses to biodata items can reflect behaviors enacted as an adaptive response, which oftentimes coincides with the acquisition of knowledge. Given the fact that vocational interests may shape experiences during early adulthood and that the noncognitive assessments studied here are intended to capture past

experience and knowledge, it is likely that the constructs assessed by interests and noncognitive measures are related.

Though there are few empirical examples relating biodata and SJT assessment methods with vocational interests, several connections between them can be made based on theory and content. Situational judgment, as assessed in this study, may be related to investigative interests, since thoughtful analysis is an attribute both constructs hold in common. Beyond investigative interests, social interests may also be related to situational judgment. Most of the situations described in SJTs are embedded within a social context, and individuals who are motivated to pursue more social experiences may be more adept at choosing effective responses in a dilemma. In addition to having an interest in being around other people, social interests also describe being concerned for the welfare of others (Holland, 1997). This quality may also be found in biodata assessments that measure social responsibility and academic values, since behaviors related to these constructs could also be motivated by preserving the welfare of others. Enterprising interests are likely related to leadership given characteristics related to leading and persuading others. The pursuit of economic growth may also relate to persistence, given the overlap in exercising commitment over a period of time. Adaptability may also be related to enterprising interests since economically favorable opportunities may be associated with capitalizing on changes in your environment. Given the role of interests in guiding behavior, I suggest that:

H4: High social interests should predict higher levels of the latent traits of social responsibility, academic values, and situational judgment

H5: High investigative interests should predict higher levels of the latent traits of knowledge, continuous learning, and situational judgment

H6: High enterprising interests should predict higher levels of the latent traits of leadership, adaptability, and perseverance

H7: High conventional interests should predict higher levels of the latent trait of knowledge

If these propositions hold, past work investigating differences in interests between demographic groups may suggest demographic differences in responses to biodata and SJT items. Su, Rounds, and Armstrong (2009) meta-analyzed over 40 technical manuals of vocational interest measures and found substantial differences across genders. Specifically, they found that men tended to have more interest in realistic ($d = 0.84$) and investigative ($d = 0.26$) domains, whereas women had stronger artistic ($d = -0.35$), social ($d = -0.68$), and conventional ($d = -0.33$) interests. In contrast, enterprising interests were not substantially different between men and women. Tracey and Robbins (2005) found similar results regarding mean differences favoring males for realistic interests and females for social interests. As mentioned above, interests affect motivation and direct individuals towards certain activities, which can translate into life choices such as choice of academic major or occupation (Eccles-Parsons, 1983; Fouad, 1999; Holland, Fritzsche, & Powell, 1994). Lending credence to the idea that interests may lead to seeking experiences relevant to particular academic domains, such as the ones evaluated by biodata and SJTs, Su, Rounds, and Armstrong (2009) also analyzed differences in interest measures related to engineering, science, and mathematics and found that men favored all three. The researchers argued that these gender differences may have contributed to some of the gender disparities observed in specific occupations in STEM fields.

Recent meta-analytic work has also found racial differences in vocational interests between Whites and African Americans (Jones, Newman, Su, & Rounds, under review). On average, White respondents tended to score higher on realistic ($d = -.22$) and investigative ($d = -.16$) scales whereas African Americans scored higher on the social ($d = .26$), enterprising ($d = .18$), and conventional ($d = .28$) interest scales. Scores on the artistic scales were not

substantially different across these groups (Jones et al., under review). As discussed with gender, differences in interests between Whites and Blacks have manifested themselves in terms of occupational choice. African Americans tend to be represented more in conventional, social, and enterprising jobs and underrepresented in STEM fields, which correspond to investigative interests (Walker & Tracey, 2012). Given the gender and race differences on vocational interests, it is also likely that different subgroups will be motivated to pursue different sets of activities, which will then manifest in group differences on biodata and SJTs:

H8: The effect of minority and gender status on the latent traits assessed by biodata and SJTs will be partially explained by differences in vocational interests such that minority and gender status will predict standing on the latent traits assessed by biodata and SJTs, and this effect will be weakened by the inclusion of vocational interests in the model.

In addition to generating race and gender differences on the latent construct, group differences on vocational interests are also likely to result in bias on biodata. Specifically, it may be the case that interests relate to the frame of reference effect in their production of DIF on biodata items (Robert, Lee, & Chan 2006). Schneider's (1987) attraction-selection-attrition (ASA) model, specifically the attraction process, may help inform how interests can shape the local reference group an individual uses when responding. Schneider's (1987) model describes how individuals are attracted to organizations where they may find similar others. Though organizations are not the focus of this investigation, the underlying process of being attracted to similar others may still be informative, and is in fact based in Holland's (1959, 1997) work arguing that individuals choose professions that match their interests. In other words, individuals may choose to participate in experiences that match their interests and the other individuals in these environments are also likely to share those interests. If this is the case, then individuals may respond to questions about their experiences, like those used in biodata measures, using

others with similar interests as their frame of reference. Given the race and gender differences in vocational interests, this suggests that interests may mediate the effects of race or gender on DIF in biodata assessments.

H9: The effect of minority and gender status on DIF in biodata items will be partially explained by vocational interests such that group status will initially predict biodata item intercepts, and this effect will be weakened by the inclusion of vocational interests in the model.

Evaluation of these hypotheses expressed above was accomplished through the methods and proposed analyses described below, using the general approach depicted in Figure 2.

METHOD

Participants and Procedures

Participants consisted of college students who were admitted and chose to attend a large, Midwestern university. During the application process, 11,637 students completed a biodata assessment and SJT along with other common admissions requirements including providing demographic information. Admitted students who chose to enroll attended one of several orientation sessions, during which further survey data was collected in paper and pencil format. A subset of these admitted students, whose selection was based on orientation scheduling, completed a survey containing a vocational interests inventory based on the RIASEC model (Holland, 1997), as well as a parallel version of the SJT they took during the application process. The final sample consisted of 1,486 students, of which 616 were Male and 827 were female (43 did not provide a response). The racial composition of the sample consisted of 1,070 White, 158 Black, 106 Asian, 78 Hispanic, 58 Multiracial, 5 Native American, and 11 participants who did not specific their race. White, Black, and Asian participants were analyzed separately whereas all other participants were included in an "Other" category.

Measures

Demographics. Demographic data included in this study comes from data collected during the application process. Of note, race, gender, high school zip code, and Pell grant eligibility status were obtained.

Biodata. The biodata assessment used in the present study was developed to measure 12 dimensions identified via content analysis of university websites describing the attributes they hoped to develop in students (Oswald et al., 2004). Seven of these attributes, defined and presented with example items in Table 1, have been retained for the current study due to their high reliabilities and past work demonstrating their validity for predicting criteria like college

GPA. This version of the biodata assessment comes from the Student Behavior and Experiences Inventory, which also includes the SJT measure below (Oswald et al., 2004). Items were designed to ask about interests, hobbies, experiences, and relevant background information related to the construct being assessed (Imus, et al., 2010). Assessing such a range of content translates into a broad assessment of the construct that includes attitudes, beliefs, and past behaviors. Each of seven scales used 10 multiple-choice items (except Social Responsibility which had 9 items) designed to assess aspects of the individual's past experiences thought to be indicative of capabilities suited for a university context.

Table 1.

Dimensions assessed with the biodata and SJT Measures

Dimension title and definition	Sample item
Knowledge: Gaining knowledge and mastering facts, ideas and theories and how they interrelate, and the relevant contexts in which knowledge is developed and applied.	For class work, how often do you tend to skim the material, reading only the important points? a. Almost all the time b. Most of the time c. Sometimes d. Rarely e. Never
Continuous Learning: Being intellectually curious and interested in continuous learning. Actively seeking new ideas and new skills, both in core areas of study as well as in peripheral or novel areas.	In the past month, how many times have you looked for more information about something that you found interesting? a. Never b. Once or twice c. 3 to 5 times d. 6 to 10 times e. More than 10 times
Social Responsibility: Being responsible to society and the community, and demonstrating good citizenship. Being actively involved in the events in one's surrounding community, which can be at the neighborhood, town/city, state, national, or college/university level.	How many hours of volunteer work did you do in high school? a. 0 b. Between 1 and 10 c. Between 11 and 30 d. Between 31 and 75 e. More than 75

Table 1 (cont'd)

Leadership: Demonstrating skills in a group, such as motivating others, coordinating groups and tasks, serving as a representative for the group, or otherwise performing a managing role in a group.

When asked to do a class project with other students, how often do you take the lead and assign tasks or roles to people in the group?

    a. I am usually the one who assigns tasks or roles to get the work done

    b. More than half the time I end up assigning the tasks and roles

    c. About half the time I take the lead in assigning tasks and roles

    d. I rarely take the lead in assigning tasks and roles

    e. I never take the lead unless I have been assigned to do so

Perseverance: Committing oneself to goals and priorities set, regardless of the difficulties that stand in the way.

When encountering problems that take a long time to solve, how impatient do you tend to become?

    a. Extremely impatient

    b. Very impatient

    c. Somewhat impatient

    d. Slightly impatient

    e. Not at all impatient

Adaptability: Adapting to a changing environment (at school or home), dealing well with gradual or sudden and expected or unexpected changes.
Being effective in planning one's everyday activities and dealing with novel problems and challenges in life.

In the past, how difficult have you found it to adjust to major changes in your life (e.g. moving, a new school, a new job)?

    a. Extremely difficult

    b. Very difficult

    c. Difficult

    d. Not very difficult

    e. Not as all difficult

Academic Values: Having a well-developed set of values, and behaving in ways consistent with those values. In everyday life, this could mean being honest, not cheating (on exams or in committed relationships), and having respect for others.

In you first three years of high school, how often did you skip classes without a legitimate reason?

    a. Most of the time

    b. A lot

    c. Sometimes

    d. Once or twice

    e. Never

Table 1 (cont'd)

Situational Judgment: Making good decisions in various academic and social situations related to each of the above areas. Analyzing and choosing from among various alternative possible actions in problem situations.	You are part of a three-person group working on a class project with a quickly approaching deadline. One member of the team is not pulling his weight. He avoids assignments, complains about the amount of work that has to be done, and says the project doesn't really matter anyway. While you are all classmates, you seem to be the group leader. What would you do?
	a. Divide the workload among members of the group, making sure everyone knows that are responsible for their share. If the group member still does not pull his own weight, bring it up with the instructor.
	b. Speak with him in private and offer him moral encouragement to complete his portion of the project. If the group member still does not pull his own weight, bring it up with the instructor.
	c. Try to get the team member motivated to do his work. If that doesn't help the situation, just put more effort into the project yourself in order to complete it.
	d. Just do the group member's portion of the assignment in addition to your own, and tell the instructor about the situation.
	e. See if the person could be removed from your group.
	f. Consult with the non-problematic group about the most appropriate course of action, and then act on whatever you jointly decide.

Note. Table reproduced with permission from Prasad, Showler, Schmitt, Ryan and Nye (2016).

SJT. The SJT included in the current study was developed as a predictor of college performance, and is part of the Student Behavior and Experiences Inventory (Oswald et al., 2004) described above. This measure was also intended to reflect academically related capabilities, but did so through presenting scenarios with a list of possible actions. Specifically,

27

each item contained a dilemma that is commonly faced by college students, along with several possible responses to that dilemma. The same twelve dimensions originally assessed by the biodata measure are also assessed by the 25 scenarios presented in the SJT. Previous findings, however, found that a unidimensional model of situational judgment best represented responses to this kind of assessment despite agreement among researchers regarding the sorting of items into different dimensions (Oswald et al., 2004). As a result, analyses treated all items as belonging to the same scale, as has been done in the past (e.g. Schmitt et al., 2009).

Interests. The vocational interest measure used during orientation activities was the brief public domain RIASEC markers scale, developed and validated by Armstrong, Allison, and Rounds (2008). This measure assessed the six RIASEC vocational interest dimensions originally proposed by Holland (1997). Six items per dimension were used which asked about activities related to a particular dimension, such as "Set up and operate machines to make products," or "Sing in a band," for a total of 36 items. Participants used a five-point Likert-scale to indicate their level of interest in an activity from "Dislike very much" to "Like very much." Scores consisted of the sum of item scores for each RIASEC dimension, individually.

Median Local Income. In addition to information gathered from subjects during their application and orientation processes, 2010 median household income of the zip code for their high school was used to describe their socioeconomic status. Localized income data was retrieved from the U.S. Census Bureau website (United States Census Bureau, 2010). Income data were divided by a factor of 1,000 before inclusion in SEM analyses. Mplus documentation indicates that model estimation may be hindered in instances where some variable variances are much larger than other variables in the model (Muthén & Muthén, 2011). By dividing income values by 1,000 the resulting variance of the Median Local Income variable was closer in magnitude to other studied variables, facilitating model estimation.

Data Analysis

The hypotheses posed in the present research were evaluated using a multiple indicator multiple cause (MIMIC) model approach (Jöreskog, & Goldberger, 1975; Muthén, 1989). In addition to the explanation that follows, Figure 2 serves as an illustration of this approach. Before such analyses were conducted, two prerequisite steps were required. First, an adequate measurement model for all scales had to be estimated to promote subsequent structural model fit. Second, scale items had to be assessed for DIF as the MIMIC approach used here would not be identified if all items, including those that did not demonstrate DIF, were regressed onto explanatory covariates. An excessive number of estimated paths would be required to assess all items simultaneously for the substantive factors contributing to DIF (Woods, Oltmanns, Turkheimer, 2009). This section describes these prerequisite steps, as well as the implementation and interpretation of the final structural models.

A measurement model that fit well for each scale was important to promote overall model fit and the interpretability of results in subsequent analyses. The adequacy of model fit for these and subsequent analyses were assessed using the following rules of thumb: SRMR < .05, NNFI > .90, CFI > .90, and RMSEA < .08. However, these rules of thumb were used as guidelines rather than cutoffs and model fit was examined holistically as conditions may arise where an individual fit index may signal misfit unnecessarily (Nye & Drasgow, 2011; Schermelleh-Engel, Moosbrugger, & Müller, 2003). White males were used as the subgroup for which measurement models were tested and modified because these individuals served as the referent group for DIF analyses across both race and gender. For each scale, an initial measurement model was fit to the data whereby all items would load onto a single latent factor.

Initial fit of the Continuous Learning ($\chi^2(35) = 122.78$, RMSEA = .07, CFI = .92, NNFI = .90, SRMR = .04) scale was acceptable and the unidimensional model was used for further

analyses. All other scales required some form of modification to identify an appropriate measurement model for analyses. In cases of unsatisfactory initial model fit, standardized residuals and modification indices were assessed to identify where model fit could be improved. In many cases, misfit was identified due to additional content that was shared after accounting for the latent factor. This issue is common in many types of noncognitive assessments (Nye, Allemand, Gosling, Potter, & Roberts, 2016) and was addressed by correlating the errors of these items if this constraint seemed theoretically justified.

For the Academic Values scale, initial fit was unsatisfactory ($\chi^2(35) = 107.06$, RMSEA = .07, CFI = .88, NNFI = .85, SRMR = .05), and was improved through the incorporation of correlated errors between the items "In the past year, how many times have you copied someone else's work and submitted it as your own (at school or at work)?" and "In high school, how many times have you cheated on a school project, assignment, or test?" both of which related to the frequency of cheating on academic assignments ($\chi^2(34) = 70.04$, RMSEA = .05, CFI = .94, NNFI = .92, SRMR = .04). Initial fit of the Social Responsibility scale ($\chi^2(27) = 129.33$, RMSEA = .09, CFI = .92, NNFI = .89, SRMR = .05) was good, but subsequent estimation of the configural models for race ($\chi^2(108) = 494.76$, RMSEA = .10, CFI = .90, NNFI = .86, SRMR = .06) and gender ($\chi^2(54) = 427.40$, RMSEA = .10, CFI = .89, NNFI = .86, SRMR = .05) did not fit well. Correlated errors between the items "How many hours of volunteer work did you do while in high school?" and "In the past year, how many hours were you engaged in community service or volunteer activities?" were also estimated and the resulting model yielded improved measurement model fit ($\chi^2(26) = 89.43$, RMSEA = .07, CFI = .95, NNFI = .93, SRMR = .04). Further, estimating this model across both gender ($\chi^2(52) = 248.78$, RMSEA = .07, CFI = .94, NNFI = .92, SRMR = .04) and race ($\chi^2(104) = 314.60$, RMSEA = .07, CFI = .94, NNFI = .92, SRMR = .05) fit well. Again, these additional model constraints were justified given the shared

content of these items. Past research has demonstrated that these constraints are necessary when justified by the content of the items and/or their theoretical relationship (Cole, Ciesla, & Steiger, 2007). All correlated residuals included to improve measurement model fit were included in subsequent analyses as well.

Beyond the addition of correlated residuals, some scales required the removal of a problematic item to achieve appropriate measurement model fit. Initial fit for the Knowledge scale was poor ($\chi^2(35) = 116.39$, RMSEA = .07, CFI = .88, NNFI = .84, SRMR = .05) and both modification indices and residuals indicated that numerous aspects of the model were misspecified. Inspection of item descriptive statistics revealed a strong ceiling effect for the following item: "In your high school courses, how effective would you say you were at learning knowledge and mastering general concepts?" The ceiling effect was reflected in the relatively weak loading of this item onto the latent Knowledge construct ($\lambda = .41$). Removal of this item resulted in acceptable model fit ($\chi^2(27) = 72.22$, RMSEA = .06, CFI = .92, NNFI = .90, SRMR = .04). Thus, this item was excluded from further analyses. The Perseverance scale also did not fit well with a single factor ($\chi^2(35) = 167.98$, RMSEA = .09, CFI = .83, NNFI = .78, SRMR = .05). Examination of standardized residual covariances revealed that the item "How often have you achieved a personal goal that seemed unattainable at first?" significantly related to several other items in a way that was not captured by the latent factor. Removal of this item meaningfully improved the fit of the Perseverance scale ($\chi^2(27) = 95.84$, RMSEA = .07, CFI = .90, NNFI = .87, SRMR = .04). Based on the findings of Oswald et al. (2004) we modelled the SJI with a single latent factor, which did not appear to fit well ($\chi^2(275) = 392.61$, RMSEA = .03, CFI = .79, NNFI = .77, SRMR = .05). Improved fit was achieved by the removal of the item "You have very much wanted to be a teacher, but you failed the entrance exam into the College of Education. This exam is not given again for a year. What would you do?" due to excessive

residual correlation with four other items. Additionally, the error terms of two pairs of items were correlated based on similarity in content[1]. As a result of these modifications, model fit was closer to acceptable levels ($\chi^2(250) = 306.36$, RMSEA = .02, CFI = .89, NNFI = .88, SRMR = .04).

Results also indicated that a two-factor model fit the Adaptability and Leadership scales best. The initial fit of a unidimensional model for the Adaptability scale was poor ($\chi^2(35) = 166.08$, RMSEA = .09, CFI = .80, NNFI = .74, SRMR = .06), and examination of residual correlations appeared to indicate distinct subsets of items. Four items that related to changes to an individual's normal routine like "In the past, how difficult have you found it to adjust to major changes in your life (e.g., moving, a new school, a new job)?" appeared to relate highly in a way not captured by a single latent construct. A two-factor model appeared to fit the Adaptability scale where the four routine-related items loaded onto one factor and all other items loaded onto the other factor, with a correlation included between factors ($\chi^2(34) = 69.25$, RMSEA = .05, CFI = .95, NNFI = .93, SRMR = .04). It should be noted that the correlation between latent factors in the two-factor model of Adaptability was quite high ($r = .57$), suggesting that these factors may not be meaningfully distinct. However, in subsequent analyses these factors were individually

---

[1] The first pair of correlated items were "You are searching for a major that interests you and think you might be interested in psychology. You do not know much about preparation to be a psychologist or what kinds of opportunities exist for careers in this area. What action would you take?" and "You are interested in several different classes/disciplines, but don't know anything about future educational or career opportunities in these areas. What steps would you take to get informed?". The second pair of correlated items were "You are part of a committee to reduce cross-cultural tension in your dorm. A group of students in your dorm complain to you that people always wish them 'Merry Christmas' or 'Happy Easter' when these holidays are not meaningful to them. They request that their differences be respected. How would you address this problem?" and "A friend on your floor is always organizing 'social' activities including trips to local bars. Aside from the fact that this person is underage and failing some classes, you realize that the individual is drinking half a dozen or more drinks at least three or four times a week. No one else seems to know or be concerned about the person. What would you do?".

resulted in good model fit, and will be referred to as Routine Adaptability (four routine-related items; $\chi^2(2) = 4.02$, RMSEA = .05, CFI = .99, NNFI = .98, SRMR = .02) and Discrete Adaptability (all other items; $\chi^2(9) = 20.04$, RMSEA = .05, CFI = .96, NNFI = .94, SRMR = .03) in further analyses.

In addition, a single-factor model of Leadership also did not fit well ($\chi^2(35) = 266.98$, RMSEA = .12, CFI = .83, NNFI = .78, SRMR = .07). Examination of residual correlations revealed the possibility of a factor representing experience with past leadership positions (e.g. "The number of high school clubs and organized activities (such as band, sports, newspapers, etc.) in which you took a leadership role was:") whereas the other seemed to relate to leadership behaviors (e.g. "During the past year, how often have you taken charge of a group that you were in, without being asked?"). Modelling Leadership with two factors ($\chi^2(34) = 148.17$, RMSEA = .09, CFI = .92, NNFI = .89, SRMR = .05) yielded acceptable model fit. However, the factor that appeared more related to leadership behaviors did not fit well on its own ($\chi^2(9) = 95.43$, RMSEA = .14, CFI = .88, NNFI = .80 SRMR = .06). Two items within this factor still demonstrated evidence of a strong residual correlation, and both related to tasks focused on organizing the group[2]. A two-factor model of the Leadership scale with the included residual correlation yielded satisfactory model fit ($\chi^2(33) = 88.91$, RMSEA = .06, CFI = .96, NNFI = .95, SRMR = .04), and the factors were highly correlated ($r = .70$). Independent models of these factors fit well, and are treated independently in further analyses as Leadership Positions ($\chi^2(2) = 0.41$, RMSEA = .00, CFI = 1.00, NNFI = 1.00, SRMR = .00) and Behavioral Leadership ($\chi^2(8) = 24.98$, RMSEA =

---

[2] The two leadership behavior items were "In the past year, how many times have you been responsible for assigning tasks and setting deadlines for other people?" and "How many times in the past year have you set the schedule (time and/or tasks) for groups in which you have worked?"

.07, CFI = .98, NNFI = .96, SRMR = .03). With suitable measurement models identified, scale items were assessed for DIF.

DIF analyses were conducted to identify which items displayed DIF that could be explained via a MIMIC model. As mentioned previously, testing all items for DIF using the MIMIC approach would require an excessive number of estimated paths (Woods, Oltmanns, Turkheimer, 2009). As such, multiple group analyses for race and gender were conducted separately to flag items for DIF and items flagged as functioning differently across in any of the groups were examined further using the MIMIC model. Before flagging items for DIF, a suitable referent item had to be found for each scale. This involved estimating a constrained baseline model (i.e. all item loadings and intercepts constrained to equality across groups) followed by models where an individual item was freely estimated across groups (Stark, Chernyshenko, & Drasgow, 2006). Models were estimated until an item was found that produced an increase in the CFI < .002 (Meade, Johnson, & Braddy, 2008), signifying an item that could serve as a suitable referent item. If an item met this condition for analyses of both race and gender, then this item was used as a referent in further analyses. A free baseline model (i.e. all item loadings and intercepts freely estimated) was then estimated for each scale, followed by models where an individual item was constrained to equality across groups. When constraining an item resulted in a decrease of CFI > .002 (i.e. constraining an item to equality across groups significantly worsened model fit), that item was flagged for DIF (for further description and rationale, see Stark, Chernyshenko, & Drasgow, 2006; Nye & Drasgow, 2011). The items flagged for DIF based on race and gender can be found in appendices A1-A10 with each table corresponding to a specific scale.

Testing for Uniform and Nonuniform DIF. The main analysis used in the current study was to model responses to biodata and SJT items using MIMIC structural equations models

(Muthén, 1989). MIMIC models are useful tools for assessing DIF, particularly when the goal is to explain DIF rather than just detect it. As with CFA more generally, the MIMIC model estimates the latent trait underlying a measure and then models participants' standings on the latent trait. This makes it possible to differentiate true differences in the latent trait from bias in the measure. The application of a MIMIC model as well as the model building approach described below are depicted in Figure 2. All of the following tests described below were conducted within the Mplus version 7.4 software package (Muthén & Muthén, 2011).

As described above, uniform DIF implies that there are consistent differences between groups across the entire response scale for a particular item. Specifically, uniform DIF is reflected in differences in the intercepts of the items. Uniform DIF can be detected if a grouping variable (e.g., race, gender) significantly predicts a response to an item while also controlling for its relationships with the latent trait (Woods, 2009).

Nonuniform DIF, on the other hand, describes how the response scale may be different across groups. This refers to differences in the factor loadings of the items on a latent trait. The detection of nonuniform DIF in the MIMIC model required the computation of an interaction term between the latent trait and the grouping variable as described by Woods and Grimm (2011). This interaction term is then used to predict responses to a particular item and a significant path suggests that the response scale for that item is dissimilar across groups. In their work, Woods and Grimm (2011) computed interaction terms using the XWITH command in Mplus as it is the approach recommended in the Mplus documentation for computing an interaction between a latent continuous variable and observed categorical variable (Muthén & Muthén, 2011). However, the XWITH command from Mplus only allows for interaction terms to be used as predictors and as such will only be used to predict item responses (i.e. cannot be used to model correlations between other predictor variables). Further, interaction terms estimated in

this way require specification of random slopes and intercepts, which limit model fit information to AIC and BIC values (Muthén & Muthén, 2011).

Explanation of DIF, a Model Building Approach. A model building approach was employed here to help explain instances where either uniform or nonuniform DIF is detected. After all scale items were individually evaluated for DIF, a baseline DIF model was estimated whereby all items flagged for DIF were regressed onto grouping variables and interaction terms to model uniform and nonuniform DIF, respectively. Regressions of the latent factor onto the dichotomous grouping variables were also included. This baseline model was used to identify the specific DIF effects as described above. A second model was then estimated where SES variables (i.e. Pell grant eligibility, median local income) were added to the baseline model. Specifically, items flagged for DIF as well as the latent factor of the target scale were regressed onto these SES variables. Finally, a third model was estimated in which the vocational interest scale scores are added to the model, with regressions of the DIF items and the latent factor of the scale onto each interest scale.

No correlations among the explanatory covariates (i.e. dichotomous grouping variables, interactions, SES variables, vocational interest scale scores) were specified in any of the three explanatory models. Muthén (1989) explains that MIMIC models should be conditioned on exogenous explanatory variables (Jöreskog & Goldberger, 1975). Specification of correlations between predictors within the MIMIC model would treat these variables as endogenous, with their variances and errors estimated as model parameters (Muthén & Muthén, 2011; Muthén 2012), violating the original suggestion by Muthén (1989). Further, inclusion of these correlations may be a source of model over-identification (Hauser, Robert, & Goldberger, 1971). However, correlations between predictors were important for evaluating the hypotheses, as explained further below. These correlations were obtained from the observed correlations

between predictors obtained outside of the estimated model (Muthén, 2012). Most relevant to hypotheses posed here, biserial correlations were reported between the dichotomous grouping variables and both SES variables and vocational interest scales. Additionally, interaction terms from the third model were saved following model estimation and these saved terms were correlated with both SES variables and vocational interest scales as well.

In instances where the third model revealed what appeared to be a meaningful change in a demographic variable predicting a latent factor, follow-up model comparisons were conducted to evaluate the strength of the explanatory effect. When a regression of the latent factor on a demographic variable appeared meaningful, this model was compared to an alternative model where the path from the demographic variable to the latent factor was fixed to the corresponding value observed in the first mimic model (i.e. did not include SES variables or vocational interests). Given that the estimation of the interaction between the latent score and demographic variables limits model fit information to AIC and BIC values (Muthén & Muthén, 2011), rules of thumb from the literature describing those model fit statistics were used to assess whether an explanation of an effect was significant. Raftery (1995) shows that a BIC decrease of less than 2 constitutes weak evidence of model improvement, a decrease 2 to 6 is good evidence, 6 to 10 is strong evidence, and greater than 10 is very strong evidence. For AIC, Burnham and Anderson (2004) suggest that a decrease of less than 2 is weak evidence, a decrease of 4 to 7 is strong evidence, and a decrease of greater than 10 is very strong evidence.

The model building approach proposed here is a novel approach to help tie observed demographic DIF effects to potential explanatory variables. Broadly speaking, an observed demographic DIF effect was considered somewhat explained if two criteria are met. First, the inclusion of explanatory variables must have led to an observable decrement in the uniform or non-uniform DIF effect from the baseline model. Second, the explanatory variable must have

both predicted responses to the DIF items and correlated with the variable signifying the uniform

or nonuniform DIF effect. Additionally, given the number of explanatory variables, the

incremental approach of including SES variables followed by vocational interest variables was

proposed to help distinguish the effects of these two predictors.

RESULTS

Descriptive statistics, intercorrelations, and Cronbach's alphas are presented in Table 2. Reliabilities for the Values ($\alpha = .57$) and SJT ($\alpha = .65$) scales were somewhat low, but all scales seem reliable enough to include in subsequent analyses. Small to moderate intercorrelations suggest that biodata, SJT, and the RIASEC interest scales measure distinct, but related constructs. Importantly, several of the categorical demographic variables correlated with some of the explanatory covariates. As expected, Black participants (coded as Black = 1, White = 0) generally came from areas with lower median income ($r = -.41, p < .001$) and were more likely to be Pell grant eligible ($r = .39, p < .001$). In addition, female participants (coded as female = 1, male = 0) were slightly more likely to come from areas with lower median income ($r = -.08, p = .018$) and Asian participants (coded as Asian = 1, White = 0) generally came from wealthier areas ($r = .18, p < .001$). Consistent with previous research (Su, Rounds, & Armstrong, 2009), gender was negatively associated with Realistic ($r = -.53, p < .001$), Investigative ($r = -.17, p < .001$), Enterprising ($r = -.34, p < .001$), and Conventional interests ($r = -.24, p < .001$). Gender was positively associated with Artistic ($r = .14, p < .001$) and Social Interests ($r = .42, p < .001$). Black respondents tended to have slightly higher Artistic ($r = .10, p = .019$) and Social ($r = .12, p = .005$) interests. Asian respondents had stronger Conventional interests ($r = .13, p = .009$). Nevertheless, the effect sizes of the correlations between race and vocational interests were generally small.

Biserial correlations could also be examined to assess the differences on biodata and SJT scales between demographic groups. Females scored significantly higher on Leadership ($r = .08, p = .013$), Values ($r = .12, p = <.001$), Social Responsibility ($r = .28, p = <.001$), Perseverance ($r = .20, p = <.001$), and SJT ($r = .28, p = <.001$). Lower scores were obtained by females on the

Table 2.
Descriptive statistics and intercorrelations of studied variables

Variable	M	SD	1	2	3	4	5	6	7	8	9
1) Leadership	3.43	0.71	(.85)								
2) Continuous Learning	3.55	0.58	.38**	(.82)							
3) Knowledge	3.54	0.44	.30**	.55**	(.72)						
4) Values	4.27	0.36	.17**	.26**	.44**	(.57)					
5) Social Responsibility	3.72	0.67	.50**	.24**	.22**	.17**	(.77)				
6) Perseverance	4.00	0.43	.42**	.46**	.56**	.41**	.32**	(.77)			
7) Adaptability	3.56	0.43	.36**	.35**	.45**	.30**	.27**	.52**	(.71)		
8) Situational Judgment	3.94	0.26	.19**	.23**	.34**	.43**	.21**	.36**	.24**	(.65)	
9) Female	0.57	0.49	.08*	-.13**	.00	.12**	.28**	.20**	.00	.28**	
10) Black	0.11	0.31	-.05	.05	-.10*	-.11*	-.07	.15**	-.06	.02	.09**
11) Asian	0.07	0.26	-.06	.01	-.10*	-.12*	.15**	-.26**	-.19**	-.13**	-.02
12) Other	0.10	0.30	-.09*	-.02	-.07	-.04	-.09*	.02	-.05	-.03	-.03
13) Pell Eligibility	0.34	0.47	-.01	.08**	.02	-.08**	-.05*	.11**	.03	.01	.06
14) Med. Local Income	66628.40	27525.11	-.01	-.04	-.06*	.02	.00	-.08**	-.05	-.05	-.08*
15) Realistic	2.20	0.82	-.06*	.09**	.00	-.11**	-.11**	-.12**	-.03	-.14**	-.53**
16) Investigative	3.12	0.96	-.01	.14**	.17**	.01	-.01	.00	.04	.03	-.17**
17) Artistic	2.79	0.92	.07**	.20**	.04	.00	.05	.02	-.07**	.06*	.14**
18) Social	3.16	0.80	.16**	.07**	.11**	.12**	.24**	.18**	.11**	.20**	.42**
19) Enterprising	3.00	0.86	.11**	-.01	-.03	-.08**	.00	.04	.08**	-.05	-.34**
20) Conventional	2.54	0.81	.02	.04	.08**	-.03	-.03	.00	.05*	-.04	-.24**

Note. Reliabilities presented along the diagonal in parentheses. Female denotes the dummy coded variable representing Gender (coded as Female = 1, Male = 0). Black, Asian, Other represent dummy coded variables representing Race categories (all coded as Minority = 1, White = 0). ** $p < .01$; * $p < .05$.

Table 2. (Cont'd)

Variable	10	11	12	13	14	15	16	17	18	19	20
1) Leadership											
2) Continuous Learning											
3) Knowledge											
4) Values											
5) Social Responsibility											
6) Perseverance											
7) Adaptability											
8) Situational Judgment											
9) Female											
10) Black											
11) Asian	-.10**										
12) Other	-.12**	-.09									
13) Pell Eligibility	.39**	-.02	.10**								
14) Med. Local Income	-.41**	.18**	-.07	-.22**							
15) Realistic	-.04	.09	.00	-.03	-.01	(.85)					
16) Investigative	-.08	.07	.03	-.04	-.05*	.36**	(.84)				
17) Artistic	.10*	.02	.10*	.01	-.06*	.08**	.13**	(.78)			
18) Social	.12**	-.00	-.01	.06*	-.04	-.09**	.08**	.28**	(.77)		
19) Enterprising	-.06	-.03	.01	-.03	.08**	.27**	-.01	-.03	.07**	(.83)	
20) Conventional	-.06	.13**	-.07	-.04	.03	.39**	.11**	-.10**	.00	.52**	(.82)

Note. Reliabilities presented along the diagonal in parentheses. Female denotes the dummy coded variable representing Gender (coded as Female = 1, Male = 0). Black, Asian, Other represent dummy coded variables representing Race categories (all coded as Minority = 1, White = 0). ** $p < .01$; * $p < .05$.

Continuous Learning scale ($r = -.13$, $p = <.001$), and no score differences were observed for Knowledge and Adaptability. Black participants scored significantly lower on Knowledge ($r = -.10$, $p = .018$), and Values ($r = -.11$, $p = .011$), but scored higher on Perseverance ($r = .15$, $p = <.001$). No differences were observed on the Leadership, Continuous Learning, Social Responsibility, Adaptability, and SJT scales between Black and non-Black participants. Finally, Asian participants scored significantly lower on Knowledge ($r = -.10$, $p = .038$), Values ($r = -.12$, $p = .013$), Adaptability ($r = -.19$, $p = <.001$), and SJT ($r = -.13$, $p = .009$), but scored higher on Social Responsibility ($r = .15$, $p = .002$). Even though significant score differences were observed, most were quite small and in the case of gender almost all favored females.

Assessment of Differential Item Functioning

Tables A1-A10 display item content, configural model fit for race and gender, as well as fit for each model used to flag items for DIF across the scales included in this study. Two scales, Routine Adaptability and SJT, could not be fully analyzed as planned. For Routine Adaptability, a suitable referent item could not be identified for analyses based on Race, as all items produced a change in CFI of greater than .005, which is too far beyond the cutoff of .002 to consider using as a referent. As a result, further analyses of this scale only examined differences across gender. A suitable referent was found for the SJT, but an acceptable configural model fit could not be achieved for race ($\chi^2(1000) = 1311.69$, RMSEA = .03, CFI = .83, NNFI = .81, SRMR = .05). This suggests that although a well-fitting measurement model could be found with white males, this model did not fit well when applied across races[3]. However, configural model fit of the SJT

---

[3] Further analyses were conducted to examine the inability to find configural invariance across races. First, the measurement model found with white males was estimated within each racial group. Acceptable model fit was found among White participants ($\chi^2(250) = 395.98$, RMSEA = .02, CFI = .89, NNFI = .88, SRMR = .03), but not among any other racial group (Black: $\chi^2(250) = 326.21$, RMSEA = .04, CFI = .44, NNFI = .38, SRMR = .07; Asian: $\chi^2(250) = 294.72$,

was acceptable for gender ($\chi^2(502) = 692.88$, RMSEA = .02, CFI = .88, NNFI = .87, SRMR = .04). Thus, further analyses of the SJT focused only on gender. Due to the inability to establish configural invariance in SJT based on race, H3, which suggested that factor loading differences across racial groups could be accounted for by SES, could not be tested. For all other scales, DIF was identified in at least one item and the MIMIC model analyses were conducted as planned to evaluate the hypotheses that account for these effects.

Evaluation of Hypotheses

Broadly, the present work sought to identify instances of differential item functioning and latent factor score differences between demographic groups, and attempted to explain why such differences occur. H1 stated that minority respondents may have lower standings on the latent factors measured by biodata and SJT scales and that these differences would be attenuated after accounting for SES. Across all scales, being Pell grant eligible only significantly predicted a higher standing on the latent Continuous Learning factor ($\beta = .21$, $p = .003$), and median local income only predicted a lower standing on the Leadership Positions factor ($\beta = -.07$, $p = .031$). In neither case did these effects coincide with meaningful differences in the latent factors across demographic groups. It should also be noted that the inclusion of SES variables rendered some demographic differences in the latent factors nonsignificant. However, in each of these cases the SES variables themselves did not significantly predict the latent factor, suggesting that the

---

RMSEA = .04, CFI = .77, NNFI = .75, SRMR = .08; Other: $\chi^2(250) = 294.78$, RMSEA = .03, CFI = .78, NNFI = .75, SRMR = .07). Examination of modification indices and standardized residuals did not indicate clear causes of model misfit in the measurement models of minority groups. However, configural invariance across races was also tested using the full applicant sample ($N = 11,637$) and yielded acceptable model fit ($\chi^2(1000) = 2159.85$, RMSEA = .02, CFI = .90, NNFI = .89, SRMR = .02), suggesting that the inability to find evidence for configural invariance is a function of using a reduced sample (for whom vocational interest data was available) more so than substantive differences in situational judgment across groups. The issue of sample size is further discussed in the limitations section of the discussion.

change in significance was likely due to error in the estimate of demographic group effects more so than a meaningful explanation of an effect by SES. With all of this in mind, there was minimal support for H1.

H2 stated that item intercepts on the biodata scales would vary across racial minority groups and this variation would be partially explained by SES. Of the 37 biodata items flagged for DIF, nine were significantly predicted by SES variables after accounting for the latent factor. However, only one item ("How often do you ask a teacher or classmate questions that go beyond the material but are still relevant to the topic (either in or out of class)?", Table B4) displayed the expected effect of attenuating an initially high item intercept among Black respondents ($\beta = .27$, $p = .001$), though the intercept difference remained significant ($\beta = .20$, $p = .023$) after inclusion of median local income ($\beta = -.06$, $p = .022$). Other intercept differences varied across demographic groups and were partially explained by SES but were not consistent with the hypothesized effects. A relatively low item intercept among Asian respondents for the aforementioned item ($\beta = -.20$, $p = .044$) was attenuated ($\beta = -.19$, $p = .079$) by inclusion of median local income ($\beta = -.06$, $p = .022$), but the change in statistical significance was associated with a minimal change in effect size. Similarly, for the item "In your first three years of high school, how often did you skip classes without a legitimate reason?" (Table B9) the inclusion of Pell grant eligibility ($\beta = -.18$, $p = .002$) decreased an initially significant intercept difference across male and female respondents (initial model: $\beta = -.13$, $p = .016$; secondary model: $\beta = -.08$, $p = .130$), but the actual change in effect size was small, suggesting that the explanatory nature of Pell grant eligibility was not meaningful in this case. For the item "When asked to do a class project with other students, how often do you take the lead and assign tasks or roles to people in the group?" (Table B5) inclusion of Pell grant status seemed to clarify an initially non-significant intercept among Asian respondents (Initial model: $\beta = -.21$, $p = .080$; Secondary model: $\beta = -.25$,

$p = .045$). However, this effect may not be practically important given the small change in effect

size and borderline statistical significance. Overall, H2 received little support as the occurrence

of SES variables predicting item responses did not have substantial effects on intercept

differences across demographic groups.

Several hypotheses also posited an association between Vocational Interests and the

latent traits measured by biodata and SJT scales (H4 – H7; summarized in Table 3).

Table 3.
Correspondence of RIASEC dimensions with biodata and SJT

RIASEC dimension	Corresponding biodata or SJT dimension(s)
Realistic	
Investigative	Continuous Learning, Knowledge, Situational Judgment
Artistic	
Social	Social Responsibility, Academic Values, Situational Judgment
Enterprising	Leadership, Adaptability, Perseverance
Conventional	Knowledge

These hypotheses were evaluated using the third MIMIC model in Figure 2 for each scale, part

of which included regressions of the scale latent factor onto each vocational interest scale. The

results of these analyses are summarized in Table 4 and the full results are provided in Appendix

B for each scale. Overall, the vocational interest scales predicted each noncognitive measure as

expected, with some notable exceptions. Social interests predicted Social Responsibility ($\beta = .13$,

$p = <.001$), Academic Values ($\beta = .12$, $p = .001$), and Situational Judgment ($\beta = .19$, $p < .001$) as

expected. Additionally, Social interests predicted Leadership Behaviors ($\beta = .16$, $p < .001$),

Leadership Positions ($\beta = .15$, $p < .001$), Knowledge ($\beta = .13$, $p = .006$), Perseverance ($\beta = .13$, p

$= .003$), Discrete Adaptability ($\beta = .14$, $p = .008$), and Routine Adaptability ($\beta = .13$, $p = .003$).

Given that the expected relationships were observed, these results support H4. Investigative

interests predicted Situational Judgment ($\beta = .07$, $p = .022$), Continuous Learning ($\beta = .10$, $p =$

$.004$), and Knowledge ($\beta = .13$, $p = .005$) as hypothesized providing support for H5. Enterprising

45

interests were hypothesized and found to predict Behavioral Leadership (β = .21, p = <.001), Leadership Positions (β = .12, p = .002), Discrete Adaptability (β = .12, p = .047), Routine Adaptability (β = .11, p = .013), and Perseverance (β = .13, p = .006). Unexpectedly, Enterprising interests also negatively related to Continuous Learning (β = -.08, p = .03). Overall, these results support H6. Finally, Conventional interests were found to predict Knowledge (β = .10, p = .049) as expected. Conventional interests also predicted Continuous Learning (β = .09, p = .020) and Discrete Adaptability (β = .15, p = .010) and were negatively related to Behavioral Leadership (β = -.12, p = .002) and Routine Adaptability (β = -.11, p = .010). In spite of the preponderance of unexpected relationships, H7 received support. Beyond examination of these hypothesized relationships, Realistic and Artistic Interests did demonstrate some predictive utility. Realistic interests negatively predicted Values (β = -.11, p = .010), Knowledge (β = -.12, p = .038), Perseverance (β = -.15, p = .001), Discrete Adaptability (β = -.21, p < .001), and Situational Judgment (β = -.12, p = .002). Artistic Interests only predicted Continuous Learning (β = .19, p < .001). Though in some cases the specific relationships between vocational interests and both biodata and SJT did not appear exactly as expected, this set of results as a whole

Table 4.
Summary of regressions of biodata latent factor scale scores on vocational interests

Vocational Interest Scale	Behavioral Leadership	Leadership Positions	Knowledge	Continuous Learning	Values
Realistic	-.08 (.04)	-.01 (.04)	-.12* (.06)	-.02 (.04)	-.11* (.04)
Investigative	-.01 (.04)	-.03 (.03)	.13** (.05)	.10** (.03)	.05 (.04)
Artistic	.05 (.03)	.04 (.03)	-.03 (.05)	.19** (.03)	.00 (.04)
Social	.16** (.04)	.15** (.03)	.13** (.05)	.05 (.04)	.12** (.04)
Enterprising	.21** (.04)	.12** (.04)	-.04 (.05)	-.08* (.04)	-.07 (.04)
Conventional	-.12** (.04)	-.01 (.04)	.10* (.05)	.09* (.04)	.07 (.04)

Note. ** p < .01, * p < .05. Effects presented are standardized regression weights with standard errors in parentheses. Each set of regression weights corresponding a particular biodata latent factor are from the final model of that biodata scale. Please see Appendix B for the full set of results for each scale.

Table 4 (cont'd)

Vocational Interest Scale	Perseverance	Discrete Adaptability	Routine Adaptability	Social Responsibility
Realistic	-.15** (.05)	-.21** (.06)	-.02 (.05)	-.03 (.04)
Investigative	.08 (.04)	.04 (.06)	.00 (.04)	.05 (.03)
Artistic	-.04 (.04)	.03 (.05)	-.06 (.04)	-.05 (.03)
Social	.13** (.04)	.14** (.05)	.13** (.04)	.13* (.03)
Enterprising	.13** (.05)	.12* (.06)	.11* (.04)	.06 (.04)
Conventional	-.01 (.05)	.15* (.06)	-.11* (.04)	-.02 (.04)

Note. ** $p < .01$, * $p < .05$. Effects presented are standardized regression weights with standard errors in parentheses. Each set of regression weights corresponding a particular biodata latent factor are from the final model of that biodata scale. Please see Appendix B for the full set of results for each scale

suggests a meaningful relationship between interests and the constructs measured here using noncognitive assessments.

H8 posited that the relationships between vocational interests and the biodata and SJT scales would help explain initially observed differences between demographic groups. This appeared to be the case for female respondents who initially demonstrated a higher standing on the latent Values factor ($\beta = .23$, $p = .001$). After inclusion of vocational interests, a higher standing on the Values factor among females was eliminated ($\beta = .02$, $p = .779$). Further, this model fit substantially better than a model where the relationship between the gender variable and the Values factor was fixed to the value observed in the first MIMIC model, which did not include SES variables or vocational interests (constrained model: AIC = 29293.07, BIC = 29851.18; unconstrained model: AIC = 29297.65, BIC = 29850.54). Of note, the change in AIC signified strong evidence. As mentioned previously, both Realistic and Social values predicted the latent Values factor, and both were associated with being female. A similar pattern was seen for the Discrete Adaptability factor, whereby females initially had a higher standing on this factor ($\beta = .36$, $p < .001$) but this higher standing was weakened ($\beta = .21$, $p = .071$) after accounting for vocational interests. Constraining this regression parameter to the value of the

first MIMIC model appeared to worsen model fit compared to when the regression was freely estimated (constrained model: AIC = 18485.55, BIC = 19054.09; unconstrained model: AIC = 18486.43, BIC = 19060.19), with the change in BIC signifying strong evidence. This reduction is likely due primarily to Social and Realistic interests because gender differences in these interest dimensions mirrored the gender differences in Discrete Adaptability (e.g. females tend to have higher Social Interests; Social Interests predict higher Discrete Adaptability; $\beta = .14$, $p = .008$). Further, though females ultimately demonstrated higher Perseverance ($\beta = .27$, $p = .005$), this effect was meaningfully larger ($\beta = .39$, $p < .001$) before the addition of vocational interests. However, freely estimating this regression did not clearly lead to better model fit than when the regression parameter was constrained (constrained model: AIC = 25538.47, BIC = 26320.87; unconstrained model: AIC = 25538.77, BIC = 26326.38). This suggests that the change in the regression weight of gender predicting perseverance is not significant. Situational Judgment demonstrated a similar set of findings whereby females had a higher standing on the latent trait ($\beta = .33$, $p < .001$) even after including vocational interests, but this effect was meaningfully larger in the initial model ($\beta = .49$, $p < .001$). However, freely estimating the parameter of the SJT factor regressed onto the gender dummy coded variable also did not clearly lead to better model fit than when it was constrained (constrained model: AIC = 73665.14, BIC = 74186.74; unconstrained model: AIC = 73663.56, BIC = 74190.37), suggesting that the observed change in regression weights between models is not significant. It should also be noted that after incorporation of SES and vocational interests, females still had a moderately lower standing on Routine Adaptability ($\beta = -.43$, $p < .001$) and a moderately higher standing on Social Responsibility ($\beta = .41$, $p < .001$). Overall, it does appear as though vocational interests help explain some of the true score gender differences across the biodata scales.

Though the explanatory power of vocational interests was demonstrated for some of the latent score differences based on gender, the effects of minority group status were largely independent of vocational interests. Across all scales, there were no observed differences in latent scores across minority groups that were meaningfully decreased after including vocational interests in the model. This is likely due to the relatively small differences in vocational interests observed across the racial subgroups. However, significant differences in latent scores as a function of minority group status should be highlighted. In the final MIMIC models that included SES and vocational interests, Black participants still had a moderately lower standing on Leadership Behaviors (β = -.40, p = .002), Knowledge (β = -.46, p = .002), and Discrete Adaptability (β = -.53, p = .003). Asian participants had a moderately higher standing on Social Responsibility (β = .40, p < .001), a moderately lower standing on Knowledge (β = -.41, p = .007) and Leadership Behaviors (β = -.46, p = .002), and a significantly lower standing on Discrete Adaptability (β = -.74, p < .001). Given some evidence of vocational interests explaining differences based on gender but not across minority groups, H8 received moderate support.

Lastly, H9 stated that observed differences in the item intercepts across demographic groups would be partially explained by vocational interests. Three items from the Discrete Adaptability scale (Table B4) demonstrated a similar pattern of intercept differences based on gender. Initially, the item "How often have you failed to meet responsibilities because you had taken on too much?" demonstrated a significantly lower item intercept for females (β = -.15, p = .018). This effect was diminished (β = -.06, p = .543) upon inclusion of vocational interests, with Artistic interests being the primary contributor (β = -.11, p = .001). It should be noted that though the overall change in effect size was small, proportionally the effect was roughly halved, which signifies that the intercept difference being explained by Artistic interests may be meaningful.

The item "How difficult has it been for you to continue with something after being interrupted and having to take care of something else?" also initially demonstrated a lower intercept for females (β = -.17, p = .013) that was attenuated (β = .00, p = .976) by vocational interests. The relevant interests here were Realistic (β = .15, p = .001), Enterprising (β = -.09, p = .041), and Conventional (β = -.09, p = .031) interests. Finally, the item "In the past, how difficult has it been for you to change your study habits to improve on a skill or to do better in a class" exhibited a lower intercept for females (β = -.13, p = .046) that was attenuated (β = -.09, p = .347) by the inclusion of Investigative (β = .08, p = .014) and Artistic (β = -.13, p < .001) interests. For this item a small observed change in effect size was also proportionally a meaningful one (roughly a reduction of 30%), which may be meaningful in terms of item functioning.

Gender-based DIF in two Continuous Learning items also seemed to be explained by interests (see Table B4). Responses to "When learning new things, some people tend to feel stressed or tired, while others tend to feel inspired or refreshed. How do you tend to feel when you learn new things?", demonstrated a significantly lower intercept for females (β = -.20, p < .001) that decreased in the final model (β = -.13, p = .11). This effect may have been due to Enterprising interests predicting responses to this item (β = -.06, p = .047). "How often do you ask a teacher or classmate questions that go beyond the material but are still relevant to the topic (either in or out of class)?" also eliminated statistically significant intercept differences across men and women (Initial model: β = -.12, p = .013; Final model: β = -.12, p = .141). However, none of the interest dimensions were related to responses for this item so the change in significance was likely due to the lower statistical power of the refined model because the actual effect size was unchanged. Lastly, "How important has it been in the past for you to be involved in community or volunteer work?" (Social Responsibility, Table B8) was initially found to have

a higher intercept for females (β = .20, p <.001), an effect that was eliminated in the final model (β = .08, p = .144) due to the effects of Realistic (β = -.06, p = .025) and Social interests (β = .08, p = .002). For the most part, the relationship between interests and an item response corresponded to the association between gender and interests. For example, an initially lower intercept for an item among females would ultimately be revealed to be a lower intercept among those with low Realistic interests, which was the case for most females. This pattern signals meaningful explanation of these effects.

Several intercept differences across minority groups also appeared to be related to vocational interests, but the explanatory pattern was less clear than for gender. For the item "To what extent would your friends describe you as someone who goes after what you want?" (Perseverance, Table B5), an intercept difference was found for respondents categorized as Other (β = -.21, p = .03) but this effect was mitigated in the final model (β = -.13, p = .377) after including interests. This was most likely due to Artistic interests predicting the item response (β = -.09, p = .002) given the relatively high Artistic interests among those in the Other race category. All other observed intercept differences based on minority status that appeared to be partially explained using vocational interests were likely not practically meaningful. Some methodological issues should be brought up before these remaining effects are summarized.

The methodological issues that likely yielded several unexpected effects were the changes in the variables that were included in each model and the sample size for each minority group. The model building approach used here involved systematically adding new variables to the model to test the various effects (i.e. SES in the second model and vocational interests in the third), which may have influenced estimates of the variance of the endogenous variables (Muthén, 1989). Though these changes likely influenced some of the effects across gender, the relatively small sample sizes of minority groups likely exacerbated the consequences of adding

variables to the models and decreased the statistical power of these models. As a result, the following effects are likely not practically meaningful but are still discussed.

The aforementioned Perseverance item, "To what extent would your friends describe you as someone who goes after what you want?" (Perseverance, Table B5) yielded a significant intercept difference among Asian respondents (β = -.26, p = .031). This effect was weakened (β = -.23, p = .11) after the inclusion of interests, and may be due to Conventional interests predicting the response to this item (β = -.09, p = .015) given the association between being Asian and Conventional interests. However, the minor change in effect size also suggests that this change in significance may not be practically meaningful. This item was also predicted by Enterprising (β = .10, p = .004) and Investigative interests (β = .06, p = .041), but Asian respondents did not show a meaningful association with either interest domain. Black respondents were predicted to have a lower item intercept for "Over the past year, how many times were you given detention (or a similar punishment)?" (Table B9; β = -.30, $S.E.$ = .11, p = .006) but after the inclusion of interests this effect became nonsignificant (β = -.41, $S.E.$ = .24, p = .095). This change in significance is primarily due to a larger standard error, as the magnitude of the effect itself increased. Conventional interests predicted the item response to this item as well (β = .08, p = .015), but were not higher among Black respondents. The item "How often do you ask a teacher or classmate questions that go beyond the material but are still relevant to the topic (either in or out of class)?" (Table B4) demonstrated intercept differences across Black and White respondents but, again, this effect decreased (Initial model: β = .27, p = .001, Final model: β = .013, p = .926) after including vocational interests in the model. Interestingly, none of the vocational interest dimensions appeared to predict responses to this item and, therefore, could not have explained this reduction. The inclusion of interests also seemed to clarify an intercept difference in "To what extent would your friends describe you as someone who goes after what

you want?" (Table B5; $\beta = .31$, $p = .030$) and "How many times in the past year have you set the schedule (time and/or tasks) for groups in which you have worked?" (Behavioral Leadership. Table B1; $\beta = .23$, $p = .032$) for Black respondents, but these effects may be due to error given the nonsignificant initial effect.

Overall, vocational interests appeared to explain intercept differences across men and women in a predictable way. The role of interests in explaining intercept differences across races was somewhat more tenuous and many of the results for these models were likely influenced by the methodological limitations of the models examined here. These issues will be further discussed in the limitations section below. These findings, alongside the fact that intercept differences were not explained for 23 other biodata items and one SJT item suggests that H9 received moderate support.

DISCUSSION

The goal of the present study was to help extend the examination of how and why bias may occur in biodata and SJTs. By looking at multiple measurement methods, as well as multiple constructs in the case of biodata, this study sought a more general understanding of the explanatory mechanisms underlying measurement bias. At a broad level, this work serves to help address the criticism of bias research that demonstration of bias too often takes precedence over its explanation (Griel, 2005). The primary explanatory factors included in this study were indicators of SES and vocational interests. MIMIC modelling was the primary analytic approach used to incorporate explanatory variables in the assessment of DIF. This analytic approach also allowed for distinguishing DIF from true demographic score differences on the latent trait. A summary of the results for each hypothesis can be found in Table 5. Overall, SES did little to explain either differences in latent scores or the preponderance of DIF. Vocational interests, on the other hand, helped explain both differences in latent scores and DIF but for gender more so than race. Further, vocational interests were consistently related to the constructs measured by biodata scales. This suggests that for both the construction of biodata scales and their interpretation, vocational interests are important to consider. Specifically, vocational interests explained variance in some item responses originally attributed to differences between males and females, suggesting that item content that is highly related to interests may be more likely to exhibit DIF. Additionally, given that some latent differences between males and females were also explained by interests, it is important to be mindful that differences between males and females may reflect differences in interests.

Several score differences between groups were found to be meaningful at the latent level of analysis. MIMIC analyses indicated that females had a moderately higher standing on the

54

latent factors of Perseverance, Values, Discrete Adaptability, Social Responsibility, and Situational Judgment. Vocational interests appeared to explain a meaningful amount of the gender differences for Values and Discrete Adaptability, but a significantly higher standing on Perseverance, Social Responsibility, and Situational Judgment remained after interests were included in the model. Despite a higher standing on Discrete Adaptability, females demonstrated a moderately lower standing on the Routine Adaptability latent factor. For racial minorities, Black respondents had a slightly lower standing on the Leadership Behaviors latent factor and Asian respondents had a much lower standing. Interestingly, despite the conceptual similarity between leadership positions and behaviors, the Positions factor was largely the same across demographic groups. The Knowledge and Discrete Adaptability scales demonstrated a similar pattern to Leadership Behaviors where both Black and Asian test-takers scored either moderately or substantially lower than White test-takers. Social Responsibility analyses demonstrated a moderate effect for Asians such that they had a higher standing on this trait. Across scales, several latent factor differences between groups were found and appeared independent of vocational interests and SES.

At the item level, meaningful patterns of DIF emerged in only two instances. Examining the pattern of DIF is important to assess the extent to which items consistently favor one group over another. The Leadership Behaviors scale contained three items flagged for DIF, two of which demonstrated significantly higher intercepts for Asian participants in the first MIMIC model controlling for demographic variables (the third item was nonsignificant) It is possible that these two items may serve to obscure the lower standing Asian participants have on the latent factor by artificially increasing their observed scores. The other instance was in the case of Discrete Adaptability, where four of five items flagged for DIF demonstrated significantly lower.

Table 5.
Summary of degree of support and relevant results for hypotheses posed in the present study

Hypothesis	Degree of Support	Summary of Relevant Results
H1: The effects of minority status on the standings of the latent traits measured by biodata and SJTs will be partially explained by SES such that minority status will initially predict lower standings on the latent traits measured by biodata and SJTs and this effect will be weakened upon inclusion of measures of SES.	Not supported	Pell grant eligibility and median local income did not explain latent score differences
H2: The effects of minority status on DIF in biodata items will be partially explained by SES such that minority status will initially predict higher item intercepts and this effect will be weakened upon inclusion of SES.	Not supported	Only one item demonstrated the expected effect of SES variables attenuating an inflated intercept difference
H3: The effects of minority status on DIF in SJT items will be partially explained by SES such that minority status will initially predict smaller item factor loadings and this effect will be weakened upon inclusion of SES.	Not evaluated	An acceptable configural model for SJT across race could not be estimated
H4: High social interests should predict higher levels of the latent traits of social responsibility, academic values, and situational judgment	Supported	Social interests predicted all expected noncognitive constructs
H5: High investigative interests should predict higher levels of the latent traits of knowledge, continuous learning, and situational judgment	Supported	Investigative interests predicted all expected noncognitive constructs
H6: High enterprising interests should predict higher levels of the latent traits of leadership, adaptability, and perseverance	Supported	Enterprising interests predicted all expected noncognitive constructs
H7: High conventional interests should predict higher levels of the latent trait of knowledge	Supported	Conventional interests predicted the latent knowledge construct

Table 5 (cont'd)

H8: The effect of minority and gender status on the latent traits assessed by biodata and SJTs will be partially explained by differences in vocational interests such that minority and gender status will predict standing on the latent traits assessed by biodata and SJTs, and this effect will be weakened by the inclusion of vocational interests in the model	Moderately supported	Interests explained a meaningful amount of the difference in Values and Discrete Adaptability factors across gender, but not race
H9: The effect of minority and gender status on DIF in biodata items will be partially explained by vocational interests such that group status will initially predict biodata item intercepts, and this effect will be weakened by the inclusion of vocational interests in the model	Moderately supported	Interests explained uniform DIF effects across gender for 3 Discrete Adaptability and 2 Continuous Learning items. Uniform DIF effects across race did not appear to be meaningfully explained by interests.

item intercepts for females. This pattern of DIF may undermine equitable use of Discrete Adaptability scale scores across gender.

Given that DIF was found for many of the items studied here yet few consistent patterns were observed, future research may seek larger samples with more power to detect DIF effects via MIMIC modelling or use other approaches (e.g. mean and covariance structure analyses, Nye & Drasgow, 2011) to distinguish bias from observed scores.

One of the stronger set of findings in the current work was that vocational interests were related to the constructs measured using biodata and situational judgment assessments. This finding has several potential implications. First, one of the main arguments posed by Oswald et al. (2004) for the utility of biodata and situational judgment assessments was that these assessments predicted incremental variance in academic performance over personality and indicators of cognitive ability. Given that the results here demonstrate that biodata. scales bear some relationship with interests, and that past work shows that interests predict academic

performance (Nye et al., 2012), the extent to which the incremental validity observed by Oswald et al. (2004) is representative of vocational interests should be examined. Second, if interests do direct behaviors and influence the development of procedural knowledge that is assessed by biodata and SJTs, the results found here may imply a more nuanced origin to the constructs assessed. Past work suggests that biodata and situational judgment assessments measure constructs that are the product of past experiences (Mael, 1991; Lievens & Motowidlo, 2016). If vocational interests help determine the experiences an individual chooses to pursue, then it should follow that constructs that are the product of such experiences are also indirectly related to interests. It should be noted that though a causal account is provided to justify the link between vocational interests and the biodata and situational judgment scales here, the evidence provided is quite limited with respect to causality. Future work should examine the relationship between interests and the constructs assessed using biodata and SJTs with a longitudinal design to at least establish temporal precedence between these constructs. Should such evidence be provided, future work may able to further examine the relationship between these constructs, as well as the long-term consequences of vocational interest change.

Though an overall connection between interests and the latent constructs assessed by biodata and SJTs was supported, several unanticipated relationships were observed that should be discussed. Of note, Social and Conventional interests predicted several biodata constructs that were not expected, and it was not anticipated that Realistic and Artistic interests would predict any noncognitive construct. For Social, Conventional, and Artistic interests, some unanticipated relationships make sense theoretically and would make sense to expect in future studies. For example, Social interests predicting Leadership Positions and Behaviors bears similarity in content related to working with other people. A similar argument could be made for those with Conventional interests not engaging in Behavioral Leadership or Routine Adaptability. Further,

58

those with Artistic interests may be quite engaged in pursuing new ideas, as is captured by

Continuous Learning (Holland, 1997; Oswald et al., 2004). Instead of specific relationships

based on content, it may be the case that Realistic interests conflict with qualities that fit well in

an academic context. This may be the case given the broad negative relationships observed

between Realistic interests and several of the noncognitive constructs assessed by biodata and

SJTs. Other relationships, such as Social interests predicting Knowledge or Conventional

interests predicting Discrete Adaptability may need to be further evaluated. Given that these

effects were not hypothesized and do not seem to align based on construct content, it is difficult

to comment on whether such relationships should be expected in future studies. In spite of not

hypothesizing certain relationships, it appears as though the general connection between interests

and the noncognitive constructs assessed here based on construct content still holds.

Vocational interests also played a meaningful role in explaining uniform DIF effects.

Five items were found that exhibited DIF as a function of gender and differences in vocational

interests partially explained these effects. In these instances, an observed uniform DIF effect was

attenuated by vocational interests in a way that corresponded to the relationship between gender

and interest. For example, the incorporation of social interests into the MIMIC model attenuated

intercept differences favoring females because social interests were positively related to both the

item and to being female.

Though interests helped explain uniform DIF effects, the observed pattern does not align

with the frame of reference effect proposed by Robert et al. (2006). The frame of reference effect

suggests that individuals respond to items that ask for a social comparison by perceiving their

own unique comparison group. Thus, comparison group differences may explain item response

differences as well. Specifically, it was thought that an individual may use their demographic

group as a referent and that their responses to biodata and SJT items would differ from their

59

referent group's vocational interests. In other words, individuals may view themselves as particularly high on a trait when their demographic group's standing is low. An individual's demographic group was thought to serve as a comparison group given that individuals tend to be attracted to similar others (e.g. Holland, 1997; Schneider, 1987). However, the results indicate that individual item responses are aligned with their demographic group's standing on vocational interests. This suggests that in instances where DIF was explained, demographic group membership served as a proxy for the individual's vocational interests more so than a description of that individual's perceived comparison group.

Future research examining the frame of reference effect may benefit from two considerations. First, demographic variables may serve as a poor indicator of one's comparison group. Robert et al. (2006) discuss local comparison groups more in terms of culture, which may be a more salient indicator of a comparison group than demographic group membership. Use of psychological variables that describe an individual's comparison group may be more likely to reveal a frame of reference effect than demographics. Second, it may be the case that biodata assessments are constructed in a way that reduces reliance on comparison groups. Though some items rely on social comparison or require evaluating some abstract amount, characteristics thought to produce the frame of reference effect (Robert et al. 2006), these qualities may not influence responding as much as they would in other assessment methods.

Though a clear pattern of statistical results were found for DIF based on gender, evaluation of item content is harder to link to vocational interests and DIF. In some instances this relationship is quite clear. For example, responses to the item "How important has it been in the past for you to be involved in community or volunteer work?" is clearly aligned with Social and Realistic vocational interests as engaging in community or volunteer work likely involves working with others and may also involve working outdoors or with primarily manual tasks.

However, most other items explained by vocational interests were less clearly linked based on content. For example, the item "How often have you failed to meet responsibilities because you had taken on too much?" was explained by Artistic interests, but the relationship between excessive responsibilities and being interested in creative pursuits is more ambiguous. Further, the small effect sizes for many of these effects, like this item in particular, invites the possibility that some of these findings were observed by chance and do not represent theoretically meaningful results. Future efforts may seek to hone in on key experiences that differ across males and females, using vocational interests and differential accessibility (e.g. Imus et al., 2010) as a guide, to help provide more specific suggestions as to how to write biodata items in a way that reduces the risk of DIF.

As was the case with some of the relationships between interests and the constructs assessed here by biodata and SJTs, interests would often explain uniform DIF in a way that was not immediately apparent. Like the example item above, a uniform DIF effect would be found with an item whose content did not seem to relate to the particular vocational interest that was found to be statistically relevant. Some of these unexpected effects may be due to Type I error. Given that many of the standardized uniform DIF effects were small, it is possible that some were observed by chance. However, it may also be the case that such unexpected effects are due to an unaccounted mechanism(s) of DIF. Even though many uniform DIF effects appeared as though demographics were serving as a proxy for interests, other mechanisms of DIF may also be at play. Given the fact that little evidence was found that the frame of reference effect was producing uniform DIF effects (Robert et al. 2006), future research should consider other mechanisms by which demographic DIF may occur in biodata and SJT assessments.

Contrary to expectation, the role of SES in accounting for differences across demographic groups was quite minimal at both the item and the scale levels. The biodata and

SJT measures studied here capture noncognitive attributes that were in part the product of past experiences (Oswald et al., 2004) and these experiences were thought to be shaped by the environment of the individual being assessed. Due to the way SES may shape one's environment and that differences in race coincide with differences in SES (e.g. Cottrell et al., 2015) it was thought that SES may be reflected in the measures assessed here. Only one item demonstrated the expected pattern of effects where SES partially explained DIF. In addition, no latent mean differences across groups were explained by SES. However, this is not to say that SES was not relevant whatsoever. Items assessed for DIF via MIMIC modelling were selected based on exhibiting evidence of DIF across demographic groups. Of the items selected, nine items were predicted by SES variables after accounting for the latent factor. It is possible that other items that did not demonstrate DIF across demographic groups may still demonstrate some form of bias related to SES. Further, both the Continuous Learning and Leadership scales were predicted by SES, though SES was negatively related to a higher standing on Continuous Learning and the effect on Leadership was quite small. Though the effects of SES at the item- and scale-levels appeared minimal, the potentially broad and high-stakes use of noncognitive assessments may warrant further examination of the influence of SES.

Finally, it is important to take stock in both the nature and scope of current explanations of measurement bias, as well as substantive group differences, in biodata assessments and SJTs. Imus et al. (2010) provide good evidence for the role of accessibility in explaining differences in item responses based on gender, but accessibility was less related to response differences based on race. Kim et al. (2004) use socioeconomic reasons to predict DIF in SJTs across racial groups, but the results indicated other major reasons existed for DIF. Cultural values seem to relate to construct differences in biodata and SJT scales across racial groups (Prasad et al., 2017) but do not explain why biodata item response differences arise as a function of race (Whitney &

Schmitt, 1997). The present study adds to this body of research by again evaluating SES as well as introducing the role of vocational interests. At the item level, both vocational interests and perceptions of accessibility seem important to item response differences across males and females, but only limited evidence of the role of SES exists for race differences in survey responding.

Limitations

A limitation of the present research could be the use of median local income as an indicator of SES. Median local income was included as a way to characterize the economic resources an individual may have experienced during high school. However, the MIMIC model, as used here, may not have been able to adequately incorporate the prediction of an individual level outcome using a group level variable (e.g. Kozlowski & Klein, 2000). Future research exploring the impacts of SES should use appropriate statistical methods that can model multilevel relationships or assess individual-level perceptions of environmental variables to examine these relationships.

In spite of efforts to ensure otherwise, sample size limitations may have had a number of negative effects on the present study. First, adequate modelling of the SJT was hindered by the small sample sizes for some minority groups. Specifically, when estimating a configural model for the SJT, unique item factor loadings and intercepts as well as latent means and variances would constitute 52 unique estimated parameters per group. For the Asian and Black groups, this would have resulted in roughly two to three participants per estimated parameter. Though the fit indices used here should be relatively robust to samples of this size, having so few participants per estimated parameter may increase the error in parameter estimates that are used for comparisons across groups. The sample sizes and the number of estimated parameters was also a concern for the MIMIC model and its ability to detect uniform and nonuniform DIF. Though the

63

subgroup sample sizes used here met or exceeded the recommendations by Woods and Grimm (2011), the models estimated were significantly more complex than those simulated in their research. First, Woods and Grimm (2011) only examined varying conditions with a single focal and referent group whereas the present research employs four focal groups. Further, the incorporation of eight explanatory variables may have also constituted a meaningful increase in model complexity.

In terms of DIF detection, there were a few instances where there were discrepancies between the DIF items identified by mean and covariance structure (MACS) analyses and MIMIC modelling (Woods & Grimm, 2011). This happened more frequently for DIF based on race, where an item would be flagged for DIF using MACS analyses but then no significant effect would be found within the MIMIC model. It may be the case that the free baseline approach is more strongly affected by the small sample sizes in some groups than a MIMIC model, but it may also be the case that the significant differences detected by MACS analyses reflected differences between two focal groups. Using the MIMIC model, DIF effects are detected when a particular focal group is significantly different than the referent, but not necessarily when focal groups are different from each other. Using the present analyses it is difficult to compare the relative likelihood of either explanation, but future research may wish to keep such considerations in mind.

Finally, in some instances, the effects varied between models in unexpected ways. For example, an initially significant intercept difference for Black participants increased in size between the first and third MIMIC models but was ultimately nonsignificant. This coincided with a relatively large increase in the estimated standard error for that effect. Across all models, there were several other instances where a particular DIF effect would fluctuate in terms of statistical significance across models even though the actual size of the effect did not change

substantially. This may be due to the prescription by Muthén (1989) that explanatory variables should be exogenous combined with the model building approach used here. Specifically, if the estimated variances of endogenous variables are conditioned on the exogenous variables in the model, then it is likely that endogenous variable variances as well as the standard errors of effects involving endogenous variables would fluctuate based on the changing set of exogenous variables. Though this did not seem to meaningfully influence the core results of this study, future research on this topic will need to address this methodological limitation.

Practical Implications

The present research demonstrated that latent score differences may be large enough to cause concern of measurement invariance when using these assessments across demographic groups. The implication of this is that in spite of the many benefits biodata and SJT assessments hold (e.g. Ployhart, 2006; Mitchell et al., 2001), additional measurement equivalence research may be advisable before expanding the use of these measures across different demographic groups. Given some of the moderate to large latent differences between demographic groups, there may be meaningful differences in the experiences individuals from different groups have beyond item level idiosyncrasies. These differences should be further understood to promote the fair and effective use of noncognitive assessments in selection procedures. Why these behaviors differ based on race and whether other behaviors could be assessed that are construct relevant are important questions to ask to further refine the use of biodata and situational judgment assessments.

Beyond a deeper understanding of biodata and SJT assessment methods, a failure to take measurement invariance into account may negatively impact the accuracy of selection procedures. Nye and Drasgow (2011) argue that the presence of DIF may artificially influence how individuals are rank ordered based on a selection instrument. Inaccuracy in rank ordering of

candidates due to bias is problematic not only for accurately and fairly hiring individuals but also because it can increase the risk of adverse impact. A failure to account for bias as a function of group membership can cause individuals from a particular demographic group to be disproportionately selected over other groups. Further, Nye and Drasgow (2011) found that the risk of adverse impact due to bias increases as higher cut scores are used. Given the size of latent score differences (particularly those that favor White males) and the number of items flagged for DIF, the results of this study indicate that bias does play a meaningful role in interpreting biodata and SJT score differences across groups. Combined with the findings of Nye and Drasgow (2011), this suggests that the risk of adverse impact may also increase as practitioners turn to noncognitive assessments to complement cognitive ability testing for highly competitive positions.

It is also important to highlight the practical implications of possible compensatory DIF (e.g. Raju, Van der Linden, & Fleer, 1995). For both the leadership positions and continuous learning scales, DIF was identified when scale level demographic differences were modest. Even in instances where significant demographic differences in scale scores were observed, they were often quite modest even while DIF was present. It is possible that many of the DIF effects observed here were either small overall or did not systematically favor one group over another. Although items within a scale may show evidence of DIF, DIF in opposite directions can cancel out at the scale level and result in scale scores that are not biased across groups (Raju, Van der Linden, & Fleer, 1995). This presents a dilemma regarding the practical use of these scales. Like past studies (e.g. Schmitt et al., 2009), observed score differences between groups were relatively minor suggesting low risk of adverse impact when used in a selection context. However, if these scales were being used operationally and items that exhibited DIF were removed, the scale level

similarity between groups may be changed as well. Thus, the practical note is to evaluate whether item removal due to DIF truly improves the equity of scale scores.

Beyond concerns about the comparability of scores, some practical guidance can be gained from this research. Following the call for explanatory mechanisms of DIF by Griel (2005), this research shows that biodata items that tap into experiences but that also relate to vocational interest domains may be likely to function differently across males and females. Though items flagged for DIF may still warrant removal whether the mechanism for DIF is understood or not, it may behoove test creators to consider the content of experiences covered by items and whether or not those experiences may differ based on gender. Further, it does not appear as though socioeconomic differences dramatically influenced DIF. As a result, test makers may not be at risk when incorporating content that may vary as a function of SES (e.g. relationships with teachers and other school related activities). Though these suggestions are relatively intuitive, such guidance may be helpful given the flexibility of constructing biodata assessments and the onus placed on test makers when creating new assessments for different constructs and contexts.

Beyond the implications for assessment, the present research also has implications for prospective college students. High social, investigative, and enterprising interests seemed to be positively related to constructs assessed in the biodata measure used here, whereas realistic interests yielded negative relationships for the most part. This suggests that individuals who have high social, investigative, and enterprising interests, as well as low realistic interests, may be more likely to engage in the experiences that yield the noncognitive qualities to do well in college (Oswald et al., 2004). Though this interest signature, so to speak, may help shape efforts to help prospective college students, what exactly those efforts should be is still a broad question. The stability of interests during high school years (Low et al., 2005) suggests that individuals

with certain interests may not naturally direct themselves towards experiences that may prepare them for college. This would imply that these individuals may benefit from external influence, such as being directed towards more volunteer opportunities or incorporating more class activities with the need to independently explore a topic. However, with the aid of future research, changing a high school student's interests may be an option to consider for more self-directed engagement in experiences that promote academic success. Low et al. (2005) found that reported interest levels increased during college years and Morris (2016) found that interest differences between males and females were larger among younger participants than older participants. These findings suggest that interest change can occur and possibly in a way that reduces demographic differences. Future work in vocational interest development may reveal strategies that lead to high school students being self-motivated to pursue experiences that may promote academic success.

Conclusion

Differences in the use of biodata and SJT assessments were evaluated between both males and females and across minority groups. The present research found that vocational interests were important to an individual's overall standing on the noncognitive attributes assessed here. Further, interests may also help explain why some latent score and item response differences exist across males and females. Further work is needed to identify additional mechanisms for DIF as many items still exhibited DIF as a function of gender when accounting for interests. Additionally, vocational interests do not seem to play a role in explaining item functioning or latent score differences across racial groups. SES was also evaluated as an explanation of DIF and latent score differences across groups, but for the most part such effects were not observed. Overall, the present work constitutes a thorough examination of differential

functioning in noncognitive assessments and establishes a meaningful relationship between the

noncognitive constructs assessed here and vocational interests.

APPENDICES

APPENDIX A:

Configural model estimation and DIF analyses for studied scales

Table A1.
Configural model estimation and DIF analyses for the Behavioral Leadership scale

Item Content	Item Responses	Fit Statistic	Gender	Race
*During the past year, how often have you taken charge of a group that you were in, without being asked?	Never Once or twice Between three and five times Between six and ten times More than ten times	χ^2 (df) RMSEA CFI NNFI SRMR	123.75 (16) .095 .957 .919 .037	127.04 (32) .090 .962 .928 .037
(Reversed) How often in the past year have you volunteered to be the spokesperson for a group project you did at school or work?	Much more often than most people Somewhat more often than most people About as often as most people Somewhat less often than most people A good bit less often than most people	χ^2 (df) RMSEA CFI NNFI SRMR	124.37 (18) .089 .957 .929 .037	132.07 (38) .082 .962 .940 .039
In the past year, how many times have you been responsible for assigning tasks and setting deadlines for other people?	Never Once Twice Three or four times Five times or more	χ^2 (df) RMSEA CFI NNFI SRMR	123.76 (18) .089 .958 .929 .037	**151.64 (38)** **.090** **.954** **.928** **.045**
How many times in the past year have you set the schedule (time and/or tasks) for groups in which you have worked?	Never Once Twice Three or four times Five times or more	χ^2 (df) RMSEA CFI NNFI SRMR	**134.21 (18)** **.093** **.954** **.923** **.041**	141.05 (38) .085 .959 .935 .044

Table A1 (cont'd)

(Reversed) When asked to do a class project with other students, how often do you take the lead and assign tasks or roles to people in the group?	I am usually the one who assigns tasks or roles to get the work done More than half the time I end up assigning the tasks and roles About half the time I take the lead in assigning tasks and roles I rarely take the lead in assigning tasks and roles I never take the lead unless I have been assigned to do so	χ^2 (df) *RMSEA* *CFI* *NNFI* *SRMR*	130.08 (18) .092 .955 .925 .039	**140.45 (38)** **.085** **.959** **.935** **.045**
When you are in a meeting for a project or activity, how do you tend to be?	You are quiet You follow others You are sometimes more of a leader and sometimes more of a follower You lead others	χ^2 (df) *RMSEA* *CFI* *NNFI* *SRMR*	126.30 (18) .090 .957 .928 .041	138.23 (38) .084 .960 .936 .043

Note. * denotes item identified as referent item. Fit statistics displayed by referent item are for the estimation of the configural model. Gender and Race denote multiple groups models comparing gender and race demographic groups, respectively. Fit statistics presented in bold denote indication of DIF as determined by a CFI decrease of > .002. (Reversed) denotes that lower item responses relate to a higher standing on the target scale, and that item responses were reversed prior to analyses.

Table A2.

Configural model estimation and DIF analyses for the Leadership Positions scale

Item Content	Item Responses	Fit Statistic	Gender	Race
*How many times in the past year have you tried to get someone to join an activity in which you were involved or leading	Never Once Twice Three or four times Five times or more	χ^2 *(df)* RMSEA CFI NNFI SRMR	5.52 (4) .023 .999 .997 .009	15.41 (8) .050 .996 .987 .015
The number of high school clubs and organized activities (such as band, sports, newspapers, etc.) in which you took a leadership role was:	I did not take a leadership role 1 2 3 4 or more	χ^2 *(df)* RMSEA CFI NNFI SRMR	**15.94 (6)** **.047** **.994** **.988** **.021**	22.42 (14) .040 .995 .992 .031
During the last two years, how many leadership positions were you offered (even if you didn't take them)?	None One Two or three Four or five More than five	χ^2 *(df)* RMSEA CFI NNFI SRMR	6.42 (6) .010 1.000 1.000 .014	17.77 (14) .027 .998 .996 .018
In the past year, how often have you been selected by a group or club to serve as an official or representative?	Never Once Twice Three or four times Five times or more	χ^2 *(df)* RMSEA CFI NNFI SRMR	11.05 (6) .034 .997 .994 .012	22.74 (14) .041 .995 .991 .029

Note. * denotes item identified as referent item. Fit statistics displayed by referent item are for the estimation of the configural model. Gender and Race denote multiple groups models comparing gender and race demographic groups, respectively. Fit statistics presented in bold denote indication of DIF as determined by a CFI decrease of > .002.

Table A3.
Configural model estimation and DIF analyses for the Knowledge scale

Item Content	Item Responses	Fit Statistic	Gender	Race
*How often have you studied for tests by trying to memorize just the basic facts and not much more?	Very often Often Sometimes Rarely Never	χ^2 (df) RMSEA CFI NNFI SRMR	179.63 (54) .050 .919 .892 .038	234.53 (108) .056 .919 .892 .044
For classwork, how often do you tend to skim the material, reading only the important points?	Almost all the time Most of the time Sometimes Rarely Never	χ^2 (df) RMSEA CFI NNFI SRMR	**192.71 (56)** **.057** **.912** **.887** **.039**	240.49 (114) .055 .919 .898 .045
(Reversed) In general, what is the lowest grade that you find acceptable for yourself?	A or equivalent B or equivalent C or equivalent D or equivalent F or equivalent	χ^2 (df) RMSEA CFI NNFI SRMR	184.54 (56) .056 .917 .894 .042	**254.70 (114)** **.058** **.910** **.887** **.050**
(Reversed) How often do you spend extra time on school assignments, even after they are turned in, so that you can gain a better understanding of the material or principles?	Very often Often Sometimes Rarely Never	χ^2 (df) RMSEA CFI NNFI SRMR	185.94 (56) .056 .917 .893 .039	**252.04 (114)** **.057** **.912** **.889** **.046**
Generally, whenever you learn about a topic or how to perform a task, how often do you learn all the details as well as the general principles?	Hardly ever Not very often Sometimes Often Almost always	χ^2 (df) RMSEA CFI NNFI SRMR	179.66 (56) .055 .921 .898 .038	**261.75 (114)** **.059** **.906** **.881** **.047**

Table A3 (cont'd)

(Reversed) When you took classes that you thought were easy, how important was it for you still to understand the concepts underlying the class material?	Extremely important Very important Rather important Sort of important Not important	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	182.00 (56) .055 .919 .896 .038	**253.09 (114)** **.057** **.911** **.888** **.046**
(Reversed) In your last year of high school, on how many tests did you "settle" for a passing grade, rather than spend significant amounts of time learning material well?	Never Once Twice Three or four times Five times or more	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	182.12 (56) .055 .919 .896 .039	239.1 (114) .054 .920 .899 .044
A year after completing a class, how much can you typically remember about what you were taught?	I tend to forget most of what was taught in class I remember the general ideas that were taught in class I remember some of the details that were taught in class I remember a lot of the details that were taught in class	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	**196.27 (56)** **.058** **.910** **.884** **.039**	238.36 (114) .054 .921 .900 .045

Table A3 (cont'd)

How do you compare your standards for learning to those of your high school teachers?	Much lower than my teachers' standards Lower than my teachers' standards About the same than my teachers' standards Higher than my teachers' standards Much higher than my teachers' standards	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	182.11 (56) .055 .919 .896 .039	**245.50 (114)** **.056** **.916** **.894** **.046**

Note. * denotes item identified as referent item. Fit statistics displayed by referent item are for the estimation of the configural model. Gender and Race denote multiple groups models comparing gender and race demographic groups, respectively. Fit statistics presented in bold denote indication of DIF as determined by a CFI decrease of > .002. (Reversed) denotes that lower item responses relate to a higher standing on the target scale, and that item responses were reversed prior to analyses.

Table A4.
Configural model estimation and DIF analyses for the Continuous Learning scale

Item Content	Item Responses	Fit Statistic	Gender	Race
*(Reversed) When it is not required to do so, how often do you read materials (e.g. books, magazines, web sites) that pertain to subjects that you are learning about in class?	Very often Often Sometimes Rarely Never	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	333.72 (70) .071 .923 .901 .040	436.06 (140) .075 .915 .891 .045
(Reversed) When curious about a particular subject, how likely were you to go out and research the subject on your own?	Extremely likely Very likely Rather likely Sort of likely Not likely	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	334.15 (72) .070 .924 .904 .040	445.47 (146) .074 .914 .894 .050
In the past month, how many times have you looked for more information about something that you found interesting?	Never Once or twice 3 to 5 times 6 to 10 times More than 10 times	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	**385.12 (72)** **.077** **.909** **.886** **.047**	429.21 (146) .074 .916 .897 .045
(Reversed) How often do you ask a teacher or classmate questions that go beyond the material but are still relevant to the topic (either in or out of class)?	Very often Often Sometimes Rarely Never	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	336.99 (72) .070 .923 .903 .041	**456.42 (146)** **.076** **.911** **.891** **.047**
In the past 6 months, how many times have you been so absorbed when learning something that you didn't realize how much time passed?	Almost Never Once Twice Three or four times Five times or more	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	334.15 (72) .070 .924 .904 .040	440.22 (146) .074 .916 .896 .045

Table A4 (cont'd)

In the past month, how many times did you go out and learn more about something simply because it seemed interesting?	Never Once Twice Three or four times Five times or more	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	**361.32 (72)** **.074** **.916** **.895** **.044**	443.76 (146) .074 .915 .895 .046
(Reversed) When a textbook or instructor mentions another source of information about a topic, how likely are you to find it and learn more on your own?	Extremely Likely Very Likely Somewhat Likely Not very likely Not at all likely	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	335.65 (72) .070 .923 .904 .041	446.26 (146) .074 .914 .894 .046
(Reversed) How likely were you to take a class or find an instructor so that you could learn more about a hobby or skill?	Much more likely than most people Somewhat more likely than most people About as likely as others Somewhat less likely than most people Much less likely than most people	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	334.00 (72) .070 .924 .905 .041	446.79 (146) .074 .914 .894 .047
(Reversed) How often have you become involved in something just for the sake of learning?	Very often Often Sometimes Rarely Never	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	333.88 (72) .070 .924 .905 .041	438.06 (146) .073 .916 .897 .045

Table A4 (cont'd)

When learning new things, some people tend to feel stressed or tired, while others tend to feel inspired or refreshed. How do you tend to feel when you learn new things?	Very stressed/tired Somewhat stressed/tired Something in between stressed/tired and inspired/refreshed Somewhat inspired/refreshed Very inspired/refreshed			
		χ^2 *(df)*	**347.03 (72)**	444.15 (146)
		RMSEA	**.072**	.074
		CFI	**.920**	.915
		NNFI	**.900**	.895
		SRMR	**.041**	.047

Note. * denotes item identified as referent item. Fit statistics displayed by referent item are for the estimation of the configural model. Gender and Race denote multiple groups models comparing gender and race demographic groups, respectively. Fit statistics presented in bold denote indication of DIF as determined by a CFI decrease of > .002. (Reversed) denotes that lower item responses relate to a higher standing on the target scale, and that item responses were reversed prior to analyses.

Table A5.
Configural model estimation and DIF analyses for the Values scale

Item Content	Item Responses	Fit Statistic	Gender	Race
*(Reversed) In high school, how many times have you cheated on a school project, assignment, or test?	0 1 2 or 3 4 to 10 More than 10	χ² (df) RMSEA CFI NNFI SRMR	186.27 (68) .048 .931 .908 .038	305.23 (136) .058 .907 .877 .048
(Reversed) In the past, how likely were you to return money that you received by accident (e.g., received extra change after buying something)?	Much more likely than most people Somewhat more likely than most people About as likely as others Somewhat less likely than most people Much less likely than most people	χ² (df) RMSEA CFI NNFI SRMR	186.68 (70) .047 .932 .912 .039	311.38 (142) .057 .907 .882 .050
During high school, how many times have you expressed disapproval or anger at a friend for behaving in a manner that you considered to be unethical or wrong?	Never Once Twice Three or four times Five times or more	χ² (df) RMSEA CFI NNFI SRMR	**194.72 (70)** **.049** **.927** **.906** **.040**	**316.03 (142)** **.057** **.904** **.879** **.050**
(Reversed) In the past year, how many times have you copied someone else's work and submitted it as your own (at school or at work)?	Never Once Twice Three or four times More than five times	χ² (df) RMSEA CFI NNFI SRMR	190.24 (70) .048 .930 .909 .039	309.42 (142) .056 .908 .883 .050
(Reversed) When you have found someone else's belongings, how often have you attempted to return them?	Always Most of the time Half of the time Less than half of the time I have never found someone's belongings	χ² (df) RMSEA CFI NNFI SRMR	187.50 (70) .048 .931 .911 .039	313.22 (142) .057 .906 .881 .051

81

Table A5 (cont'd)

(Reversed) Over the past year, how many times were you given detention (or a similar punishment)?	Never	χ^2 (df)	**201.54 (70)**	**316.72 (142)**
	Once	RMSEA	**.050**	**.058**
	Twice	CFI	**.923**	**.904**
	Three or four times	NNFI	**.901**	**.878**
	Five times or more	SRMR	**.044**	**.052**
In your first three years of high school, how often did you skip classes without a legitimate reason?	Most of the time	χ^2 (df)	**196.47 (70)**	316.93 (142)
	A lot	RMSEA	**.049**	.058
	Sometimes	CFI	**.926**	.904
	Once or twice	NNFI	**.905**	.878
	Never	SRMR	**.043**	.053
If a fellow student offered you a copy of upcoming exam questions that he had retrieved from the teacher's recycling bin, how likely would you be to accept a copy?	Extremely likely	χ^2 (df)	187.45 (70)	315.29 (142)
	Quite likely	RMSEA	.048	.057
	Somewhat unlikely	CFI	.931	.905
	Not at all likely	NNFI	.912	.879
		SRMR	.039	.052
If you were struggling with a school assignment, and a fellow student with more expertise offered to finish it for you, how likely is it that you would accept the offer?	Extremely likely	χ^2 (df)	190.48 (70)	314.84 (142)
	Quite likely	RMSEA	.048	.057
	Somewhat likely	CFI	.929	.905
	Not at all likely	NNFI	.909	.879
		SRMR	.039	.052
How many times have you been accused of acting unethically?	Very often	χ^2 (df)	**208.60 (70)**	**323.11 (142)**
	Often	RMSEA	**.052**	**.059**
	Sometimes	CFI	**.919**	**.900**
	Rarely	NNFI	**.896**	**.874**
	Never	SRMR	**.044**	**.058**

Note. * denotes item identified as referent item. Fit statistics displayed by referent item are for the estimation of the configural model. Gender and Race denote multiple groups models comparing gender and race demographic groups, respectively. Fit statistics presented in bold denote indication of DIF as determined by a CFI decrease of > .002. (Reversed) denotes that lower item responses relate to a higher standing on the target scale, and that item responses were reversed prior to analyses.

Table A6.
Configural model estimation and DIF analyses for the Social Responsibility scale

Item Content	Item Responses	Fit Statistic	Gender	Race
*How many hours of volunteer work did you do while in high school?	0 Between 1 and 10 Between 11 and 30 Between 31 and 75 More than 75	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	248.78 (52) .071 .944 .922 .043	314.6 (104) .074 .943 .922 .049
How many times in the past year have you volunteered in social service or charity organizations?	Never Once Twice Three Four times or more	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	251.4 (54) .070 .943 .925 .044	317.98 (110) .071 .944 .927 .049
During the past two years, how many times did you work with not-for-profit groups?	0 1 2 3 or 4 5 or more	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	250.91 (54) .070 .944 .925 .044	323.08 (110) .072 .943 .925 .050
During the last year, how many times have you given money, food, or clothes to a charity or a poor person in need?	0 1 2 3 More than 3	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	251.22 (54) .070 .943 .925 .044	**333.06 (110)** **.074** **.940** **.922** **.055**
In the past year, how many hours were you engaged in community service or volunteer activities?	None Less than 10 hours 11 - 40 hours 41 - 80 hours More than 80 hours	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	251.8 (54) .070 .943 .924 .044	**336.09 (110)** **.074** **.939** **.921** **.051**
(Reversed) How important has it been in the past for you to be involved in community or volunteer work?	Extremely important Very important Important Not very important Not at all important	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	**267.71 (54)** **.073** **.939** **.918** **.044**	326.44 (110) .073 .942 .924 .053

Table A6 (cont'd)

(Reversed) In the past, how likely were you to stop to help a stranger in need (e.g., giving directions to a lost person)?	Extremely Likely Very Likely Somewhat Likely Not very likely Not at all likely	χ^2 (df) RMSEA CFI NNFI SRMR	248.84 (54) .070 .944 .926 .043	324.78 (110) .072 .942 .925 .051
In the past year, in how many fundraisers have you participated?	None 1 2 3 4 or more	χ^2 (df) RMSEA CFI NNFI SRMR	254.89 (54) .071 .942 .923 .045	**334.26 (110)** **.074** **.940** **.921** **.055**
During the past year, how often have you recycled?	Never Not very often Sometimes Often Always	χ^2 (df) RMSEA CFI NNFI SRMR	250.9 (54) .070 .944 .925 .044	**376.5 (110)** **.081** **.928** **.906** **.054**

Note. * denotes item identified as referent item. Fit statistics displayed by referent item are for the estimation of the configural model. Gender and Race denote multiple groups models comparing gender and race demographic groups, respectively. Fit statistics presented in bold denote indication of DIF as determined by a CFI decrease of > .002. (Reversed) denotes that lower item responses relate to a higher standing on the target scale, and that item responses were reversed prior to analyses.

Table A7.
Configural model estimation and DIF analyses for the Perseverance scale

Item Content	Item Responses	Fit Statistic	Gender	Race
(Reversed) How important is it to you to succeed in whatever task you are engaged in?	Extremely important	χ^2 *(df)*	270.65 (54)	323.8 (108)
	Very important	RMSEA	.073	.073
	Important	CFI	.889	.891
	Not very important	NNFI	.852	.855
	Not at all important	SRMR	.044	.048
To what extent would your friends describe you as someone who goes after what you want?	Not at all	χ^2 *(df)*	274.07 (56)	**335.13 (114)**
	A slight extent	RMSEA	.072	**.072**
	A moderate extent	CFI	.888	**.888**
	A large extent	NNFI	.856	**.859**
	A great extent	SRMR	.045	**.055**
How frequently do you fail to get what you want because you did not put in enough effort?	Very often	χ^2 *(df)*	**277.88 (56)**	**373.04 (114)**
	Often	RMSEA	**.073**	**.078**
	Sometimes	CFI	**.886**	**.869**
	Rarely	NNFI	**.854**	**.835**
	Never	SRMR	**.049**	**.053**
(Reversed) To what extent has it been important to you to do your very best whenever you take on a project?	Extremely important	χ^2 *(df)*	276.89 (56)	329.62 (114)
	Very important	RMSEA	.073	.071
	Important	CFI	.887	.891
	Not very important	NNFI	.854	.862
	Not at all important	SRMR	.046	.050
(Reversed) How often have you accomplished something you initially thought was very difficult or almost impossible?	Very often	χ^2 *(df)*	**277.95 (56)**	331.98 (114)
	Often	RMSEA	**.073**	.072
	Sometimes	CFI	**.886**	.890
	Rarely	NNFI	**.854**	.861
	Never	SRMR	**.045**	.050
(Reversed) How often have you finished a project when faced with difficult circumstances?	Very often	χ^2 *(df)*	272.37 (56)	**336.77 (114)**
	Often	RMSEA	.072	**.073**
	Sometimes	CFI	.889	**.888**
	Rarely	NNFI	.857	**.858**
	Never	SRMR	.044	**.052**

Table A7 (cont'd)

(Reversed) How often do others tend to compliment you on your determination to continue with a project under difficult circumstances?	Very often	χ^2 *(df)*	**283.94 (56)**	**339.48 (114)**
	Often	*RMSEA*	**.074**	**.073**
	Sometimes	*CFI*	**.883**	**.886**
	Rarely	*NNFI*	**.850**	**.856**
	Never	*SRMR*	**.045**	**.054**
How often do you tend to give up on a task after being told that you were not doing well?	Almost all the time	χ^2 *(df)*	**286.15 (56)**	333.77 (114)
	Most of the time	*RMSEA*	**.074**	.072
	Sometimes	*CFI*	**.882**	.889
	Rarely	*NNFI*	**.848**	.860
	Never	*SRMR*	**.054**	.051
When encountering problems that take a long time to solve, how impatient do you tend to become?	Extremely impatient	χ^2 *(df)*	**289.19 (56)**	325.02 (114)
	Very impatient	*RMSEA*	**.075**	.071
	Somewhat impatient	*CFI*	**.881**	.893
	Slightly impatient	*NNFI*	**.846**	.865
	Not at all impatient	*SRMR*	**.050**	.049

Note. * denotes item identified as referent item. Fit statistics displayed by referent item are for the estimation of the configural model. Gender and Race denote multiple groups models comparing gender and race demographic groups, respectively. Fit statistics presented in bold denote indication of DIF as determined by a CFI decrease of > .002. (Reversed) denotes that lower item responses relate to a higher standing on the target scale, and that item responses were reversed prior to analyses.

Table A8.
Configural model estimation and DIF analyses for the Discrete Adaptability scale

Item Content	Item Responses	Fit Statistic	Gender	Race
(Reversed) How effective would others say you are at handling multiple projects simultaneously?	Much more effective than most people Somewhat more effective than most people About as effective as most people Somewhat less effective than most people Much less effective than most people	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	89.99 (18) .073 .924 .873 .037	115.44 (36) .077 .914 .856 .042
How often have you failed to meet responsibilities because you had taken on too much?	Very often Often Sometimes Rarely Never	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	**96.36 (20)** **.072** **.919** **.879** **.041**	121.88 (42) .072 .913 .876 .047
(Reversed) How difficult has it been for you to continue with something after being interrupted and having to take care of something else?	Very easy Easy Not easy but not difficult Difficult Very difficult	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	**96.64 (20)** **.072** **.919** **.878** **.037**	119.51 (42) .070 .916 .880 .046
(Reversed) How often do you plan ahead and make a specific schedule of things you need or want to do?	Very often Often Sometimes Rarely Never	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	**130.46 (20)** **.086** **.883** **.824** **.052**	118.37 (42) .070 .917 .882 .043
In the past, how difficult has it been for you to change your study habits to improve on a skill or to do better in a class	Very difficult Difficult Not easy but not difficult Easy Very easy	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	**94.27 (20)** **.071** **.921** **.882** **.038**	**124.38 (42)** **.073** **.911** **.872** **.048**

Table A8 (cont'd)

When you are working on a serious and relatively difficult task and something or someone interrupts you, how do you usually react?	With a great deal of annoyance - it is hard to get back to the original task You are irritated - it's hard to stay on task when you are interrupted It bothers you just a little - you'd really prefer not to be interrupted It doesn't bother you - you feel one of the challenges of any job is the ability to "juggle" several things at a time			
		χ^2 (df)	**104.31 (20)**	119.84 (42)
		RMSEA	**.075**	.071
		CFI	**.911**	.916
		NNFI	**.866**	.879
		SRMR	**.040**	.043

Note. * denotes item identified as referent item. Fit statistics displayed by referent item are for the estimation of the configural model. Gender and Race denote multiple groups models comparing gender and race demographic groups, respectively. Fit statistics presented in bold denote indication of DIF as determined by a CFI decrease of > .002. (Reversed) denotes that lower item responses relate to a higher standing on the target scale, and that item responses were reversed prior to analyses.

Table A9.
Configural model estimation and DIF analyses for the Routine Adaptability scale

Item Content	Item Responses	Fit Statistic	Gender
*Compared with others, how long does it take you to feel comfortable in new situations or places?	A very long time A long time Neither a short nor a long time A short time A very short time	χ^2 *(df)* RMSEA CFI NNFI SRMR	4.083 (4) .005 1.000 1.000 .009
In the past, how difficult have you found it to adjust to major changes in your life (e.g., moving, a new school, a new job)?	Extremely difficult Very difficult Difficult Not very difficult Not at all difficult	χ^2 *(df)* RMSEA CFI NNFI SRMR	**9.47 (6)** **.028** **.996** **.992** **.016**
How difficult has it been for you to deal with situations that forced you to make adjustments in your daily life (e.g., a broken leg, illness, or family crisis)?	Very difficult Difficult Not easy but not difficult Easy Very Easy	χ^2 *(df)* RMSEA CFI NNFI SRMR	8.01 (6) .021 .998 .996 .015
To what extent have you been bothered by sudden changes in your schedule?	To a great extent To a large extent To a moderate extent To a slight extent Not at all	χ^2 *(df)* RMSEA CFI NNFI SRMR	4.19 (6) .000 1.000 1.000 .010

Note. * denotes item identified as referent item. Fit statistics displayed by referent item are for the estimation of the configural model. Gender and Race denote multiple groups models comparing gender and race demographic groups, respectively. Fit statistics presented in bold denote indication of DIF as determined by a CFI decrease of > .002. (Reversed) denotes that lower item responses relate to a higher standing on the target scale, and that item responses were reversed prior to analyses.

Table A10.

Configural model estimation and DIF analyses for the situational judgment scale

Item Content	Item Responses	Fit Statistic	Gender
*You have been standing in line for the restroom for some time after a campus event, and someone cuts into the line ahead of you. What would you do?	Politely inform the person that there is a line and hopefully he/she will move to the back. (Best) Say aloud to someone near you how rude it is that people cut in line. Give them dirty looks, and try to squeeze them out of line. Scold the person for not respecting other people. (Worst) Be annoyed but not do anything. It's just one more person. Calmly cut back in front of them. (Best)	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	691.43 (500) .023 .881 .861 .035
You are part of a three-person group working on a class project with a quickly approaching deadline. One member of the team is not pulling his weight. He avoids assignments, complains about the amount of work that has to be done, and says the project doesn't really matter anyway. While you are all classmates, you seem to be the group leader. What would you do?	Divide the workload evenly among members of the group, making sure everyone knows they are responsible for their share. If the group member still does not pull his own weight, bring it up with the instructor. (Best) Speak with him in private and offer him moral encouragement to complete his portion of the project. If the group member still does not pull his own weight, bring it up with the instructor. Try to get the team member motivated to do his work. If that doesn't help the situation, just put more effort into the project yourself in order to complete it. Just do the group member's portion of the assignment in addition to your own, and tell the instructor about the situation. (Worst) See if the person could be removed from your group. Consult with the non-problematic group member about the most appropriate course of action, and then act on whatever you jointly decide.	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	693.63 (502) .023 .881 .869 .035

Table A10 (cont'd)

A fellow student allows you to listen to threatening phone messages that have been placed on the person's voicemail by another student. The student does not want you to tell anyone, but thinks the caller may be capable of causing physical harm. What would you do?	Try to talk them into calling the police and warn them not to walk around alone. (Best) Talk to the resident assistant about it. Contact the police yourself if you think there is any real threat of physical harm. Find out who is making the calls, if it is another student, confront them – singly or jointly. (Worst) Unless the friend knows something that they're not saying, there is no reason not to call the police – so call them if your friend won't. Have the friend change their phone number.	χ^2 (df) RMSEA CFI NNFI SRMR	**703.34 (502)** **.023** **.875** **.862** **.036**
As a leader of a student organization, you asked a committee member to track the use of important and costly supplies. In response, she developed forms requiring the organization's committee members to indicate when and how they used various supplies. The coordinating individual now complains that no committee members are complying with her request for information on the use of supplies. How would you handle this situation?	Explain the importance of tracking to the committee, and request that everyone comply with the request. (Best) Ask everyone to respect the coordinating individual's hard work and effort by cooperating. Limit access to the supplies until people start filling out the forms, or have penalties for not complying. Designate someone else to be in charge of tracking and enforcing the information requests. (Worst) Ask the committee if there is a misunderstanding about the forms and for suggestions on improving them.	χ^2 (df) RMSEA CFI NNFI SRMR	693.91 (502) .023 .881 .869 .035

Table A10 (cont'd)

Your roommate, usually a tidy person, has recently experienced some personal difficulties. As a result, he/she has become quite distracted and has left much of the household responsibilities to you. You have talked to him/her about your concerns, and empathetically requested that he/she resume his/her share of the responsibilities as soon as possible. A month passes and you are still doing too much of his/her work. What would you do?	Find out more about his/her problem and try to deal with that first. (Best)	χ^2 *(df)*	694.88 (502)
	Stop doing all of the household responsibilities to show him/her what it's like.	*RMSEA*	.023
		CFI	.880
	Talk with him/her again and explain that you are suffering as a result of his behavior. (Best)	*NNFI*	.868
		SRMR	.035
	Tell him/her that if he/she doesn't help, you will move out. (Worst)		
	Do your share of the work, and put anything of his/hers that affects you in his/her area of the room.		

Table A10 (cont'd)

After you arrive on campus, you begin to socialize with a group of students who drink regularly even though all are underage. By the end of the term, you realize that you are drinking several drinks at least three nights a week, but you don't know how to withdraw from the group in which this is normal routine behavior. What action would you take?	Ask a close friend to help watch out for your best interests, and pursue other activities with other people. As long as you keep your grades up it is not a problem. (Worst) Explain to the group that you are concerned about falling behind if you continue the behavior and concentrate more on your studies instead. Join alternative groups such as campus clubs and sports, or maybe even take an evening or early morning job. (Best) Just socialize with the group less frequently. Continue socializing with the group, but don't always drink when they do.	χ^2 (df) RMSEA CFI NNFI SRMR	691.91 (502) .023 .882 .870 .035
You have been having trouble with a class in which everyone else seems to be doing well. Your homework comes back with unsatisfactory grades week after week, and your test scores have been marginally passing. How would you proceed?	Find a study group to work with you. Talk to the professor, and to friends in the class, and read more. Get tutoring, and study more frequently for this class. Seek help from someone in the class who is doing well. Talk to the professor or TA to find out what you are doing wrong, compare notes with others and seek out tutoring. (Best) Stay calm and continue to do the best you can. (Worst)	χ^2 (df) RMSEA CFI NNFI SRMR	691.85 (502) .023 .882 .870 .035

Table A10 (cont'd)

There is a seminar being held on campus that would expand your understanding of a class topic, but the seminar time conflicts with the class schedule. What would you do?	Skip the class, and go to the seminar because it is related to the class. (Worst) Go to class because it might cover what the seminar would cover. Go to class and talk to someone that went to the seminar. Get advice from the professor and then decide what to do. (Best)	χ^2 (df) RMSEA CFI NNFI SRMR	695.58 (502) .023 .880 .868 .035
You are the student coordinator for the gym, and it's 4:30 P.M. You have just been informed that there is no heat in the gym. As it is the middle of winter and very cold, you know this will be a problem. There is a student dance being held in the gym at 7:00 P.M., and there are no alternative facilities in which to hold the number of people expected at this event. What would you do?	Let everyone know that it's postponed or called off. Call maintenance, and see if they can fix it. (Best) Look for small heaters to fill the room. Call people and check the consensus opinion about what to do. Find a group of rooms as an alternative location. Inform the students to dress warmly. (Worst)	χ^2 (df) RMSEA CFI NNFI SRMR	694.09 (502) .023 .881 .869 .035

Table A10 (cont'd)

You and five other students must have a report ready within 48 hours. The last time the six of you worked together, you became the leader. You know that one of the group members did no work whatsoever on the last occasion, yet she is in your group again. This time it is necessary that all members pull their own weight. What would you do?	Let her know that you are aware that she did not do any work last time, and that this time it is necessary that she fully contribute. Do all of your end of the work and ensure that the instructor is aware that you did your share, regardless of what the other members do. Explain to the group that the professor will be made aware of who contributed what to the project, and ensure that this happens. (Best) Stress the importance that everyone fully contributes his or her share to the project. Work as closely with her as possible (e.g. assign both of you a related task) so as to offer encouragement and ensure that her work gets done. Assign her a specific task with a specific timeframe. If she does not do the work, ask to have her re-assigned, and have the group pick up her work.	χ^2 (df) RMSEA CFI NNFI SRMR	694.11 (502) .023 .881 .869 .035
You are collaborating with other classmates on a project. The group of you keeps running into a variety of problems that threaten to cause the project to be late. The other group members want to just plan to submit it late. Another option would be to devote much more time than planned to the project and possibly get it in on time. What would you do?	Try to get it done, but plan to submit it late. (Worst) Ask the instructor for help or for an extension. If that doesn't work, just try your best and do what you can or turn it in late. Motivate the group to devote more time and work together to get it done. (Best) Have the group decide what to do. (Worst) Work hard to finish it because there are consequences for being late and meeting deadlines is important to you. (Best) Tell the instructor your situation, and ask for advice.	χ^2 (df) RMSEA CFI NNFI SRMR	691.65 (502) .023 .882 .870 .035

Table A10 (cont'd.)

You know that a group of students in your class cheats on exams by putting formulas into scientific calculators, cell phones, or some electronic device. The professor has clearly warned against such activity, but you are not sure what she would do if she knew what these students were doing. What action would you take?	Try doing the same thing until people start getting caught. (Worst) Study the way you know best, don't cheat, but don't turn in the other students either. (Best) You would do nothing; it's none of your business. You would mention it to the professor so she can deal with the problems in the class. Don't tell the professor, but make sure it is clear you are not involved in case they get caught. Send the professor an anonymous message about what is going on. (Best)	χ^2 (df) RMSEA CFI NNFI SRMR	692.42 (502) .023 .882 .870 .035
Because of family problems, you realize that your parents can no longer support you financially at the same level as they have and you do not have enough money to continue in school. What plans would you make?	Apply for student financial aid or get a part-time job. (Best) Ask other family members for money to finish school. Drop out of school and save money for going back. (Best) Take fewer classes because of the lower level of finances.	χ^2 (df) RMSEA CFI NNFI SRMR	694.40 (502) .023 .880 .869 .035

Table A10 (cont'd)

An event in the news makes you wonder about the history behind the news incident. What would you do?	Do some research, looking up all the facts for yourself. Do a quick Internet search to see if you could find any information. (Best) Think about it briefly, then move on. (Worst) Ask others what they know about the topic. (Best) Resolve to read the newspaper more often.	χ^2 (df) RMSEA CFI NNFI SRMR	692.43 (502) .023 .882 .870 .035
You are finding a particular class dull and boring, and are having difficulty staying awake. What would you do?	Do what you can to stay awake, such as drinking caffeine or sitting toward the front of the class. (Best) Read the class material beforehand to make the lecture more interesting. During the lecture, do some studying that is required for the course. Make sure you are getting enough sleep every school night. (Best) Skip the class if it is that dull and boring to you. (Worst)	χ^2 (df) RMSEA CFI NNFI SRMR	**698.38 (502)** **.023** **.878** **.866** **.035**
Your grade for a particular class is based on three exams, with no class attendance requirement. All of the homework requirements for the class are posted on the professor's web site. What would you do?	Attend class for as long as you feel that it is helping your grades. Do all the homework but only go to some of the lectures. It's the exams that count. Go to all the classes anyway. The professor may say something important. (Best) Skip classes, but if you did poorly on the first exam, start going to classes. There is no need to go to classes. Just get the homework done, and pass the exams. (Worst)	χ^2 (df) RMSEA CFI NNFI SRMR	692.18 (502) .023 .882 .870 .035

Table A10 (cont'd)

You share a dorm room with three other students. One half-hour before you are expecting a guest, you get home to find the place completely trashed. There is no sign of any of your roommates. What would you do?	Clean up the mess as much as possible before the guest arrives. Then speak with your roommates immediately upon their return, so your guest knows how concerned you were about the mess. Leave the mess and explain the situation to your guest. (Worst) Leave the mess and take the guest somewhere else. Clean up the mess as much as possible before the guest arrives. Then, without the guest around, ask the roommates why the place was trashed so badly and what can be done in the future to avoid this situation. (Best)	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	692.46 (502) .023 .882 .870 .035
One of your friends' roommates frequently parties until late at night, often returning to the room after drinking, engaging in loud and obnoxious behavior. Your friend finds that she cannot study or sleep well, but also feels reluctant or afraid to talk with the dorm authorities. What action would you take?	Approach the dorm authorities on behalf of your friend. Talk to the roommate yourself, and explain that her behavior bothers your friend. (Worst) Tell your friend to talk with her roommate and let her know that the behavior is not acceptable. Offer to let your friend stay with you when necessary. Suggest to your friend that she talk it out with the roommate, and offer to be available as a neutral third party when the two have the conversation. (Best)	χ^2 *(df)* *RMSEA* *CFI* *NNFI* *SRMR*	693.66 (502) .023 .881 .869 .035

Table A10 (cont'd)

You are searching for a major that interests you and think you might be interested in psychology. You do not know much about preparation to be a psychologist or what kinds of opportunities exist for careers in this area. What action would you take?	Talk to an advisor in psychology to see what career options are available. (Best) Talk with a friend who is a psychology major to see what it is about. Take an introductory psychology course to see what areas in psychology there are. Look up job listings for psychologists on the Internet. (Worst)	χ^2 (df) RMSEA CFI NNFI SRMR	691.93 (502) .023 .882 .870 .035
You are interested in several different classes/disciplines, but don't know anything about future educational or career opportunities in these areas. What steps would you take to get informed?	Go to an advisor or knowledgeable professional who might tell you more and answer your questions. (Best) Research topics using available resources like relevant books and Internet web sites. Attempt to obtain some hands-on experience, like internships. (Best) Use the school career services and career counselors. Take some introductory classes in the area of interest to see if you want to pursue that area further. Think about your interests and try to figure out which of them fit with the different disciplines. Ask friends and family for advice and information. If possible ask a friend who is familiar with the area.	χ^2 (df) RMSEA CFI NNFI SRMR	696.83 (502) .023 .879 .867 .035

Table A10 (cont'd)

In a class of 50 students, you discover that a group of your friends have worked out a scheme to share answers on an exam. The professor has vision problems and will likely never notice. You are not doing very well in the course. What would you do in these circumstances?	Avoid being around these friends. It is not exactly honest but under the circumstances, the scheme is OK. You would join them. Do your own work and not tell the professor about the scheme because it is not your problem. (Best) Cheat and get a good grade. (Worst) Tell the professor about the scheme. Study for the exam, but join the scheme as a backup strategy for the test.	χ^2 (df) RMSEA CFI NNFI SRMR	692.85 (502) .023 .881 .870 .035
Your professor has just given you a project that will obviously require the whole semester to complete. She gave you all the details you need to get started, but you are not sure how the project should proceed from there. She does not appear to intend to give you any more information in class. What would you do?	Work out the project to the best of your ability and approach the professor if you get stuck. (Worst) Generate some ideas, and then go to office hours to see how the professor responds to them. Ask the professor about the project after class. Visit the professor or a teaching assistant during office hours to discuss the project. (Best) Talk to other students to get an idea of what they are doing. Try to get an idea of whether or not other students seem confused. If so, bring the issue up with the professor during class. (Worst)	χ^2 (df) RMSEA CFI NNFI SRMR	692.97 (502) .023 .881 .869 .035

Table A10 (cont'd)

You are part of a committee to reduce cross-cultural tension in your dorm. A group of students in your dorm complain to you that people always wish them "Merry Christmas" or "Happy Easter" when these holidays are not meaningful to them. They request that their differences be respected. How would you address this problem?	Ask the group to politely ignore the greetings with the realization that the people had good intentions. (Worst) Tell the well-wishers to please respectfully refrain from making specific holiday greetings. (Worst) Have a meeting at which people can discuss their differences and hopefully work out an understanding. (Best) As part of the committee, make all cultural holidays visible so that people can be aware of diversity. (Best) Tell them to respond with a meaningful greeting of their own.	χ^2 (df) RMSEA CFI NNFI SRMR	692.58 (502) .023 .882 .870 .035
A friend on your floor is always organizing "social" activities including trips to local bars. Aside from the fact that this person is underage and failing some classes, you realize that the individual is drinking half a dozen or more drinks at least three or four times a week. No one else seems to know or be concerned about the person. What would you do?	Talk to him/her about easing up on the alcohol, explaining that it will not help with his/her classes, which should be the main reason why he/she is in college. Use humor to broach the topic and offer alternatives to his/her usual "social" activities. Bring up the situation with the floor's resident assistant. Try to get him/her involved in other activities. (Best) Talk to the person to subtly determine if there are other issues that need to be addressed, and refer him/her to help if appropriate. (Best) Talk to other people on the floor, and discuss ways to address the situation. Ask him/her once about this behavior and see where the discussion leads, then leave him/her to his/her own course of action. (Worst)	χ^2 (df) RMSEA CFI NNFI SRMR	692.88 (502) .023 .881 .870 .035

Table A10 (cont'd)

Note. * denotes item identified as referent item. Fit statistics displayed by referent item are for the estimation of the configural model. Gender and Race denote multiple groups models comparing gender and race demographic groups, respectively. Fit statistics presented in bold denote indication of DIF as determined by a CFI decrease of > .002. (Reversed) denotes that lower item responses relate to a higher standing on the target scale, and that item responses were reversed prior to analyses. (Best) and (Worst) denote the responses rated as best and worst by subject matter experts. For further scoring information see Oswald et al. (2004).

APPENDIX B:

MIMIC model analyses for studied scales

Table B1.
MIMIC model of the Behavioral Leadership scale

Outcome	Item Responses	Predictor	Model 1 β (S.E.)	Model 1 p	Model 2 β (S.E.)	Model 2 p	Model 3 β (S.E.)	Model 3 p
Behavioral Leadership Factor		Gender	.05 (.06)	.400	.06 (.07)	.345	-.06 (.08)	.409
		Black	-.32 (.11)	.003	-.36 (.12)	.003	-.40 (.13)	.002
		Asian	-.53 (.13)	<.001	-.53 (.14)	<.001	-.46 (.13)	<.001
		Other	-.30 (.10)	.004	-.35 (.11)	.001	-.37 (.10)	<.001
		Pell Eligibility			.08 (.07)	.296	.04 (.07)	.560
		High School			-.02 (.04)	.598	-.04 (.04)	.286
		Realistic					-.08 (.04)	.052
		Investigative					-.01 (.04)	.836
		Artistic					.05 (.03)	.162
		Social					.16 (.04)	<.001
		Enterprising					.21 (.04)	<.001
		Conventional					-.12 (.04)	.002

Table B1 (cont'd)

Effect on Items

How many times in the past year have you set the schedule (time and/or tasks) for groups in which you have worked?	Never Once Twice Three or four times Five times or more	Variable	Est (SE)	p	Est (SE)	p	Est (SE)	p
		Gender	.04 (.05)	.390	.04 (.05)	.391	.03 (.07)	.650
		Black	-.08 (.08)	.350	-.04 (.09)	.638	.06 (.11)	.555
		Asian	.45 (.07)	<.001	.44 (.08)	<.001	.40 (.10)	<.001
		Other	.11 (.08)	.193	.11 (.09)	.213	.03 (.10)	.791
		Gender X Factor	-.01 (.02)	.795	-.01 (.03)	.569	-.01 (.03)	.829
		Black X Factor	-.04 (.03)	.154	-.03 (.03)	.218	-.03 (.03)	.253
		Asian X Factor	.01 (.02)	.667	.01 (.02)	.614	.01 (.02)	.742
		Other X Factor	.03 (.02)	.216	.04 (.02)	.062	.04 (.02)	.078
		Pell Eligibility			.00 (.06)	.997	.00 (.06)	.944
		High School			.06 (.03)	.031	.06 (.03)	.024
		Realistic					-.01 (.03)	.767
		Investigative					.00 (.03)	.879
		Artistic					.00 (.03)	.885
		Social					-.01 (.03)	.758
		Enterprising					-.06 (.03)	.042
		Conventional					.10 (.03)	.001

Table B1 (cont'd)

In the past year, how many times have you been responsible for assigning tasks and setting deadlines for other people?							
Never Once Twice Three or four times Five times or more	Gender	.16 (.05)	<.001	.18 (.05)	<.001	.24 (.07)	.001
	Black	.05 (.07)	.491	.14 (.08)	.071	.23 (.11)	.032
	Asian	.33 (.08)	<.001	.33 (.08)	<.001	.33 (.09)	<.001
	Other	.04 (.08)	.598	.03 (.08)	.694	.05 (.10)	.586
	Gender X Factor	-.02 (.02)	.343	-.04 (.02)	.110	-.03 (.02)	.192
	Black X Factor	-.04 (.02)	.073	-.03 (.02)	.169	-.03 (.02)	.169
	Asian X Factor	-.01 (.02)	.670	-.01 (.02)	.632	-.01 (.02)	.741
	Other X Factor	-.01 (.03)	.655	.00 (.03)	.895	.00 (.03)	.913
	Pell Eligibility			-.07 (.05)	.196	-.07 (.05)	.196
	High School			.09 (.03)	<.001	.10 (.03)	<.001
	Realistic					.03 (.03)	.295
	Investigative					-.02 (.03)	.370
	Artistic					.01 (.03)	.584
	Social					-.07 (.03)	.010
	Enterprising					-.11 (.03)	<.001
	Conventional					.06 (.03)	.034

Table B1 (cont'd)

In the past year, how many times have you been responsible for assigning tasks and setting deadlines for other people?	I am usually the one who assigns tasks or roles to get the work done More than half the time I end up assigning the tasks and roles About half the time I take the lead in assigning tasks and roles I rarely take the lead in assigning tasks and roles I never take the lead unless I have been assigned to do so	Gender	.13 (.04)	.004	.12 (.04)	.010	.15 (.06)	.017
		Black	.09 (.07)	.233	.09 (.08)	.272	.17 (.11)	.134
		Asian	-.02 (.08)	.821	-.03 (.09)	.712	-.07 (.10)	.466
		Other	.07 (.07)	.314	.09 (.08)	.210	.08 (.10)	.394
		Gender X Factor	.02 (.02)	.344	.02 (.02)	.417	.01 (.02)	.720
		Black X Factor	-.03 (.02)	.265	-.02 (.02)	.459	-.01 (.03)	.575
		Asian X Factor	-.01 (.02)	.618	.01 (.02)	.757	.01 (.02)	.531
		Other X Factor	.00 (.02)	.959	.01 (.02)	.677	.01 (.02)	.577
		Pell Eligibility			-.02 (.05)	.639	-.02 (.05)	.758
		High School			-.01 (.02)	.774	-.01 (.02)	.606
		Realistic					-.02 (.03)	.369
		Investigative					.02 (.02)	.398
		Artistic					-.03 (.02)	.196
		Social					-.07 (.03)	.008
		Enterprising					-.02 (.03)	.485
		Conventional					.08 (.03)	.005

Note. Effects listed with "X Factor" denote an interaction between particular demographic grouping variable and the standing on the latent factor score.

Table B2.
MIMIC model of the Leadership Positions scale

Outcome	Item Responses	Predictor	Model 1 β (S.E.)	p	Model 2 β (S.E.)	p	Model 3 β (S.E.)	p
Leadership Positions Factor		Gender	.06 (.06)	.352	.06 (.06)	.325	-.01 (.07)	.888
		Black	.02 (.10)	.872	-.01 (.12)	.928	-.02 (.12)	.838
		Asian	.02 (.12)	.887	.03 (.13)	.797	.05 (.13)	.712
		Other	-.21 (.10)	.031	-.26 (.10)	.009	-.27 (.10)	.006
		Pell Eligibility			-.03 (.07)	.636	-.05 (.07)	.485
		High School			-.07 (.03)	.031	-.08 (.03)	.009
		Realistic					-.01 (.04)	.788
		Investigative					-.03 (.03)	.333
		Artistic					.04 (.03)	.200
		Social					.15 (.03)	<.001
		Enterprising					.12 (.04)	.002
		Conventional					-.01 (.04)	.735

Table B2 (cont'd)

Effect on Items

The number of high school clubs and organized activities (such as band, sports, newspapers, etc.) in which you took a leadership role was:	I did not take a leadership role 1 2 3 4 or more	Effect	Model 1		Model 2		Model 3	
		Gender	.15 (.04)	.001	.14 (.05)	.001	.18 (.06)	.005
		Black	.03 (.07)	.729	-.02 (.08)	.852	.08 (.13)	.540
		Asian	-.06 (.08)	.490	-.05 (.08)	.549	-.02 (.10)	.840
		Other	.04 (.07)	.587	.00 (.08)	.977	.07 (.10)	.461
		Gender X Factor	-.01 (.02)	.715	-.01 (.02)	.607	-.01 (.02)	.517
		Black X Factor	-.04 (.02)	.071	-.04 (.02)	.073	-.04 (.02)	.132
		Asian X Factor	.00 (.02)	.909	.00 (.02)	.839	-.01 (.02)	.756
		Other X Factor	-.01 (.02)	.549	-.02 (.02)	.305	-.03 (.02)	.241
		Pell Eligibility			.02 (.05)	.758	.01 (.05)	.914
		High School			.00 (.02)	.963	-.01 (.02)	.667
		Realistic					-.03 (.03)	.324
		Investigative					.02 (.02)	.490
		Artistic					-.01 (.02)	.551
		Social					-.02 (.03)	.542
		Enterprising					.07 (.03)	.009
		Conventional					-.04 (.03)	.206

Note. Effects listed with "X Factor" denote an interaction between particular demographic grouping variable and the standing on the latent factor score.

109

Table B3.
MIMIC model of the Knowledge scale

Outcome	Item Responses	Predictor	Model 1 β (S.E.)	p	Model 2 β (S.E.)	p	Model 3 β (S.E.)	p
Knowledge Factor		Gender	.00 (.08)	.983	-.05 (.09)	.547	-.16 (.10)	.131
		Black	-.42 (.13)	.001	-.50 (.15)	.001	-.46 (.15)	.002
		Asian	-.35 (.14)	.013	-.38 (.15)	.012	-.41 (.15)	.007
		Other	-.24 (.14)	.072	-.22 (.14)	.134	-.20 (.14)	.157
		Pell Eligibility			.06 (.10)	.562	.07 (.10)	.449
		High School			-.06 (.05)	.210	-.06 (.05)	.246
		Realistic					-.12 (.06)	.038
		Investigative					.13 (.05)	.005
		Artistic					-.03 (.05)	.507
		Social					.13 (.05)	.006
		Enterprising					-.04 (.05)	.500
		Conventional					.10 (.05)	.049

Table B3 (cont'd)

Effect on Items

For classwork, how often do you tend to skim the material, reading only the important points?	Almost all the time Most of the time Sometimes Rarely Never						
		Gender	.20 (.05)	<.001	.22 (.06)	<.001	.16 (.07)	.032
		Black	-.28 (.10)	.004	-.29 (.11)	.011	-.15 (.11)	.167
		Asian	-.05 (.10)	.600	-.06 (.11)	.563	-.03 (.10)	.743
		Other	-.09 (.08)	.243	-.13 (.09)	.133	-.05 (.12)	.664
		Gender X Factor	.04 (.03)	.255	.03 (.03)	.433	.02 (.03)	.562
		Black X Factor	-.04 (.03)	.244	-.04 (.04)	.219	-.05 (.04)	.208
		Asian X Factor	-.02 (.03)	.451	-.01 (.03)	.690	-.02 (.04)	.619
		Other X Factor	-.04 (.03)	.275	-.04 (.04)	.293	-.04 (.04)	.274
		Pell Eligibility			.08 (.06)	.175	.09 (.06)	.161
		High School			.06 (.03)	.052	.06 (.03)	.036
		Realistic					.00 (.03)	.916
		Investigative					.03 (.03)	.346
		Artistic					.00 (.03)	.928
		Social					.00 (.03)	.952
		Enterprising					-.04 (.03)	.270
		Conventional					-.04 (.03)	.244

Table B3 (cont'd)

(Reversed) In general, what is the lowest grade that you find acceptable for yourself?			B (SE)	p	B (SE)	p	B (SE)	p
(Reversed) In general, what is the lowest grade that you find acceptable for yourself?	A or equivalent	Gender	.00 (.05)	.981	-.01 (.06)	.812	.11 (.07)	.121
	B or equivalent	Black	-.42 (.09)	<.001	-.50 (.11)	<.001	-.38 (.10)	<.001
	C or equivalent	Asian	-.12 (.13)	.357	-.20 (.16)	.193	-.05 (.17)	.766
	D or equivalent	Other	-.16 (.08)	.058	-.25 (.10)	.008	-.07 (.12)	.574
	F or equivalent	Gender X Factor	-.04 (.03)	.153	-.03 (.03)	.267	-.04 (.03)	.156
		Black X Factor	-.06 (.03)	.070	-.04 (.03)	.203	-.02 (.04)	.507
		Asian X Factor	-.07 (.04)	.093	-.08 (.04)	.039	-.10 (.04)	.015
		Other X Factor	-.08 (.03)	.006	-.07 (.03)	.027	-.07 (.03)	.025
		Pell Eligibility			-.07 (.06)	.291	-.04 (.06)	.458
		High School			-.07 (.03)	.011	-.06 (.03)	.018
		Realistic					-.03 (.03)	.358
		Investigative					.12 (.03)	<.001
		Artistic					-.04 (.03)	.225
		Social					-.04 (.03)	.191
		Enterprising					-.06 (.03)	.102
		Conventional					.15 (.03)	<.001

Table B3 (cont'd)

			Model 1		Model 2		Model 3	
(Reversed) How often do you spend extra time on school assignments, even after they are turned in, so that you can gain a better understanding of the material or principles?	Very often Often Sometimes Rarely Never	Gender	.13 (.06)	.029	.15 (.06)	.019	.17 (.08)	.026
		Black	.19 (.11)	.070	.13 (.12)	.275	.19 (.11)	.100
		Asian	.28 (.10)	.007	.27 (.12)	.023	.28 (.10)	.006
		Other	.15 (.09)	.109	.11 (.11)	.315	.12 (.10)	.246
		Gender X Factor	.03 (.03)	.234	.03 (.03)	.356	.01 (.03)	.622
		Black X Factor	-.03 (.03)	.320	-.02 (.03)	.531	-.04 (.03)	.163
		Asian X Factor	-.02 (.03)	.511	-.02 (.03)	.530	-.02 (.03)	.390
		Other X Factor	-.03 (.03)	.324	-.01 (.03)	.802	-.02 (.03)	.597
		Pell Eligibility			.13 (.07)	.053	.14 (.07)	.041
		High School			-.03 (.03)	.373	-.03 (.03)	.364
		Realistic					.05 (.04)	.222
		Investigative					.04 (.03)	.266
		Artistic					.03 (.03)	.351
		Social					-.01 (.03)	.779
		Enterprising					-.02 (.04)	.541
		Conventional					.02 (.04)	.671

Table B3 (cont'd)

Generally, whenever you learn about a topic or how to perform a task, how often do you learn all the details as well as the general principles?	Hardly ever Not very often Sometimes Often Almost always						
		Gender	-.03 (.06)	.602	-.01 (.06)	.857	-.04 (.08)	.665
		Black	.46 (.09)	<.001	.47 (.11)	<.001	.56 (.12)	<.001
		Asian	.07 (.12)	.550	.04 (.14)	.800	.11 (.12)	.364
		Other	.07 (.09)	.430	.08 (.10)	.438	.05 (.11)	.613
		Gender X Factor	.03 (.03)	.353	.02 (.03)	.427	.02 (.03)	.559
		Black X Factor	-.02 (.04)	.596	-.03 (.05)	.545	-.05 (.03)	.144
		Asian X Factor	-.03 (.04)	.334	-.03 (.04)	.386	-.04 (.04)	.277
		Other X Factor	.00 (.03)	.925	.00 (.03)	.937	.00 (.03)	.909
		Pell Eligibility			.04 (.07)	.496	.05 (.07)	.469
		High School			.02 (.03)	.638	.01 (.03)	.756
		Realistic					.01 (.04)	.724
		Investigative					.04 (.03)	.215
		Artistic					.03 (.03)	.327
		Social					.01 (.03)	.693
		Enterprising					.03 (.04)	.483
		Conventional					-.03 (.03)	.451

Table B3 (cont'd)

			Est (SE)	p	Est (SE)	p	Est (SE)	p
(Reversed) When you took classes that you thought were easy, how important was it for you still to understand the concepts underlying the class material?	Extremely important Very important Rather important Sort of important Not important	Gender	.05 (.05)	.327	.07 (.06)	.185	.12 (.08)	.123
		Black	.31 (.11)	.004	.33 (.12)	.004	.32 (.11)	.004
		Asian	.15 (.10)	.142	.13 (.11)	.226	.13 (.12)	.259
		Other	.16 (.08)	.057	.17 (.09)	.068	.16 (.10)	.113
		Gender X Factor	.00 (.03)	.936	-.01 (.03)	.705	-.01 (.03)	.677
		Black X Factor	.00 (.04)	.904	.00 (.04)	.934	-.02 (.03)	.613
		Asian X Factor	.00 (.03)	.866	.00 (.03)	.961	.01 (.03)	.854
		Other X Factor	-.02 (.03)	.426	.00 (.03)	.907	-.01 (.03)	.811
		Pell Eligibility			.05 (.06)	.431	.04 (.06)	.523
		High School			.00 (.03)	.886	.01 (.03)	.866
		Realistic					.07 (.04)	.060
		Investigative					-.04 (.03)	.273
		Artistic					.05 (.03)	.114
		Social					-.02 (.03)	.535
		Enterprising					-.01 (.03)	.706
		Conventional					-.02 (.04)	.602

Table B3 (cont'd)

A year after completing a class, how much can you typically remember about what you were taught?	I tend to forget most of what was taught in class I remember the general ideas that were taught in class I remember some of the details that were taught in class I remember a lot of the details that were taught in class	Predictor						
		Gender	-.25 (.05)	<.001	-.24 (.06)	<.001	-.23 (.08)	.005
		Black	-.07 (.11)	.505	-.10 (.13)	.457	-.07 (.11)	.562
		Asian	-.12 (.12)	.306	-.18 (.13)	.172	-.20 (.11)	.060
		Other	.02 (.10)	.816	.02 (.11)	.844	.02 (.10)	.869
		Gender X Factor	-.04 (.03)	.145	-.02 (.03)	.463	-.03 (.03)	.391
		Black X Factor	-.01 (.04)	.836	-.02 (.05)	.724	-.01 (.05)	.859
		Asian X Factor	.00 (.03)	.960	-.01 (.03)	.880	-.01 (.03)	.858
		Other X Factor	.01 (.03)	.703	.00 (.04)	.988	-.01 (.04)	.890
		Pell Eligibility			.07 (.06)	.290	.07 (.06)	.246
		High School			.02 (.03)	.541	.03 (.03)	.366
		Realistic					.02 (.04)	.678
		Investigative					.02 (.03)	.640
		Artistic					.06 (.03)	.046
		Social					.02 (.03)	.624
		Enterprising					-.05 (.04)	.171
		Conventional					.05 (.04)	.140

Table B3 (cont'd)

How do you compare your standards for learning to those of your high school teachers?	Much lower than my teachers' standards Lower than my teachers' standards About the same than my teachers' standards Higher than my teachers' standards Much higher than my teachers' standards							
		Gender	-.09 (.06)	.117	-.11 (.06)	.053	-.04 (.08)	.581
		Black	.17 (.10)	.088	.11 (.12)	.381	.19 (.10)	.063
		Asian	.07 (.11)	.541	.04 (.14)	.763	.08 (.12)	.511
		Other	-.09 (.09)	.321	-.09 (.09)	.307	-.02 (.10)	.849
		Gender X Factor	-.03 (.03)	.407	-.03 (.03)	.436	-.03 (.03)	.323
		Black X Factor	-.03 (.04)	.480	-.02 (.04)	.597	-.03 (.04)	.413
		Asian X Factor	-.04 (.04)	.288	-.03 (.04)	.412	-.03 (.04)	.436
		Other X Factor	-.04 (.03)	.161	-.03 (.03)	.276	-.03 (.03)	.251
		Pell Eligibility			.00 (.06)	.974	.02 (.06)	.797
		High School			-.05 (.03)	.119	-.04 (.03)	.213
		Realistic					-.03 (.03)	.437
		Investigative					.15 (.03)	<.001
		Artistic					.04 (.03)	.211
		Social					.00 (.03)	.949
		Enterprising					.01 (.03)	.881
		Conventional					.07 (.03)	.052

Note. Effects listed with "X Factor" denote an interaction between particular demographic grouping variable and the standing on the latent factor score.

Table B4.
MIMIC model of Continuous Learning scale

Outcome	Item Responses	Predictor	Model 1		Model 2		Model 3	
			β (S.E.)	p	β (S.E.)	p	β (S.E.)	p
Continuous Learning Factor		Gender	-.05 (.06)	.457	-.05 (.06)	.395	-.13 (.08)	.079
		Black	.14 (.09)	.145	.03 (.10)	.798	.01 (.11)	.923
		Asian	.10 (.13)	.431	.04 (.13)	.736	-.03 (.13)	.791
		Other	.04 (.11)	.706	.01 (.11)	.925	-.03 (.11)	.764
		Pell Eligibility			.21 (.07)	.003	.23 (.07)	.001
		High School			-.02 (.03)	.500	-.01 (.03)	.662
		Realistic					-.02 (.04)	.695
		Investigative					.10 (.03)	.004
		Artistic					.19 (.03)	<.001
		Social					.05 (.04)	.149
		Enterprising					-.08 (.04)	.030
		Conventional					.09 (.04)	.020

Table B4 (cont'd)

Effect on Items

In the past month, how many times have you looked for more information about something that you found interesting?	Never Once or twice 3 to 5 times 6 to 10 times More than 10 times						
		Gender	-.42 (.05)	<.001	-.43 (.05)	<.001	-.54 (.08)	<.001
		Black	-.10 (.08)	.227	-.07 (.09)	.431	-.08 (.15)	.627
		Asian	.04 (.09)	.686	.06 (.09)	.518	.09 (.12)	.467
		Other	.00 (.08)	.993	-.01 (.08)	.918	-.04 (.11)	.755
		Gender X Factor	.02 (.02)	.306	.03 (.02)	.249	.02 (.02)	.310
		Black X Factor	.00 (.03)	.994	-.02 (.03)	.553	.00 (.03)	.962
		Asian X Factor	-.01 (.02)	.801	-.01 (.02)	.586	-.01 (.02)	.765
		Other X Factor	.00 (.02)	.841	.01 (.02)	.708	.01 (.02)	.823
		Pell Eligibility			-.08 (.06)	.170	-.08 (.05)	.157
		High School			.00 (.03)	.882	.00 (.03)	.881
		Realistic					-.02 (.03)	.425
		Investigative					-.02 (.03)	.421
		Artistic					.07 (.03)	.011
		Social					.02 (.03)	.520
		Enterprising					-.03 (.03)	.377
		Conventional					.02 (.03)	.439

Table B4 (cont'd)

			b (SE)	p	b (SE)	p	b (SE)	p
(Reversed) How often do you ask a teacher or classmate questions that go beyond the material but are still relevant to the topic (either in or out of class)?	Very often Often Sometimes Rarely Never Almost Never	Gender	-.12 (.05)	.013	-.13 (.05)	.008	-.12 (.08)	.141
		Black	.27 (.08)	.001	.20 (.09)	.023	.01 (.14)	.926
		Asian	-.20 (.10)	.044	-.18 (.10)	.079	.01 (.16)	.974
		Other	-.11 (.08)	.172	-.16 (.09)	.059	-.27 (.13)	.043
		Gender X Factor	.02 (.02)	.306	-.01 (.03)	.569	-.01 (.03)	.682
		Black X Factor	.00 (.03)	.994	.03 (.03)	.299	.04 (.03)	.132
		Asian X Factor	-.01 (.02)	.801	-.05 (.02)	.060	-.05 (.03)	.069
		Other X Factor	.00 (.02)	.841	.04 (.03)	.155	.03 (.03)	.203
		Pell Eligibility			.00 (.06)	.955	-.01 (.06)	.827
		High School			-.06 (.03)	.022	-.06 (.03)	.024
		Realistic					.01 (.03)	.680
		Investigative					.01 (.03)	.773
		Artistic					.00 (.03)	.987
		Social					.05 (.03)	.063
		Enterprising					.04 (.03)	.155
		Conventional					-.03 (.03)	.297

Table B4 (cont'd)

In the past month, how many times did you go out and learn more about something simply because it seemed interesting?	Never Once Twice Three or four times Five times or more	Gender	-.31 (.04)	<.001	-.30 (.05)	<.001	-.36 (.08)	<.001
		Black	-.14 (.08)	.079	-.08 (.09)	.364	-.04 (.15)	.781
		Asian	-.04 (.09)	.699	-.02 (.09)	.870	-.05 (.15)	.732
		Other	-.19 (.07)	.010	-.21 (.08)	.011	-.33 (.11)	.002
		Gender X Factor	.03 (.02)	.098	.03 (.02)	.172	.03 (.02)	.195
		Black X Factor	-.01 (.03)	.838	-.01 (.03)	.619	-.01 (.03)	.668
		Asian X Factor	.01 (.02)	.642	.01 (.02)	.660	.01 (.02)	.527
		Other X Factor	.04 (.02)	.018	.04 (.02)	.036	.03 (.02)	.081
		Pell Eligibility			-.04 (.05)	.433	-.04 (.05)	.438
		High School			.02 (.02)	.350	.02 (.02)	.390
		Realistic					.02 (.03)	.395
		Investigative					-.01 (.03)	.578
		Artistic					.04 (.03)	.120
		Social					.03 (.03)	.328
		Enterprising					.04 (.03)	.168
		Conventional					-.05 (.03)	.061

Table B4 (cont'd)

When learning new things, some people tend to feel stressed or tired, while others tend to feel inspired or refreshed. How do you tend to feel when you learn new things?	Very stressed/tired Somewhat stressed/tired Something in between stressed/tired and inspired/refreshed Somewhat inspired/refreshed Very inspired/refreshed						
		Gender	-.21 (.05)	<.001	-.20 (.05)	<.001	-.13 (.08)	.105
		Black	.12 (.08)	.153	.07 (.09)	.413	.24 (.17)	.173
		Asian	-.13 (.08)	.113	-.19 (.09)	.029	-.04 (.12)	.731
		Other	.02 (.08)	.829	.02 (.08)	.773	.13 (.11)	.213
		Gender X Factor	-.01 (.02)	.751	-.02 (.03)	.494	-.02 (.03)	.385
		Black X Factor	-.04 (.03)	.222	-.05 (.04)	.197	-.05 (.04)	.229
		Asian X Factor	-.03 (.02)	.219	-.04 (.02)	.120	-.05 (.02)	.048
		Other X Factor	-.02 (.02)	.288	-.02 (.02)	.337	-.03 (.02)	.195
		Pell Eligibility			.02 (.05)	.746	.02 (.05)	.782
		High School			-.03 (.02)	.194	-.03 (.03)	.205
		Realistic					.06 (.03)	.055
		Investigative					.04 (.03)	.137
		Artistic					-.04 (.03)	.138
		Social					.02 (.03)	.486
		Enterprising					-.06 (.03)	.047
		Conventional					.01 (.03)	.697

Note. Effects listed with "X Factor" denote an interaction between particular demographic grouping variable and the standing on the latent factor score.

Table B5.

MIMIC model of the Perseverance scale

Outcome	Item Responses	Predictor	Model 1 β (S.E.)	p	Model 2 β (S.E.)	p	Model 3 β (S.E.)	p
Perseverance		Gender	.39 (.07)	<.001	.38 (.08)	<.001	.27 (.10)	.005
Factor		Black	.23 (.11)	.037	.18 (.12)	.143	.23 (.13)	.071
		Asian	-.29 (.14)	.037	-.28 (.15)	.058	-.26 (.15)	.076
		Other	.04 (.12)	.771	-.03 (.13)	.845	-.02 (.13)	.888
		Pell Eligibility			.13 (.09)	.129	.12 (.09)	.178
		High School			-.05 (.04)	.280	-.06 (.04)	.185
		Realistic					-.15 (.05)	.001
		Investigative					.08 (.04)	.062
		Artistic					-.04 (.04)	.270
		Social					.13 (.04)	.003
		Enterprising					.13 (.05)	.006
		Conventional					-.01 (.05)	.816

Table B5 (cont'd)

Effect on Items

To what extent would your friends describe you as someone who goes after what you want?	Not at all A slight extent A moderate extent A large extent A great extent	Gender	.03 (.06)	.672	.02 (.06)	.784	.04 (.08)	.677
		Black	-.12 (.10)	.209	-.18 (.10)	.088	-.31 (.14)	.030
		Asian	-.26 (.12)	.031	-.28 (.13)	.026	-.23 (.14)	.107
		Other	.01 (.09)	.955	-.03 (.09)	.753	.00 (.12)	.975
		Gender X Factor	-.02 (.03)	.602	-.02 (.03)	.523	-.01 (.03)	.786
		Black X Factor	.05 (.02)	.025	.04 (.02)	.064	.05 (.02)	.064
		Asian X Factor	-.01 (.03)	.884	-.01 (.03)	.715	-.01 (.03)	.734
		Other X Factor	-.03 (.03)	.331	-.04 (.03)	.221	-.04 (.03)	.236
		Pell Eligibility			.13 (.06)	.031	.13 (.06)	.040
		High School			.01 (.03)	.817	-.01 (.03)	.845
		Realistic					.01 (.04)	.897
		Investigative					-.05 (.03)	.082
		Artistic					.01 (.03)	.753
		Social					-.01 (.03)	.827
		Enterprising					.10 (.04)	.004
		Conventional					-.09 (.04)	.015

Table B5 (cont'd)

			Coef. (SE)	p	Coef. (SE)	p	Coef. (SE)	p
How frequently do you fail to get what you want because you did not put in enough effort?	Very often Often Sometimes Rarely Never	Gender	-.01 (.06)	.830	-.02 (.06)	.746	.01 (.09)	.944
		Black	-.28 (.10)	.004	-.37 (.11)	.001	-.33 (.17)	.049
		Asian	-.75 (.10)	<.001	-.77 (.11)	<.001	-.75 (.13)	<.001
		Other	-.21 (.10)	.030	-.21 (.1)	.040	-.13 (.15)	.377
		Gender X Factor	.03 (.03)	.358	.03 (.04)	.390	.02 (.04)	.613
		Black X Factor	-.03 (.03)	.290	-.04 (.04)	.306	-.01 (.04)	.738
		Asian X Factor	-.02 (.04)	.568	-.02 (.04)	.665	-.02 (.04)	.569
		Other X Factor	-.03 (.04)	.432	-.04 (.05)	.321	-.04 (.05)	.404
		Pell Eligibility			.08 (.06)	.184	.08 (.06)	.209
		High School			-.04 (.03)	.184	-.04 (.03)	.108
		Realistic					-.01 (.04)	.765
		Investigative					.06 (.03)	.041
		Artistic					-.09 (.03)	.002
		Social					-.01 (.03)	.879
		Enterprising					-.01 (.03)	.695
		Conventional					.06 (.03)	.070

Table B5 (cont'd)

(Reversed) How often have you accomplished something you initially thought was very difficult or almost impossible?	Very often Often Sometimes Rarely Never						
		Gender	.08 (.06)	.158	.10 (.06)	.105	.08 (.08)	.307
		Black	.19 (.11)	.089	.17 (.12)	.157	.07 (.17)	.710
		Asian	.07 (.10)	.483	.01 (.1)	.903	-.05 (.12)	.682
		Other	.03 (.10)	.786	.00 (.1)	.969	.04 (.13)	.763
		Gender X Factor	-.01 (.03)	.681	-.01 (.03)	.815	-.01 (.03)	.848
		Black X Factor	.01 (.03)	.818	.02 (.04)	.642	.01 (.04)	.798
		Asian X Factor	.03 (.03)	.253	.03 (.03)	.229	.04 (.03)	.205
		Other X Factor	.00 (.03)	.914	-.02 (.04)	.675	-.03 (.04)	.508
		Pell Eligibility			.10 (.06)	.086	.10 (.06)	.105
		High School			.06 (.03)	.044	.06 (.03)	.046
		Realistic					.06 (.04)	.098
		Investigative					-.06 (.03)	.058
		Artistic					.06 (.03)	.034
		Social					.05 (.03)	.133
		Enterprising					-.03 (.03)	.334
		Conventional					-.01 (.03)	.775

Table B5 (cont'd)

(Reversed) How often have you finished a project when faced with difficult circumstances?	Very often Often Sometimes Rarely Never						
		Gender	.02 (.06)	.703	.01 (.06)	.832	.01 (.08)	.866
		Black	-.29 (.10)	.006	-.26 (.12)	.025	-.40 (.17)	.016
		Asian	-.12 (.11)	.262	-.15 (.12)	.214	-.15 (.12)	.210
		Other	.06 (.11)	.583	.01 (.11)	.901	-.02 (.15)	.875
		Gender X Factor	-.01 (.03)	.676	-.03 (.03)	.334	-.03 (.03)	.286
		Black X Factor	.03 (.03)	.239	.04 (.03)	.221	.05 (.03)	.134
		Asian X Factor	-.01 (.03)	.671	.00 (.04)	.918	-.01 (.04)	.756
		Other X Factor	.00 (.04)	.922	.03 (.04)	.490	.02 (.04)	.617
		Pell Eligibility			.01 (.06)	.867	.01 (.06)	.828
		High School			.03 (.03)	.268	.03 (.03)	.226
		Realistic					.02 (.04)	.617
		Investigative					.02 (.03)	.446
		Artistic					.05 (.03)	.062
		Social					.06 (.03)	.089
		Enterprising					-.03 (.03)	.410
		Conventional					.04 (.03)	.228

Table B5 (cont'd)

(Reversed) How often do others tend to compliment you on your determination to continue with a project under difficult circumstances?	Very often Often Sometimes Rarely Never						
		Gender	.12 (.06)	.035	.09 (.06)	.138	.09 (.08)	.264
		Black	.14 (.11)	.188	.15 (.12)	.216	.08 (.16)	.603
		Asian	.04 (.11)	.679	.02 (.11)	.872	.05 (.12)	.684
		Other	.09 (.08)	.306	.09 (.09)	.282	.16 (.11)	.141
		Gender X Factor	-.01 (.03)	.836	.00 (.03)	.968	-.01 (.03)	.672
		Black X Factor	.05 (.03)	.056	.06 (.03)	.057	.04 (.03)	.172
		Asian X Factor	-.03 (.03)	.311	-.01 (.03)	.633	-.01 (.03)	.652
		Other X Factor	-.03 (.03)	.182	-.04 (.03)	.169	-.04 (.03)	.168
		Pell Eligibility			.05 (.06)	.459	.04 (.06)	.503
		High School			.02 (.03)	.373	.02 (.03)	.525
		Realistic					.01 (.04)	.801
		Investigative					-.06 (.03)	.064
		Artistic					.03 (.03)	.267
		Social					.04 (.03)	.206
		Enterprising					.02 (.03)	.646
		Conventional					.05 (.03)	.167

128

Table B5 (cont'd)

How often do you tend to give up on a task after being told that you were not doing well?	Almost all the time Most of the time Sometimes Rarely Never						
		Gender	-.27 (.06)	<.001	-.23 (.06)	<.001	-.24 (.08)	.004
		Black	.17 (.09)	.063	.08 (.10)	.396	.10 (.14)	.447
		Asian	-.21 (.12)	.080	-.25 (.13)	.045	-.35 (.14)	.010
		Other	.08 (.10)	.423	.08 (.10)	.443	.04 (.13)	.773
		Gender X Factor	.06 (.03)	.036	.04 (.03)	.248	.04 (.03)	.198
		Black X Factor	-.02 (.03)	.485	-.02 (.03)	.420	-.02 (.03)	.436
		Asian X Factor	.05 (.03)	.123	.05 (.03)	.110	.04 (.03)	.234
		Other X Factor	.03 (.03)	.375	.03 (.03)	.397	.03 (.03)	.398
		Pell Eligibility			.13 (.06)	.034	.12 (.06)	.044
		High School			-.04 (.03)	.207	-.04 (.03)	.235
		Realistic					.07 (.03)	.044
		Investigative					-.03 (.03)	.363
		Artistic					-.03 (.03)	.373
		Social					.03 (.03)	.393
		Enterprising					-.02 (.04)	.677
		Conventional					-.01 (.03)	.754

Table B5 (cont'd)

When encountering problems that take a long time to solve, how impatient do you tend to become?	Extremely impatient Very impatient Somewhat impatient Slightly impatient Not at all impatient							
		Gender	-.28 (.06)	<.001	-.27 (.06)	<.001	-.33 (.08)	<.001
		Black	-.04 (.10)	.689	-.07 (.11)	.554	-.09 (.15)	.533
		Asian	-.08 (.10)	.416	-.14 (.10)	.182	-.19 (.12)	.126
		Other	-.07 (.11)	.526	-.03 (.10)	.759	.02 (.13)	.860
		Gender X Factor	.03 (.03)	.328	.05 (.03)	.130	.03 (.03)	.392
		Black X Factor	.02 (.03)	.559	.01 (.03)	.650	.02 (.03)	.604
		Asian X Factor	.01 (.03)	.878	.01 (.03)	.799	-.01 (.03)	.872
		Other X Factor	.01 (.04)	.818	-.03 (.04)	.461	-.03 (.03)	.412
		Pell Eligibility			.07 (.06)	.289	.07 (.06)	.237
		High School			.01 (.03)	.705	.01 (.03)	.758
		Realistic					.05 (.04)	.187
		Investigative					.00 (.03)	.902
		Artistic					-.03 (.03)	.362
		Social					.07 (.03)	.028
		Enterprising					-.13 (.03)	<.001
		Conventional					.08 (.03)	.015

Note. Effects listed with "X Factor" denote an interaction between particular demographic grouping variable and the standing on the latent factor score.

Table B6.
MIMIC model of the Discrete Adaptability scale

Outcome	Item Responses	Predictor	Model 1 β (S.E.)	p	Model 2 β (S.E.)	p	Model 3 β (S.E.)	p
Discrete		Gender	.36 (.10)	<.001	.36 (.10)	<.001	.21 (.12)	.071
Adaptability		Black	-.42 (.16)	.007	-.47 (.17)	.006	-.53 (.18)	.003
Factor		Asian	-.68 (.20)	.001	-.73 (.21)	.001	-.74 (.21)	<.001
		Other	-.27 (.15)	.069	-.38 (.15)	.014	-.37 (.15)	.014
		Pell Eligibility			.05 (.11)	.677	.03 (.11)	.808
		High School			-.04 (.05)	.441	-.07 (.05)	.164
		Realistic					-.21 (.06)	<.001
		Investigative					.04 (.06)	.445
		Artistic					.03 (.05)	.512
		Social					.14 (.05)	.008
		Enterprising					.12 (.06)	.047
		Conventional					.15 (.06)	.010

Table B6 (cont'd)

Effect on Items

How often have you failed to meet responsibilities because you had taken on too much?	Very often Often Sometimes Rarely Never	Gender	-.15 (.06)	.018	-.15 (.06)	.015	-.06 (.10)	.543
		Black	.03 (.11)	.780	-.01 (.13)	.919	-.15 (.17)	.375
		Asian	-.14 (.13)	.287	-.11 (.16)	.492	-.02 (.14)	.879
		Other	-.02 (.11)	.893	.06 (.12)	.636	-.07 (.17)	.688
		Gender X Factor	.02 (.03)	.564	.01 (.04)	.804	-.01 (.04)	.790
		Black X Factor	.04 (.03)	.191	.05 (.03)	.121	.06 (.03)	.081
		Asian X Factor	-.03 (.04)	.404	-.02 (.04)	.542	-.02 (.04)	.595
		Other X Factor	.07 (.04)	.072	.05 (.04)	.237	.04 (.04)	.347
		Pell Eligibility			.03 (.07)	.621	.05 (.07)	.486
		High School			-.04 (.03)	.269	-.03 (.03)	.369
		Realistic					.07 (.04)	.055
		Investigative					.03 (.03)	.440
		Artistic					-.11 (.03)	.001
		Social					-.01 (.04)	.689
		Enterprising					-.03 (.04)	.448
		Conventional					-.07 (.04)	.072

Table B6 (cont'd)

(Reversed) How difficult has it been for you to continue with something after being interrupted and having to take care of something else?	Categories	Variable						
	Very easy	Gender	-.17 (.07)	.013	-.18 (.07)	.011	.00 (.10)	.976
	Easy	Black	.06 (.11)	.573	-.06 (.12)	.617	.08 (.15)	.600
	Not easy but not difficult	Asian	.01 (.12)	.936	-.05 (.14)	.741	.16 (.16)	.296
	Difficult	Other	.11 (.10)	.253	.12 (.11)	.279	.23 (.12)	.056
	Very difficult	Gender X Factor	-.04 (.03)	.155	-.03 (.03)	.201	-.05 (.03)	.049
		Black X Factor	-.03 (.03)	.328	-.02 (.03)	.512	-.02 (.03)	.549
		Asian X Factor	-.05 (.03)	.055	-.05 (.03)	.064	-.05 (.03)	.038
		Other X Factor	-.04 (.03)	.196	-.02 (.03)	.422	-.03 (.03)	.310
		Pell Eligibility			.12 (.08)	.108	.14 (.08)	.076
		High School			-.03 (.04)	.472	-.01 (.04)	.840
		Realistic					.15 (.05)	.001
		Investigative					.01 (.04)	.891
		Artistic					-.07 (.04)	.068
		Social					-.04 (.04)	.261
		Enterprising					-.09 (.04)	.041
		Conventional					-.09 (.04)	.031

Table B6 (cont'd)

(Reversed) How often do you plan ahead and make a specific schedule of things you need or want to do?							
	Very often	Gender	.38 (.06)	<.001	.41 (.06)	<.001	.48 (.09)	<.001
	Often	Black	.06 (.09)	.494	.06 (.10)	.525	.13 (.14)	.375
	Sometimes	Asian	.06 (.12)	.618	.04 (.14)	.794	.10 (.15)	.524
	Rarely	Other	.07 (.08)	.404	.10 (.09)	.251	.11 (.13)	.380
	Never	Gender X Factor	-.01 (.03)	.824	.00 (.03)	.978	.00 (.03)	.978
		Black X Factor	-.02 (.03)	.578	-.03 (.03)	.455	-.02 (.03)	.522
		Asian X Factor	-.01 (.04)	.774	-.01 (.05)	.905	-.02 (.05)	.630
		Other X Factor	.01 (.03)	.723	.00 (.03)	.923	.00 (.03)	.897
		Pell Eligibility			-.01 (.06)	.898	-.02 (.06)	.761
		High School			.01 (.03)	.644	.01 (.03)	.685
		Realistic					.07 (.03)	.033
		Investigative					-.03 (.03)	.259
		Artistic					-.08 (.03)	.011
		Social					.01 (.03)	.811
		Enterprising					.02 (.04)	.481
		Conventional					-.03 (.03)	.459

134

Table B6 (cont'd)

In the past, how difficult has it been for you to change your study habits to improve on a skill or to do better in a class	Very difficult / Difficult / Not easy but not difficult / Easy / Very easy						
		Gender	-.13 (.06)	.046	-.14 (.07)	.041	-.09 (.09)	.347
		Black	.00 (.11)	.988	-.04 (.12)	.718	-.06 (.15)	.695
		Asian	-.20 (.11)	.077	-.20 (.15)	.176	.02 (.14)	.907
		Other	-.05 (.10)	.624	-.07 (.11)	.549	.23 (.16)	.148
		Gender X Factor	-.01 (.03)	.671	.00 (.03)	.912	-.02 (.03)	.468
		Black X Factor	.01 (.03)	.641	.01 (.03)	.793	.02 (.03)	.418
		Asian X Factor	-.07 (.03)	.035	-.07 (.04)	.060	-.06 (.03)	.062
		Other X Factor	-.05 (.04)	.187	-.09 (.04)	.014	-.08 (.04)	.030
		Pell Eligibility			.04 (.07)	.543	.05 (.07)	.504
		High School			-.02 (.03)	.468	-.02 (.03)	.548
		Realistic					-.01 (.04)	.717
		Investigative					.08 (.03)	.014
		Artistic					-.13 (.03)	<.001
		Social					.04 (.04)	.323
		Enterprising					-.05 (.04)	.180
		Conventional					.03 (.04)	.449

Table B6 (cont'd)

When you are working on a serious and relatively difficult task and something or someone interrupts you, how do you usually react?	With a great deal of annoyance - it is hard to get back to the original task You are irritated - it's hard to stay on task when you are interrupted It bothers you just a little - you'd really prefer not to be interrupted It doesn't bother you - you feel one of the challenges of any job is the ability to "juggle" several things at a time							
		Gender	-.24 (.06)	<.001	-.23 (.07)	.001	-.24 (.10)	.016
		Black	.22 (.10)	.037	.11 (.12)	.354	.21 (.14)	.149
		Asian	.14 (.15)	.326	.12 (.17)	.486	.09 (.16)	.592
		Other	-.01 (.09)	.877	-.04 (.10)	.675	.10 (.14)	.457
		Gender X Factor	.02 (.03)	.570	.03 (.03)	.390	.01 (.03)	.693
		Black X Factor	-.01 (.03)	.764	-.01 (.03)	.809	-.01 (.03)	.755
		Asian X Factor	.01 (.04)	.761	.01 (.04)	.857	.02 (.04)	.664
		Other X Factor	-.02 (.03)	.578	-.04 (.03)	.264	-.05 (.03)	.141
		Pell Eligibility			.11 (.07)	.119	.12 (.07)	.081
		High School			-.04 (.03)	.235	-.03 (.03)	.288
		Realistic					.11 (.04)	.007
		Investigative					.00 (.04)	.937
		Artistic					-.05 (.03)	.123
		Social					.03 (.03)	.331
		Enterprising					-.06 (.04)	.112
		Conventional					-.07 (.04)	.077

Note. Effects listed with "X Factor" denote an interaction between particular demographic grouping variable and the standing on the latent factor score.

Table B7.
MIMIC model of the Routine Adaptability scale

Outcome	Item Responses	Predictor	Model 1 β (S.E.)	p	Model 2 β (S.E.)	p	Model 3 β (S.E.)	p
Routine		Gender	-.41 (.07)	<.001	-.36 (.07)	<.001	-.43 (.08)	<.001
Adaptability		Pell Eligibility			.02 (.08)	.829	.01 (.08)	.868
Factor		High School			.00 (.04)	.968	-.02 (.04)	.684
		Realistic					-.02 (.05)	.690
		Investigative					.00 (.04)	.946
		Artistic					-.06 (.04)	.117
		Social					.13 (.04)	.003
		Enterprising					.11 (.04)	.013
		Conventional					-.11 (.04)	.010
Effect on Items								
How often	Very often	Gender	.19 (.05)	<.001	.16 (.05)	.002	.16 (.07)	.020
have you	Often	Gender X Factor	.04 (.03)	.194	.03 (.03)	.286	.03 (.03)	.317
failed to meet	Sometimes	Pell Eligibility			.09 (.05)	.075	.09 (.05)	.075
responsibilities	Rarely	High School			-.05 (.03)	.080	-.05 (.03)	.065
because you	Never	Realistic					-.02 (.03)	.498
had taken on		Investigative					-.01 (.03)	.709
too much?		Artistic					.01 (.03)	.690
		Social					-.05 (.03)	.143
		Enterprising					-.02 (.03)	.537
		Conventional					.04 (.03)	.245

Note. Effects listed with "X Factor" denote an interaction between particular demographic grouping variable and the standing on the latent factor score.

Table B8.
MIMIC model of the Social Responsibility scale

Outcome	Item Responses	Predictor	Model 1 β (S.E.)	p	Model 2 β (S.E.)	p	Model 3 β (S.E.)	p
Social		Gender	.45 (.06)	<.001	.48 (.06)	<.001	.41 (.07)	<.001
Responsibility		Black	-.20 (.10)	.045	-.21 (.11)	.066	-.18 (.12)	.135
Factor		Asian	.33 (.09)	<.001	.38 (.10)	<.001	.40 (.10)	<.001
		Other	-.19 (.10)	.048	-.16 (.10)	.112	-.17 (.10)	.111
		Pell Eligibility			-.09 (.07)	.192	-.11 (.07)	.123
		High School			-.05 (.03)	.130	-.05 (.03)	.072
		Realistic					-.03 (.04)	.379
		Investigative					.05 (.03)	.150
		Artistic					-.05 (.03)	.127
		Social					.13 (.03)	<.001
		Enterprising					.06 (.04)	.131
		Conventional					-.02 (.04)	.613

Table B8 (cont'd)

Effect on Items

During the last year, how many times have you given money, food, or clothes to a charity or a poor person in need?							
0							
1							
2							
3							
More than 3							

	Estimate (SE)	p	Estimate (SE)	p	Estimate (SE)	p
Gender	.08 (.06)	.140	.06 (.06)	.318	.03 (.08)	.727
Black	.28 (.08)	.001	.28 (.09)	.002	.35 (.12)	.004
Asian	.04 (.12)	.739	.00 (.12)	.977	-.05 (.16)	.757
Other	.13 (.09)	.131	.08 (.09)	.419	.14 (.12)	.222
Gender X Factor	-.02 (.03)	.507	-.01 (.03)	.639	-.01 (.03)	.706
Black X Factor	-.05 (.03)	.115	-.04 (.03)	.157	-.05 (.03)	.141
Asian X Factor	.03 (.03)	.215	.04 (.03)	.162	.04 (.03)	.136
Other X Factor	-.05 (.03)	.110	-.05 (.03)	.118	-.05 (.03)	.127
Pell Eligibility			-.02 (.06)	.727	-.04 (.06)	.473
High School			.01 (.03)	.757	.00 (.03)	.891
Realistic					.04 (.03)	.294
Investigative					-.09 (.03)	.002
Artistic					-.01 (.03)	.857
Social					.10 (.03)	.001
Enterprising					.06 (.03)	.084
Conventional					-.05 (.03)	.109

Table B8 (cont'd)

In the past year, how many hours were you engaged in community service or volunteer activities?	None Less than 10 hours 11 - 40 hours 41 - 80 hours More than 80 hours						
		Gender	-.07 (.04)	.085	-.07 (.04)	.090	-.14 (.06)	.010
		Black	.25 (.08)	.001	.29 (.09)	.001	.15 (.10)	.130
		Asian	.13 (.09)	.157	.15 (.10)	.120	.06 (.12)	.627
		Other	.08 (.06)	.223	.10 (.07)	.159	.09 (.07)	.207
		Gender X Factor	.04 (.02)	.013	.04 (.02)	.034	.04 (.02)	.039
		Black X Factor	.05 (.02)	.002	.06 (.02)	.001	.06 (.02)	<.001
		Asian X Factor	.04 (.02)	.026	.04 (.02)	.024	.04 (.02)	.041
		Other X Factor	.01 (.02)	.450	.01 (.02)	.613	.01 (.02)	.508
		Pell Eligibility			.08 (.05)	.081	.08 (.05)	.109
		High School			.03 (.02)	.141	.04 (.02)	.090
		Realistic					.01 (.03)	.779
		Investigative					-.01 (.02)	.754
		Artistic					.02 (.02)	.335
		Social					.04 (.02)	.132
		Enterprising					-.04 (.03)	.140
		Conventional					.03 (.03)	.201

Table B8 (cont'd)

(Reversed) How important has it been in the past for you to be involved in community or volunteer work?	Extremely important Very important Important Not very important Not at all important							
		Gender	.20 (.04)	<.001	.21 (.05)	<.001	.08 (.06)	.166
		Black	.23 (.07)	.001	.22 (.08)	.004	.23 (.08)	.004
		Asian	-.02 (.08)	.842	-.04 (.08)	.658	.00 (.10)	.996
		Other	.03 (.02)	.164	.02 (.02)	.305	.01 (.02)	.433
		Gender X Factor	-.01 (.02)	.768	-.01 (.02)	.510	.00 (.02)	.871
		Black X Factor	-.01 (.02)	.598	-.01 (.02)	.739	-.01 (.02)	.502
		Asian X Factor	-.03 (.02)	.105	-.02 (.02)	.235	-.01 (.02)	.490
		Other X Factor			.07 (.05)	.170	.05 (.05)	.286
		Pell Eligibility			.01 (.02)	.795	.00 (.02)	.896
		High School					-.06 (.03)	.025
		Realistic					-.02 (.02)	.513
		Investigative					-.02 (.02)	.395
		Artistic					.08 (.02)	.002
		Social					-.04 (.03)	.111
		Enterprising					.01 (.03)	.605
		Conventional					.06 (.07)	.380

Table B8 (cont'd)

In the past year, in how many fundraisers have you participated?

None
1
2
3
4 or more

Predictor	Estimate (SE)	p	Estimate (SE)	p	Estimate (SE)	p
Gender	.08 (.05)	.140	.05 (.05)	.377	-.05 (.07)	.481
Black	-.02 (.09)	.836	-.06 (.09)	.539	-.09 (.11)	.398
Asian	-.21 (.10)	.038	-.19 (.11)	.067	-.29 (.14)	.031
Other	-.18 (.08)	.030	-.20 (.09)	.020	-.22 (.09)	.018
Gender X Factor	.04 (.02)	.118	.04 (.03)	.144	.04 (.03)	.155
Black X Factor	-.01 (.03)	.613	.00 (.03)	.929	.00 (.03)	.986
Asian X Factor	.06 (.02)	.002	.07 (.02)	.001	.07 (.02)	.001
Other X Factor	.01 (.03)	.669	.01 (.03)	.788	.01 (.03)	.702
Pell Eligibility			-.04 (.06)	.458	-.07 (.06)	.182
High School			-.04 (.03)	.123	-.06 (.03)	.021
Realistic					.00 (.03)	.915
Investigative					-.06 (.03)	.033
Artistic					-.01 (.03)	.630
Social					.12 (.03)	<.001
Enterprising					.06 (.03)	.047
Conventional					-.01 (.03)	.708

Table B8 (cont'd)

During the past year, how often have you recycled?	Never Not very often Sometimes Often Always							
		Gender	-.01 (.05)	.819	.01 (.06)	.804	.01 (.08)	.869
		Black	-.82 (.10)	<.001	-.62 (.11)	<.001	-.62 (.13)	<.001
		Asian	-.21 (.12)	.095	-.18 (.12)	.140	-.26 (.16)	.113
		Other	-.13 (.09)	.126	-.05 (.09)	.551	-.08 (.1)	.421
		Gender X Factor	.02 (.03)	.522	.02 (.03)	.489	.03 (.03)	.395
		Black X Factor	.03 (.04)	.505	.02 (.04)	.593	.01 (.04)	.866
		Asian X Factor	.04 (.03)	.125	.03 (.03)	.305	.03 (.03)	.307
		Other X Factor	.01 (.03)	.755	.01 (.03)	.775	.00 (.03)	.948
		Pell Eligibility			-.12 (.06)	.044	-.10 (.06)	.087
		High School			.14 (.03)	<.001	.14 (.03)	<.001
		Realistic					.05 (.03)	.124
		Investigative					.00 (.03)	.880
		Artistic					.05 (.03)	.076
		Social					-.03 (.03)	.393
		Enterprising					-.01 (.03)	.681
		Conventional					-.01 (.03)	.702

Note. Effects listed with "X Factor" denote an interaction between particular demographic grouping variable and the standing on the latent factor score.

Table B9.
MIMIC model of the Values scale

Outcome	Item Responses	Predictor	Model 1 β (S.E.)	p	Model 2 β (S.E.)	p	Model 3 β (S.E.)	p
Values Factor		Gender	.23 (.07)	.001	.20 (.07)	.004	.02 (.08)	.779
		Black	-.14 (.10)	.181	-.11 (.11)	.328	-.06 (.12)	.619
		Asian	-.29 (.14)	.031	-.27 (.15)	.066	-.28 (.15)	.052
		Other	-.01 (.11)	.930	.02 (.11)	.830	.03 (.11)	.784
		Pell Eligibility			-.03 (.08)	.740	-.03 (.08)	.732
		High School			-.01 (.04)	.855	.00 (.04)	.997
		Realistic					-.11 (.04)	.010
		Investigative					.05 (.04)	.204
		Artistic					.00 (.04)	.929
		Social					.12 (.04)	.001
		Enterprising					-.07 (.04)	.099
		Conventional					.07 (.04)	.131

Table B9 (cont'd)

Effect on Items

During high school, how many times have you expressed disapproval or anger at a friend for behaving in a manner that you considered to be unethical or wrong?	Never	Gender	-.12 (.05)	.029	-.12 (.06)	.025	-.24 (.07)	<.001
	Once	Black	-.15 (.09)	.098	-.15 (.10)	.142	-.11 (.11)	.332
	Twice	Asian	-.21 (.11)	.056	-.28 (.12)	.016	-.24 (.11)	.037
	Three or four times	Other	-.10 (.09)	.250	-.09 (.09)	.322	-.09 (.10)	.095
	Five times or more	Gender X Factor	-.04 (.03)	.223	-.03 (.03)	.374	-.03 (.03)	.314
		Black X Factor	-.02 (.04)	.566	-.02 (.04)	.648	-.03 (.04)	.375
		Asian X Factor	-.05 (.03)	.089	-.03 (.03)	.265	-.02 (.03)	.522
		Other X Factor	-.02 (.03)	.580	-.01 (.03)	.806	-.01 (.04)	.757
		Pell Eligibility			-.01 (.06)	.858	-.02 (.06)	.813
		High School			.05 (.03)	.069	.05 (.03)	.070
		Realistic					-.08 (.03)	.018
		Investigative					.00 (.03)	.950
		Artistic					.05 (.03)	.072
		Social					.08 (.03)	.012
		Enterprising					-.04 (.03)	.209
		Conventional					.01 (.03)	.718

Tables B9 (cont'd)

(Reversed) Over the past year, how many times were you given detention (or a similar punishment)?	Never Once Twice Three or four times Five times or more	Gender	.25 (.05)	<.001	.26 (.06)	<.001	.33 (.10)	.001
		Black	-.30 (.11)	.006	-.32 (.13)	.010	-.41 (.24)	.095
		Asian	.11 (.09)	.248	.18 (.09)	.038	.22 (.11)	.044
		Other	-.06 (.10)	.520	-.12 (.11)	.282	-.14 (.15)	.337
		Gender X Factor	-.11 (.05)	.031	-.11 (.05)	.033	-.12 (.05)	.033
		Black X Factor	.03 (.08)	.750	.08 (.10)	.460	.12 (.12)	.320
		Asian X Factor	-.05 (.03)	.072	-.07 (.02)	<.001	-.07 (.02)	<.001
		Other X Factor	.01 (.05)	.924	.05 (.06)	.418	.05 (.06)	.390
		Pell Eligibility			-.06 (.06)	.269	-.06 (.06)	.293
		High School			.00 (.03)	.995	.00 (.03)	.957
		Realistic					-.02 (.03)	.514
		Investigative					.02 (.03)	.426
		Artistic					-.02 (.03)	.490
		Social					-.03 (.03)	.336
		Enterprising					-.04 (.03)	.200
		Conventional					.08 (.03)	.015

Table B9 (cont'd)

In your first three years of high school, how often did you skip classes without a legitimate reason?	Most of the time A lot Sometimes Once or twice Never						
		Gender	-.13 (.06)	.016	-.08 (.06)	.130	-.07 (.08)	.413
		Black	-.06 (.09)	.521	.02 (.10)	.854	-.04 (.15)	.800
		Asian	-.10 (.10)	.349	-.07 (.11)	.519	-.03 (.13)	.796
		Other	-.21 (.11)	.057	-.21 (.12)	.095	-.29 (.17)	.084
		Gender X Factor	.06 (.04)	.173	.04 (.04)	.389	.05 (.04)	.295
		Black X Factor	.06 (.04)	.207	.07 (.05)	.177	.07 (.06)	.240
		Asian X Factor	-.04 (.03)	.288	-.06 (.04)	.124	-.06 (.04)	.123
		Other X Factor	.07 (.07)	.267	.09 (.07)	.191	.09 (.07)	.179
		Pell Eligibility			-.18 (.06)	.002	-.18 (.06)	.004
		High School			-.02 (.03)	.551	-.02 (.03)	.605
		Realistic					.03 (.03)	.357
		Investigative					.00 (.03)	.968
		Artistic					.00 (.03)	.985
		Social					-.04 (.03)	.236
		Enterprising					.01 (.03)	.856
		Conventional					-.02 (.03)	.624

Table B9 (cont'd)

How many times have you been accused of acting unethically?	Very often Often Sometimes Rarely Never		Estimate (SE)	p	Estimate (SE)	p	Estimate (SE)	p
		Gender	.33 (.06)	<.001	.36 (.06)	<.001	.40 (.08)	<.001
		Black	-.04 (.09)	.644	-.01 (.10)	.909	-.04 (.13)	.747
		Asian	-.04 (.10)	.647	-.03 (.10)	.741	-.06 (.12)	.634
		Other	-.13 (.10)	.187	-.07 (.10)	.487	-.13 (.13)	.333
		Gender X Factor	-.06 (.04)	.143	-.07 (.04)	.084	-.07 (.04)	.104
		Black X Factor	-.03 (.04)	.508	-.03 (.04)	.551	.00 (.05)	.961
		Asian X Factor	.04 (.04)	.361	.03 (.05)	.562	.02 (.05)	.629
		Other X Factor	.08 (.04)	.039	.07 (.04)	.090	.07 (.04)	.082
		Pell Eligibility			-.06 (.06)	.321	-.05 (.06)	.345
		High School			.04 (.03)	.141	.04 (.03)	.158
		Realistic					.01 (.03)	.744
		Investigative					-.01 (.03)	.762
		Artistic					-.05 (.03)	.070
		Social					-.01 (.03)	.724
		Enterprising					-.01 (.03)	.736
		Conventional					-.03 (.03)	.443

Note. Effects listed with "X Factor" denote an interaction between particular demographic grouping variable and the standing on the latent factor score.

Table B10.
MIMIC model of the Situational Judgment scale

Outcome	Predictor	Model 1 β (S.E.)	p	Model 2 β (S.E.)	p	Model 3 β (S.E.)	p
Situational Judgment Factor	Gender	.49 (.07)	<.001	.52 (.07)	<.001	.33 (.08)	<.001
	Pell Eligibility			.03 (.07)	.638	.04 (.07)	.583
	High School			-.02 (.04)	.538	-.03 (.03)	.357
	Realistic					-.12 (.04)	.002
	Investigative					.07 (.03)	.022
	Artistic					.01 (.04)	.815
	Social					.19 (.04)	<.001
	Enterprising					-.01 (.04)	.793
	Conventional					.05 (.04)	.211
Effect on Items							
A fellow student allows you to listen to threatening phone messages that have been placed on the person's voicemail by another student. The student does not want you to tell anyone, but thinks the caller may be capable of causing physical harm. What would you do?	Gender	.47 (.06)	<.001	.45 (.06)	<.001	.49 (.09)	<.001
	Gender X Factor	-.02 (.03)	.484	-.01 (.04)	.756	-.02 (.04)	.512
	Pell Eligibility			.04 (.06)	.504	.04 (.06)	.397
	High School			-.05 (.03)	.063	-.04 (.03)	.110
	Realistic					-.03 (.03)	.385
	Investigative					.05 (.03)	.076
	Artistic					.02 (.03)	.489
	Social					-.01 (.03)	.643
	Enterprising					-.03 (.03)	.369
	Conventional					.05 (.03)	.106

Table B10 (cont'd)

You are finding a particular class dull and boring, and are having difficulty staying awake. What would you do?							
	Gender	-.06 (.07)	.370	-.07 (.07)	.304	.11 (.13)	.419
	Gender X Factor	-.11 (.05)	.034	-.12 (.06)	.030	.03 (.03)	.317
	Pell Eligibility			.07 (.06)	.205	.09 (.06)	.100
	High School			-.01 (.03)	.744	-.01 (.03)	.757
	Realistic					-.05 (.03)	.113
	Investigative					.05 (.03)	.092
	Artistic					.00 (.03)	.971
	Social					-.06 (.03)	.047
	Enterprising					-.04 (.03)	.266
	Conventional					.05 (.03)	.125

Note. Effects listed with "X Factor" denote an interaction between particular demographic grouping variable and the standing on the latent factor score. For item responses please see Table A10.

REFERENCES

REFERENCES

Aguinis, H., & Smith, M. A. (2007). Understanding the impact of test validity and bias on selection errors and adverse impact in human resource selection. *Personnel Psychology*, 60(1), 165-199.

Armstrong, P. I., Allison, W., & Rounds, J. (2008). Development and initial validation of brief public domain RIASEC marker scales. *Journal of Vocational Behavior*, 73, 287-299.

Bliesener, T. (1996). Methodological moderators in validating biographical data in personnel selection1. *Journal of Occupational and Organizational Psychology*, 69(1), 107-120.

Bobko, P., & Roth, P. L. (2013). Reviewing, categorizing, and analyzing the literature on Black–White mean differences for predictors of job performance: Verifying some perceptions and updating/correcting others. *Personnel Psychology*, 66(1), 91-126.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261-304.

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, 12(4), 381-398.

Cottrell, J. M., Newman, D. A., & Roisman, G. I. (2015). Explaining the Black–White Gap in Cognitive Test Scores: Toward a Theory of Adverse Impact. *Journal of Applied Psychology*, 100(6), 1713-1736.

Dean, M. A. (2013). Examination of ethnic group differential responding on a biodata instrument. *Journal of Applied Social Psychology*, 43(9), 1905-1917.

De Corte, W., Lievens, F., & Sackett, P. R. (2007). Combining predictors to achieve optimal trade-offs between selection quality and adverse impact. *Journal of Applied Psychology*, 92(5), 1380-1393.

Deutsch, M., & Brown, B. (1964). Social Influences in Negro-White Intelligence Differences. *Journal of Social Issues*, 20(2), 24-35.

Drasgow, F. (1987). Study of the measurement bias of two standardized psychological tests. *Journal of Applied Psychology*, 72(1), 19-29.

Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps?. *The Future of Children*, 15(1), 35-54.

Eccles-Parsons, J. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and Achievement Motivations* (pp. 75–121). San Francisco, CA: Freeman.

Fouad, N. A. (1999). Validity evidence for interest inventories. In M. L. Savickas & R. L. Spokane (Eds.), *Vocational interests: Meaning, measurement, and counseling use* (pp. 193–209). Palo Alto, CA: Davis-Black.

Gierl, M. J. (2005). Using a dimensionality-based DIF analysis paradigm to identify and interpret constructs that elicit group differences. *Educational Measurement: Issues and Practices*, 24, 3–14.

Hauser, Robert M., & Goldberger (1971). The treatment of unobservable variables in path analysis. *Sociological Methodology*, 3, 81-117.

Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology*, 6, 35–45.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Holland, J. L., Fritzsche, B., & Powell, A. (1994). *Self-directed search: Technical manual.* Odessa, FL: Psychological Assessment Resources.

Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future--Remembering the past. *Annual Review of Psychology*, 51(1), 631-664.

Hough, L., & Paullin, C. (1994). Construct-oriented scale construction: The rational approach. In Stokes, G. S., Mumford, M. D., Owens, W. A., (Eds.). *Biodata handbook: Theory, Research, and Use of Biographical Information in Selection and Performance Prediction*. (pp. 109-145). Palo Alto, CA: CPP Books.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Sage publications.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96(1), 72.

Imus, A., Schmitt, N., Kim, B., Oswald, F. L., Merritt, S., & Westring, A. F. (2010). Differential item functioning in biodata: Opportunity access as an explanation of gender-and race-related DIF. *Applied Measurement in Education*, 24(1), 71-94.

Jachuck, K., & Mohanty, A. K. (1974). Low socio-economic status and progressive retardation in cognitive skills: A test of cumulative deficit hypothesis. *Indian Journal of Mental Retardation*, 7(1), 36-45.

Jones, K.S., Newman, D. A. Su, R. & Rounds, J. (under review). Vocational Interests and Adverse Impact: A Meta-Analysis of Black-White Differences in Vocational Interests.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351a), 631-639.

Kim, B. H., Schmitt, N., Friede, A., Oswald, F. L., Ramsay, L. J., & Gillespie, M. A. (2004). Differential item functioning in situational judgment tests: Is it a function of the scoring procedure? Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Chicago, Illinois

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions* (pp. 3-90). San Francisco: Jossey-Bass.

Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology*, 9(1), 3-22.

Low, K. D., Yoon, M., Roberts, B. W., & Rounds, J. (2005). The stability of vocational interests from early adolescence to middle adulthood: a quantitative review of longitudinal studies. *Psychological Bulletin*, 131(5), 713-737.

Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology*, 44(4), 763-792.

Mael, F. A., & Ashforth, B. E. (1995). Loyal from day one: Biodata, organizational identification, and turnover among newcomers. *Personnel Psychology*, 48(2), 309-333.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. (2007). Situational judgment tests, response instructions, and validity: a meta-analysis. *Personnel Psychology*, 60(1), 63-91.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 86(4), 730-740.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, 93(3), 568.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13(2), 127-143.

Morris, M. L. (2016). Vocational interests in the United States: Sex, age, ethnicity, and year effects. *Journal of Counseling Psychology*, 63(5), 604.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75(6), 640.

Mumford, M. D., & Owens, W. A. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement*, 11(1), 1-31.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54(4), 557-585.

Muthén, L. K. (2012, August 30). MIMIC Modeling [Msg 17]. Message posted to http://www.statmodel.com/discussion/messages/11/650.html?1454806417

Muthén, L. K., & Muthén, B. O. (2011). *Mplus User's Guide. 2010*. Los Angeles, CA: Muthén & Muthén, 6.

Nye, C. D., & Drasgow, F. (2011). Effect size indices for analyses of measurement equivalence: Understanding the practical importance of differences between groups. *Journal of Applied Psychology*, 96(5), 966.

Nye, C. D., Allemand, M., Gosling, S. D., Potter, J., & Roberts, B. W. (2016). Personality Trait Differences Between Young and Middle-Aged Adults: Measurement Artifacts or Actual Trends?. *Journal of Personality*, 84(4), 473-492.

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance a quantitative summary of over 60 years of research. *Perspectives on Psychological Science*, 7(4), 384-403.

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, 89(2), 187.

Ployhart, R. E. (2006). Staffing in the 21st century: New challenges and strategic opportunities. *Journal of Management*, 32(6), 868-897.

Prasad, J. J., Showler, M. B., Schmitt, N., Ryan, A. M., & Nye, C. D. (2016). Using Biodata and Situational Judgment Inventories across Cultural Groups. *International Journal of Testing*, 1-24.

Raju, N. S., Van der Linden, W. J., & Fleer, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement*, 19(4), 353-368.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111-163.

Robertson, I. T., & Smith, M. (2001). Personnel selection. *Journal of Occupational and Organizational Psychology*, 74(4), 441-472.

Robert, C., Lee, W. C., & Chan, K. Y. (2006). An empirical analysis of measurement equivalence with the INDCOL measure of individualism and collectivism: Implications for valid cross-cultural inference. *Personnel Psychology*, 59(1), 65-99.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23-74.

Schmitt, N., Keeney, J., Oswald, F. L., Pleskac, T. J., Billington, A. Q., Sinha, R., & Zorzie, M. (2009). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology*, 94(6), 1479.

Schmitt, N., & Quinn, A. (2010). Reductions in measured subgroup mean differences: What is possible. *Adverse Impact: Implications for Organizational Staffing and High Stakes Selection*, 425-451.

Schneider, B., & Schmitt, N. (1986). *Staffing Organizations.* Glenview, IL: Scott, Foresman.

Schneider, B. (1987). The people make the place. *Personnel Psychology*, 40(3), 437-453.

Shackleton, V., & Newell, S. (1997). International assessment and selection. In N. Anderson & P. Herriot (Eds.), *International Handbook of Selection and Assessment*. Chichester, UK: Wiley.

Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, 91(1), 25.

Su, R. & Nye, C. D. (in press). Interests and person-environment fit: A new perspective on workforce readiness and success. In J. Burrus, K. D. Mattern, B. Naemi, & R. D. Roberts (Eds.) *Building Better Students: Preparation for the Workforce.*

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: a meta-analysis of sex differences in interests. *Psychological Bulletin*, 135(6), 859.

Tracey, T. J., & Robbins, S. B. (2005). Stability of interests across ethnicity and gender: A longitudinal examination of grades 8 through 12. *Journal of Vocational Behavior,* 67(3), 335-364.

United States Census Bureau (2015). *Income and Poverty in the United States: 2014*. Retrieved from: https://www.census.gov/content/dam/Census/library/publications/2015/demo/p60-252.pdf

United States Census Bureau (2010). *2010 Demographic Profile Data*. Retrieved from: https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml

Van Iddekinge, C. H., Putka, D. J., & Campbell, J. P. (2011). Reconsidering vocational interests for personnel selection: the validity of an interest-based selection test in relation to job knowledge, job performance, and continuance intentions. *Journal of Applied Psychology*, 96(1), 13.

Walker, T. L., & Tracey, T. J. G. (2012). Perceptions of occupational prestige: Differences between African American and White college students. *Journal of Vocational Behavior*, 80, 76-81.

Weekley, J. A., Ployhart, R. E., & Harold, C. M. (2004). Personality and situational judgment tests across applicant and incumbent settings: An examination of validity, measurement, and subgroup differences. *Human Performance*, 17(4), 433-461.

Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, 21(3), 291-309.

Whitney, D. J., & Schmitt, N. (1997). Relationship between culture and responses to biodata employment items. *Journal of Applied Psychology*, 82(1), 113.

Woods, C. M. (2009). Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research*, 44(1), 1-27.

Woods, C. M., & Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement*, 35(5), 339-361.

Woods, C. M., Oltmanns, T. F., & Turkheimer, E. (2009). Illustration of MIMIC-model DIF testing with the Schedule for Nonadaptive and Adaptive Personality. *Journal of psychopathology and behavioral assessment*, 31(4), 320-330.

Zedeck, S. (2010). Adverse impact: History and evolution. *Adverse impact: Implications for organizational staffing and high stakes selection*, 3-27.