Criminal Justice and Behavior

http://cjb.sagepub.com/

Where Does Gender Fit in the Measurement of Self-Control?

Chris L. Gibson, Jeffrey T. Ward, John Paul Wright, Kevin M. Beaver and Matt Delisi

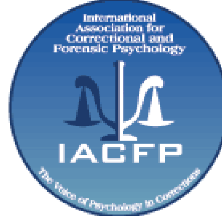
Criminal Justice and Behavior 2010 37: 883 originally published online 19 May 2010
DOI: 10.1177/0093854810369082

The online version of this article can be found at:
http://cjb.sagepub.com/content/37/8/883

Published by:

http://www.sagepublications.com

On behalf of:



International Association for Correctional and Forensic Psychology

Additional services and information for *Criminal Justice and Behavior* can be found at:

Email Alerts: http://cjb.sagepub.com/cgi/alerts

Subscriptions: http://cjb.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://cjb.sagepub.com/content/37/8/883.refs.html

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

WHERE DOES GENDER FIT IN THE MEASUREMENT OF SELF-CONTROL?

CHRIS L. GIBSON
University of Florida
JEFFREY T. WARD
University of Texas at San Antonio
JOHN PAUL WRIGHT
University of Cincinnati
KEVIN M. BEAVER
Florida State University
MATT DELISI
Iowa State University

Gender's role in self-control measures has been largely neglected. Although studies show that males have lower self-control than females, rarely have researchers questioned whether items used to measure self-control should be used for both groups. This study uses a Rasch rating scale analysis to assess item functioning of Grasmick et al.'s 24-item self-control scale for males and females. Using a sample of young adults, results indicate that 33% of the scale items showed differential functioning or item bias; that is, after controlling for self-control, females found one third of the items to be either more or less agreeable than males. Once biased items were removed from the scale, males, on average, still had lower self-control than females. In addition, after excluding biased items from the scale, the effect of self-control on criminal behavior and other outcomes was similar to the effect found with the full 24-item scale. Suggestions for future research on Grasmick et al.'s self-control scale are offered, and limitations of the current study are discussed.

Keywords: self-control; gender; Rasch model; psychometrics; measurement bias

More than two decades have now passed since the publication of Gottfredson and Hirschi's (1990) theory of self-control. Today, self-control theory stands as one of the most empirically supported theories of criminal and delinquent behavior (Pratt & Cullen, 2000), but not all of Gottfredson and Hirschi's propositions have been adequately tested, specifically those related to gender. For instance, Gottfredson and Hirschi state that males possess lower self-control than females—a key proposition from self-control theory. Although the reasons for this difference are likely complicated, studies testing the gender hypothesis still have not shown that measures of self-control are reliable and valid for both males and females (Higgins, 2007; Piquero, MacIntosh, & Hickman, 2000; Piquero & Rosay, 1998).

As it relates to gender, valid measurement of self-control is important because Gottfredson and Hirschi (1990) developed a series of propositions concerning the relationship among gender, self-control, and crime and analogous behaviors. First, they argue that gender differences in delinquent and criminal behaviors appear to be invariant over space and time. Consequently, Gottfredson and Hirschi argue that males have substantially lower self-control than females, which they argue explains the gender difference in criminal behavior. Second,

AUTHORS' NOTE: *Correspondence may be addressed to Chris L. Gibson, Department of Sociology and Criminology & Law, University of Florida, Gainesville, FL 32611-5950; e-mail: clgibson@ufl.edu.*

CRIMINAL JUSTICE AND BEHAVIOR, Vol. 37 No. 8, August 2010 883-903
DOI: 10.1177/0093854810369082
© 2010 International Association for Correctional and Forensic Psychology

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

Gottfredson argues that self-control is the "most important individual-difference cause of crime and delinquency" (Gottfredson, 2006, p. 83); therefore, the effect of gender on these outcomes should be reduced or even nullified after controlling for self-control. Third, the effect of self-control on such behaviors should be present for both males and females. Although several studies have garnered preliminary evidence that supports and refutes these claims (Arneklev, Grasmick, Tittle, & Bursik, 1993; Burton, Cullen, Evans, Alarid, & Dunaway, 1998; LaGrange & Silverman, 1999; Nagin & Paternoster, 1993; Tittle, Ward, & Grasmick, 2003), it is unknown if some of the most commonly used measures of self-control can or should be used unequivocally across genders (Piquero et al., 2000).

Each of the above propositions cannot be investigated with much precision if commonly used measures of self-control consist of items that are biased or function differently across genders. For instance, several studies have found that males, on average, have lower self-control than females (Gibbs, Giever, & Martin, 1998; Keane, Maxim, & Teevan, 1993; Tittle et al., 2003). It could be the case, however, that these observed differences are partially misleading. Gender may interact with a self-control instrument in a way that produces biased measurement across groups (Piquero et al., 2000), possibly resulting in an invalid portrayal of self-control differences between males and females. Items used to measure self-control may not be valid indicators for assessing differences in self-control across genders.

An invalid and/or biased measure of self-control may also have consequences for the status of the empirical validity of self-control theory. For example, some studies have shown that self-control has effects on crime and delinquency for both males and females (Blackwell & Piquero, 2005; Burton et al., 1998; LaGrange & Silverman, 1999; Tittle et al., 2003). Although these findings lend tentative support to Gottfredson and Hirschi's theory, such effects may be misleading if the instrument used to measure self-control is measuring it differently for males and females. For example, consider a male and female who possess similar levels of self-control who are asked to respond to a statement designed to measure low self-control. If the female is less likely to agree with the statement than the male despite having the same self-control, it would indicate differential item functioning (DIF; i.e., item bias). Similarly, if the male is less likely to agree than the female, it would also indicate item bias. DIF is highly problematic when several items from a self-control instrument are biased in the same direction because they can accumulate. In other words, when a test or instrument is biased on the whole, differential test functioning is present. As a consequence, a biased self-control instrument could potentially deflate or inflate the true effect that self-control has on crime and delinquency across the genders. It is particularly important to examine DIF for the Grasmick et al. scale because it has been the most commonly used self-control instrument (Pratt & Cullen, 2000).

GOTTFREDSON AND HIRSCHI'S CONCEPT OF SELF-CONTROL: WHAT ROLE DOES GENDER PLAY?

Although Gottfredson and Hirschi (1990) conclude that low self-control is the most important individual-level cause of crime for all people, they also develop specific propositions concerning self-control and gender. Relying on a long history of empirical evidence that supports gender differences in criminal behavior, they argue that females will have

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

more self-control than males. Past studies support Gottfredson and Hirschi's position by consistently showing that females have higher self-control than males (Gibbs et al., 1998; Keane et al., 1993; Tittle et al., 2003). This finding is supported through both behavioral and attitudinal measures, including Grasmick et al.'s self-control scale.

Gottfredson and Hirschi state that gender differences are largely a function of variations in the application of parental management across male and female children. Accordingly, they argue that parents are more attentive to behaviors of females than they are to the behaviors of males, which they believe results in gender differences in self-control. In addition, other key institutions such as schools play a role in the socialization process, leading to differences in self-control across genders. It is this difference in self-control that can explain why females are less likely to participate in crime and delinquency than males; for both genders, however, Gottfredson and Hirschi argue that self-control is the central mechanism underlying such participation. Research examining Gottfredson and Hirschi's theoretical propositions related to gender is mixed. First, studies show that gender remains an important control variable when testing self-control theory (Arneklev et al., 1993; Grasmick, Tittle, Bursik, & Arneklev, 1993; Piquero & Tibbetts, 1996). When measures of self-control are included in models predicting crime and analogous behaviors, gender remains a significant variable, suggesting that self-control does not completely explain the effect of gender. On the other hand, some studies show that measures of self-control can reduce or completely explain the effect of gender (Burton et al., 1998; LaGrange & Silverman, 1999). Second, additional research shows that low self-control has an effect on both female and male criminal and delinquent behavior (Blackwell & Piquero, 2005; Tittle et al., 2003).

Taken together, previous studies indicate that gender differences in self-control exist, that crime and delinquency are partially a function of self-control for both gender groups, and that self-control can reduce, and in some studies account for, the effect of gender on criminal and delinquent behavior. What is not known, however, is whether measures of self-control should be used unequivocally for both males and females.

THE CREATION OF GRASMICK ET AL.'S SELF-CONTROL INSTRUMENT: WHERE DOES GENDER FIT?

Shortly after publication of Gottfredson and Hirschi's theory, Grasmick and his colleagues created and tested a measure of self-control. In doing so, they derived an operational definition to reflect self-control's theoretical elements. Appendix A lists the items used by Grasmick and his colleagues to operationalize this concept. Grasmick et al. (1993) were also the first to assess the validity of their own self-control instrument. Using exploratory factor analysis, their results led them to conclude that "the strongest case can be made for a one-factor unidimensional model" (Grasmick et al., 1993, p. 17). Using different samples and methodologies, several studies have since produced evidence to both refute and support their findings (see DeLisi, Hochstetler, & Murphy, 2003; Longshore, Turner, & Stein, 1996; Marcus, 2003; Nagin & Paternoster, 1993; Piquero & Rosay, 1998; Piquero & Tibbetts, 1996; Vazsonyi, Pickering, Junger, & Hessing, 2001).

Although they found what appeared to be support for the psychometric properties of their instrument, Grasmick et al. (1993) failed to discuss the instrument's applicability to different groups. They validated their items using a community sample but did not provide researchers

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

with any advice for using the instrument on groups such as college students, juveniles, offenders, or across genders. Consequently, several researchers have used this instrument as a measure of self-control disregarding the potential limitations of its use across various groups.

Although some evidence for the reliability and validity of Grasmick et al.'s self-control scale has been reported (Arneklev, Grasmick, & Bursik, 1999; DeLisi et al., 2003; Gibson, 2005; Longshore et al., 1996; Piquero et al., 2000; Piquero & Rosay, 1998; Vazsonyi et al., 2001), it is not yet clear whether Grasmick et al.'s items should be used to measure self-control for different groups of people. Although Grasmick and his colleagues (1993) advise readers not to accept their own work as the definitive operationalization of self-control, their scale remains the measuring instrument of choice for researchers attempting to quantify this concept for various groups (see DeLisi et al., 2003).

Some research has focused on gender groups when assessing the reliability and validity of Grasmick et al.'s scale and measurement of self-control (Longshore et al., 1996; Piquero & Rosay, 1998; Vazsonyi et al., 2001), but other psychometric studies on Grasmick et al.'s scale have neglected gender comparisons (Arneklev et al., 1999; DeLisi et al., 2003; Grasmick et al., 1993; Marcus, 2003). The ones that have incorporated gender have largely centered on standard reliability estimates such as Cronbach's alpha estimates or internal validity issues such as factor loadings and dimensionality.

Generally, most studies reporting on the reliability of Grasmick et al.'s measure draw supportive conclusions (alpha ranging from .71 to .91; DeLisi et al., 2003; Grasmick et al., 1993; Longshore et al., 1996; Piquero & Rosay, 1998), but less is known about the instrument's reliability across gender. One study found that the scale and subscale reliabilities were better for males (Piquero & Rosay, 1998), but these differences were not substantial. Specifically, Piquero and Rosay (1998) reported gender specific alphas of .72 and .68 for male and female offenders, respectively. Furthermore, they reported gender specific alphas for each subscale (e.g., Impulsivity and Risk Taking), but they were not as encouraging, indicating that subscales have low internal consistency, especially for females.

Several studies have assessed the internal validity of Grasmick et al.'s scale, and more specifically how the internal structure is similar or different across gender groups. For example, Longshore et al. (1996) found that unidimensional and multidimensional models did not work well for female offenders, indicating that female responses across scale items did not covary in a way in which a discernible factor structure emerged. A multidimensional model seemed to work well for males because responses across variables did covary in a way that indicated multiple dimensions of self-control were being measured. Piquero and Rosay (1998) reanalyzed the same data employing item parceling and found that a unidimensional model fit the data well and held for both males and females. Vazsonyi and his colleagues (2001) used a sample of adolescents from four different countries and found that a six-factor model was better than a unidimensional model, and this held for males and females. Regardless of these results, researchers continue to use Grasmick et al.'s scale as if it is measuring a single trait (Piquero, Gomez-Smith, & Langton, 2004).

With two exceptions (see Higgins, 2007; Piquero et al., 2000), studies have not questioned whether Grasmick et al.'s scale items are well suited to measure self-control for males and females. Piquero and colleagues (2000) used a Rasch analysis on data collected from a college sample to assess how levels of self-control influence item responses on the Grasmick et al. scale. More important to the current study, they also assessed how gender influenced

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

item responses to Grasmick et al.'s scale items. Piquero et al. asked the following question and made the following statements:

> Does the Grasmick et al. self-control scale measure the same trait dimensions in the same way when administered to two qualitatively distinct groups, males and females? In effect we ask if test scores administered to individuals belonging to two different populations are comparable on the same measurement scale. The answer to this question is important because if trait scores are not on the same measurement scale across groups, differences among groups in mean levels with regard to external variables may be artificial and misleading. (p. 918)

To address the above question, Piquero and colleagues (2000) conducted a DIF analysis across gender and found that three items exhibited bias after controlling for levels of ability (i.e., self-control). Women in this sample were more likely to say they avoid difficult tasks than would be expected but were less likely to endorse risk over security and hurt the targets of their anger.

Piquero and colleagues used a modified version of the Grasmick et al. scale with a 5-point Likert-type response set prompting Higgins (2007) to conduct a second Rasch rating scale analysis of the Grasmick et al. scale using the original 4-point Likert-type response set. In this analysis, five items were found to function differently for males and females; importantly, four of the five were more difficult for females to endorse (Higgins, 2007), suggesting that differential test functioning may be a problem with the scale. Attending to DIF and item fit statistics, Higgins (2007) concluded that a 16-item modified Grasmick et al. scale both fits the Rasch rating scale model (RSM) and functions similarly across genders.

We believe that these findings are interesting and should be replicated, but, importantly, they raise two unanswered questions regarding gender and Grasmick et al.'s self-control measure that our study attempts to answer. First, are gender differences in self-control (as measured by Grasmick et al.'s scale) substantively meaningful differences, or are they partially or totally a function of gender-biased items? Second, would a gender-neutral self-control measure (excluding biased items) produce different effects on criminal and analogous behavior outcomes across gender groups? In summary, it becomes clear that an agreement has not yet been reached on the psychometric properties and utility of Grasmick et al.'s scale for measuring self-control for males and females. It remains unclear which items of the Gramsick et al.'s scale should be used for both gender groups (Higgins, 2007; Piquero et al., 2000) and, more importantly, what the consequences of using gender-biased items are. As Piquero et al. (2000) stated, understanding this implication is important because "differences among groups in mean levels with regard to external variables may be artificial and misleading" (p. 918). The current study builds on prior empirical work and advances our understanding of the measurement of self-control as well as Gottfredson and Hirschi's theoretical propositions related to gender.

METHOD

PARTICIPANTS

Data used for the current study were taken from a project on the relationship among individual differences, delinquency, and other risky behaviors among college students (see Piquero, Gibson, & Tibbetts, 2002). Data were drawn from students in freshman-level

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

classes attending a university in a southern U.S. state. A complete list of freshman-level classes offered during a fall semester was generated and served as our sampling frame; classes were sequentially numbered and then selected at random into the sample. Students attending class on days of data collection were asked to complete a self-administered survey instrument voluntarily during class time and were informed that responses would be both anonymous and confidential. The original sample consisted of 337 students; however, 4 cases did not have valid information on several of Grasmick et al.'s scale items, resulting in a final sample of 333 students.

Students in our sample closely resembled the demographic characteristics of the larger university, with a somewhat greater quantity of females sampled. Specifically, the analysis sample consists of 61.6% females ($n = 205$) and 38.4% males ($n = 128$). In total, 90.0% of the sample ($n = 300$) was White, whereas the remainder was non-White (i.e., African American, Asian, Hispanic, Native American, and Other). As would be expected in a random sample of freshman-level courses, more than half (55.0%; $n = 183$) of the participants identified themselves as being freshman, 19.5% ($n = 65$) sophomores, 15.3% ($n = 51$) juniors, and 10.2% ($n = 34$) seniors. The age range was 17 to 44 ($M = 20.85$); however, 88% of the students ($n = 293$) reported being 23 years of age or younger.

For obvious reasons, college student samples are often viewed with skepticism and raise several concerns when used in criminological studies (see Arneklev et al., 1999; Piquero et al., 2000). Piquero and colleagues (2000), however, indicate that college student samples are acceptable for some methodologically driven research questions, especially those centered on the measurement of self-control. Furthermore, it would make sense for the current study to use a college sample because it is expanding on two studies using similar samples.

MEASURES

Remaining consistent with past psychometric assessments (Arneklev et al., 1999; Piquero et al., 2000), the Grasmick et al. 24-item attitudinal scale was employed to examine the validity of its items across gender (see Appendix A for item description and Appendix B for item distributional properties). Items have a Likert-type response set of 1 (*strongly disagree*), 2 (*disagree*), 3 (*agree*), and 4 (*strongly agree*). Gender was coded 0 (*male*) or 1 (*female*).

The current study also assessed the effect of Grasmick et al.'s self-control scale on self-reported criminal behavior, which is a dichotomous variable indicating whether participants engaged in at least one of the following six behaviors in the past year: stole something worth less than $50, stole something worth more than $50, participated in a fight, hit someone hard enough that he or she needed bandages or medical attention, purposely damaged or destroyed school property, and purposely damaged or destroyed someone else's property. Approximately 22% of the sample reported engaging in one or more these behaviors. We use age and race (1 = *minority*) as control variables.

ANALYSIS

The analytic strategy for the current study comprised five steps. First, we assessed the reliability of the full scale and each subscale of Grasmick et al.'s instrument across gender groups. Second, we used independent samples *t* tests to assess gender differences on Grasmick et al.'s scale. Based on theoretical expectations, we investigated whether, on average, differences

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

existed on the summated scale and subscale scores for males and females. Third, Rasch RSMs were estimated to help determine whether any of Grasmick et al.'s items exhibit DIF. Fourth, independent samples t test analysis was replicated with any items that showed DIF excluded. If Grasmick et al.'s scale is producing *substantive* self-control differences, we expected to still find these differences when biased items had been excluded. On the other hand, if self-control differences are because of items that exhibit DIF, we expected to find no significant differences—or at least reduced differences—in scale scores across males and females. Finally, we estimated the effect of the biased and unbiased self-control measures on self-reported crime across gender to assess the impact biased scale items have for males and females.

THE RASCH RSM

Given that Rasch models have been infrequently used in criminology (Hickman, Piquero, & Piquero, 2004; Higgins, 2007; Piquero et al., 2000; Piquero, MacIntosh, & Hickman, 2001, 2002; Raudenbush, Johnson, & Sampson, 2003; Ward, Gibson, Boman, & Leite, 2010), it is important to elaborate on this statistical technique and how results from it will inform research on gender and self-control. Rasch models are used to transform nonlinear measures into linear measures by using simple mathematical procedures (see Wright & Masters, 1982, pp. 33-37). In doing so, the Rasch method converts raw scores from test items and persons into log-odds units or logits, thus resulting in two parameters: ability of person n, (B_n), and the difficulty of item i, (D_i). The ability of a person represents his or her level on some latent trait (e.g., self-control) relative to the difficulty of items that compose the measurement instrument. For dichotomous models, item difficulty is defined as the ability level in which a person has a 50% chance of answering an item correctly (Bond & Fox, 2001; B. D. Wright & Masters, 1982). Person ability and item difficulty are placed on the same linear scale consisting of logit units that range from $\pm \infty$ (Bond & Fox, 2001; B. D. Wright & Masters, 1982). This calibration allows the distance between person abilities (high or low self-control) and item difficulties (easy or hard to endorse self-control items) to be articulated and compared with relative ease. Luce and Tukey (1964) refer to this property as additive conjoint measurement.

The RSM is an expansion of the original Rasch model equipped to handle items that have more than two ordered categories (Bond & Fox, 2001; Embretson & Reise, 2000). Given that the Grasmick et al. items have a common four-category response format (*strongly disagree*, *disagree*, etc.), the RSM is an appropriate a priori model choice that has the advantage of parsimony when compared to other more general, direct, polytomous item response models (Embretson & Reise, 2000). Moreover, the RSM has been used in previous psychometric analyses of the Grasmick et al. scale (Higgins, 2007; Piquero et al., 2000), permitting comparisons across studies. This model creates interval-level measures from ordinal-level data using person ability and item difficulty transformations that are similar to the model for dichotomous items.

The RSM has a similar equation to the original Rasch model, with the exception of an additional threshold parameter. The number of thresholds for an item is contingent on the number of categories. For k categories, the number of thresholds equals j $(k - 1)$. Thresholds represent the ability (e.g., self-control) in which two categories are equally likely to be selected. In other words, thresholds mark the place on the ability scale where the next highest category

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

transitions to become the most likely response. For example, for ability levels that are below threshold 1, the most probable response is 1 (*strongly disagree*). Similarly, for ability levels between thresholds 1 and 2, the most probable response is 2 (*disagree*), and so forth. Each threshold parameter is estimated only once across the entire set of items, meaning that the relative distances between the thresholds is common to all items. However, the location of the thresholds on the ability scale for each item is determined by the item difficulty parameter, D_i. The RSM is expressed as follows (Bond & Fox, 2001),

$$Ln\ [P_{nik} / 1 - P_{ni\,(k-1)}] = B_n - D_i - F_k, \tag{1}$$

where the log odds of agreeing or endorsing category k relative to $k - 1$ is determined by the difference between B_n, person ability, and D_i, item difficulty, where F_k is the threshold parameter for category k.

The log odds in Equation 1 can be converted to probabilities using Equation 2 (Bond & Fox, 2001),

$$P_{ni1}(\mathrm{x} = 1 | B_n, D_i, F_1) = \frac{e^{(B_n - D_i - F_1)}}{1 + e^{(B_n - D_i - F_1)}}, \tag{2}$$

where P_{ni1} is the probability of person n choosing Category 2 (*disagree*) over Category 1 (*strongly disagree*) on item i. The difficulty parameter for the first threshold is represented by F_1.

Once a Rasch model is estimated, person abilities and item difficulties are determined and several types of output and post analytic tools are available. In the current study, we used two of these: person–item maps and DIF analysis. As discussed earlier, person abilities and item difficulties are estimated by a Rasch model and then expressed on a common scale (i.e., logit scores ranging from $\pm\infty$). This allows for an examination of item functioning relative to the sample of respondents.

RASCH PERSON–ITEM MAPS

The distribution of item difficulties can be compared to the distribution of person abilities graphically by creating a map in WINSTEPS (Version 3.42). Creating a visual map of the two distributions is important for several reasons. Most importantly, a map can be used to determine the extent to which item difficulties—easy to hard agreeability—match the range of person abilities. Several problems may exist if the distribution of scale items does not resemble the person ability distribution on the logit ruler (Bond & Fox, 2001; B. D. Wright & Masters, 1982). For instance, items could be too difficult for the sample, such that few individuals agree with the items, or items could be too easy for the sample, such that nearly all individuals agree with the items. This can result in an inaccurate depiction of person abilities. This is particularly important for the current study because until now it has been unknown whether the Grasmick et al. scale items are appropriate for both male and female samples. It could be that these items are too easy or difficult to endorse for one or both of these groups.

DIF

DIF is a condition that exists when an item from a test or instrument functions differently from one group to another. In other words, respondents with similar levels of a latent trait,

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

but who belong to different populations, have a different likelihood of agreeing to or endorsing an item. The idea behind item functioning is that scale items should not measure people differently if the measuring instrument is unbiased. Badia, Prieto, and Linacre (2002, p. 889) describe DIF as follows:

> Differential Item Functioning (DIF) investigates the items in a test, one at a time, for signs of interactions with sample characteristics. In the widely used Mantel–Haenszel procedure, reference and focal groups are identified which differ in a discernible way. These groups are stratified into matching ability levels and their relative performance on each item is quantified. The ability levels are usually determined by the total scores on the test. In this way, DIF analysis for one item is as independent as possible of the DIF analyses of the other items.

A conceptual example of DIF can be given using a ruler. A ruler provides a measure of height in inches and is not constrained to measure the height of certain groups; therefore, a ruler would not show differential functioning in measuring height across different people. DIF is used to test this idea when applied to multiple item scales or tests. When applied to these instruments, the question becomes, do items of a scale or test measure people the same or differently depending on the group being measured?

A DIF analysis requires item difficulties to be estimated for males and females separately. Then, comparisons of difficulty estimates for each item are made across groups. If an instrument is unbiased, item functioning should remain invariant across groups; thus, no statistically significant differences in item difficulties should be detected. A standard z score is used when estimating statistical significance of item differences across groups as follows (see Hickman et al., 2004),

$$Z = \frac{(d_1 - d_2)}{(se_1^2 + se_2^2)^{1/2}}, \quad (3)$$

where d indicates item difficulty and se indicates the standard error associated with the difficulty estimate being compared across groups. This test is similar to the Mantel–Haenszel significance test (Badia et al., 2002).

We elected to conduct DIF analysis on the complete Grasmick et al. scale without first assessing the unidimensional structure of the scale for two primary reasons. First, we sought to assess whether observed gender differences in self-control, as typically measured in the literature, were the result of biased items. Second, we wanted to keep the focus of the current study on gender differences in self-control and not on the unidimensional structure of the scale, the latter of which already has a long history in the literature. Nevertheless, we recognize that some may disagree with this strategy. Therefore, in a separate analysis (not shown), we conducted item fit analysis and eliminated nonfitting items in an iterative process and then conducted DIF analysis on the resulting unidimensional scale. The results were substantively similar, and consequently we present only the findings from the analysis of the full scale below.

RESULTS

Table 1 contains Cronbach's reliability coefficients for Grasmick et al.'s scale and subscales across gender groups. Estimates indicate that the 24-item self-control measure has

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

TABLE 1: Reliability Analysis and Independent Samples *t* Tests Assessing Gender Differences in Self-Control

	Males			*Females*		
Variable	M	SD	α	M	SD	α
Self-control	52.13	8.64	.84	48.39	8.72*	.85
Impulsivity	8.12	2.02	.65	7.43	2.07*	.69
Simple tasks	7.90	2.15	.75	7.91	2.07	.77
Risk taking	9.41	2.46	.78	8.35	2.27*	.75
Physical activities	11.09	2.57	.81	10.23	2.51*	.80
Self-centered	7.15	2.24	.77	6.17	1.91*	.65
Temper	8.46	2.68	.79	8.31	2.70	.76

*$p < .05$.

good internal consistency for males and females at .84 and .85, respectively. Furthermore, Table 1 shows gender-specific reliability coefficients for the subscales reflecting each of the six elements of self-control. Overall, each subscale has acceptable reliability for both males and females, and there are relatively minor differences in the subscale reliabilities across gender. More specifically, the reliabilities for the Impulsivity and Simple Tasks subscales are slightly greater for females, whereas the reliabilities of the Risk Taking, Physical Activities, and Temper subscales are slightly larger for males. One exception is the Self-Centered subscale, which has reliability estimates of .77 and .65 for males and females, respectively. In comparison to other studies (Piquero & Rosay, 1998), results from our reliability analyses for subscales are much higher. In sum, these results lend support to the internal consistency of Grasmick et al.'s scale for males and females.

Table 1 also shows results from a series of independent samples *t* tests that were used to investigate initial differences on Grasmick et al.'s self-control measure and its subscales for the male and female samples. Results from these analyses support prior studies that show males, on average, have lower self-control than females. Specifically, our results indicate that males ($M = 52.13$, $SD = 8.64$) score significantly higher on Grasmick et al.'s scale than females ($M = 48.39$, $SD = 8.72$), indicating lower self-control for males. Furthermore, several gender differences are found on subscale scores. Males, on average, were significantly more likely than females to be impulsive, to be risk takers, to be self-centered, and to prefer physical activities. Statistically significant gender differences, however, were not observed for the simple task and temper dimensions. Although these results are consistent with those of previous studies, it is unknown if these initial gender differences are real or because of self-control items that are biased.

Figures 1 and 2 show Rasch person–item maps for the male and female samples, respectively. In both figures, the distribution of person ability estimates (self-control scores) is on the left side of the vertical line and the distribution of item difficulty estimates (item agreeability) is on the right side. The numbers to the far left represent the metric for transformed person and item scores; this logit continuum can be thought of as a ruler that allows researchers to compare person ability and item difficulty estimates in tandem, as they are on the same metric. On the left side of the vertical line, each # represents two males and three females in Figures 1 and 2, respectively, and each . represents one person. Those at the

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

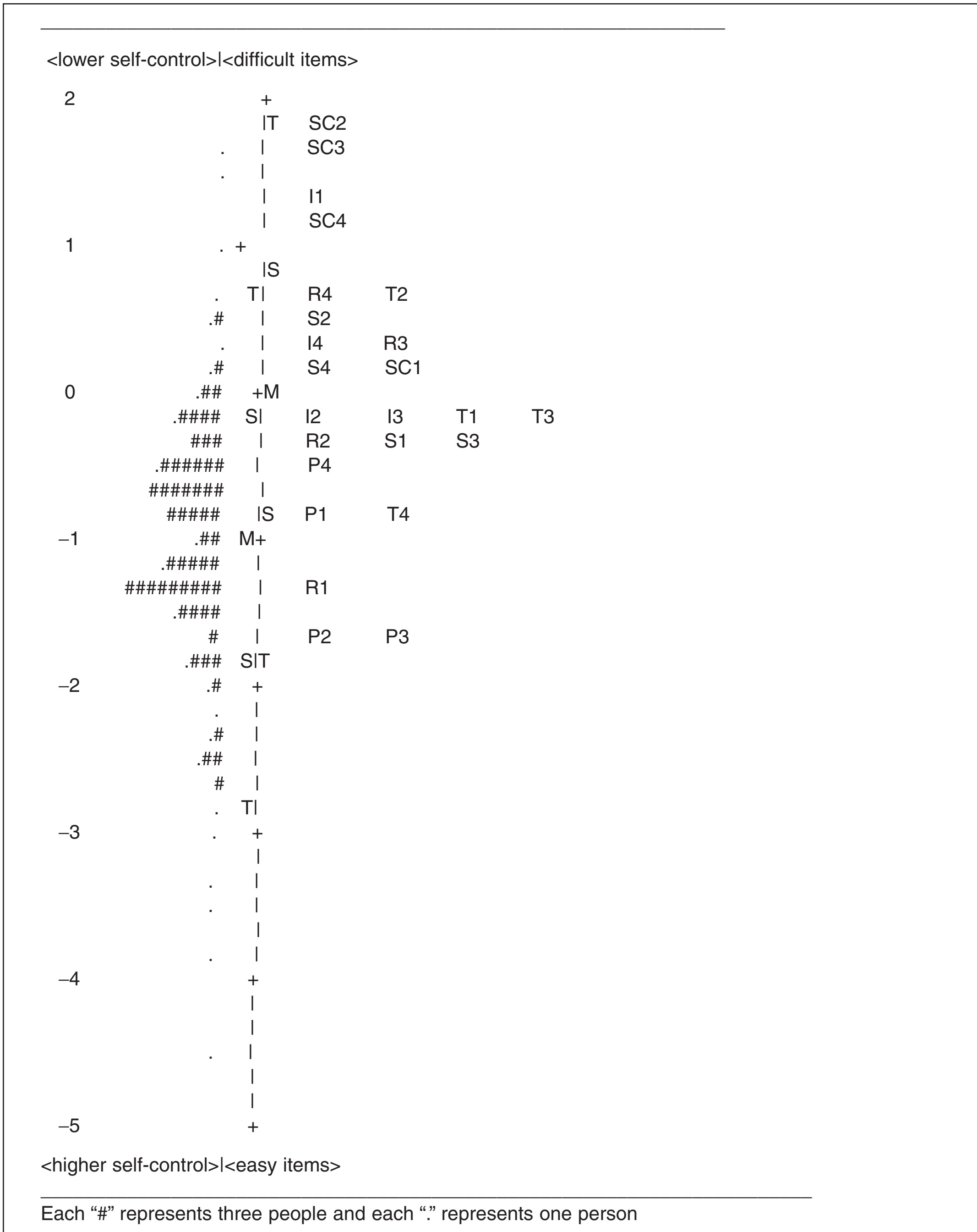


Figure 1: Rasch Person–Item Map for Females

upper end of the scale (i.e., larger positive logit scores) agree on more items and agree more strongly, reflecting lower self-control. Those with more negative logits agree on fewer items and agree less strongly, reflecting higher self-control. Items are shown on the left side

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

```
<lower self-control>|<difficult items>
 2              +
                |
                |
                |T
              . |   SC2
                |   S2
 1             T+   SC3
            .#  |S  I1        SC4
            ##  |
             .  |   S4        T2
            .#  |   R3
         ### S|     I2        I4        S1       SC1       T3
 0       ###    +M  R4        T1
   .#########   |   I3        S3        T4
        ####    |
      .######   |
       #####  M|    R2
      ######    |S  P4
-1    .#####    +   P1
       #####    |   R1
          .#    |   P2
         .##  S|T   P3
           #    |
          .#    |
-2         #    +
           #    |
              T|
           .    |
                |
                |
-3         .    +
                |
                |
           .    |
                |
                |
-4              +
           .    |
                |
                |
                |
                |
-5              +
<higher self-control>|<easy items>
Each "#" represents two people and each "." represents one person.
```

Figure 2: Rasch Person–Item Map for Males

of the vertical line. Items at the upper end of the scale (i.e., larger positive logit scores) are more difficult to endorse, whereas items at the lower end of the scale are easier to endorse. Several letters are reported on the vertical line that divides persons and items. *M* marks the

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

person and item means (average logit scores for items and persons), *S* is one standard deviation from the mean, and *T* is two standard deviations from the mean.

If Grasmick et al.'s self-control items were too easy for either gender, items would be expected to have a mean logit score substantially lower than the mean logit score for the person ability estimates. Figures 1 and 2, however, clearly show that this is not the case. In turn, if items were too hard to endorse for either sample, items would be expected to have a mean logit score substantially higher than the mean logit score for person ability estimates. Figures 1 and 2 support the latter scenario more closely, in that male and female ability distributions are, on average, lower than the item difficulty distributions. However, this is not a worst-case scenario; there is still substantial overlap between the item and person distributions, suggesting that the items target the sample reasonably well. For males and females, the average person ability estimates are approximately one logit lower than the average item difficulty estimates.

Although males and females appear to be somewhat similar in regard to person–item distributions, some important preliminary gender differences become apparent with a closer comparison of the two maps. For instance, females, on average, have lower person ability estimates than males, and several items appear to be more difficult for females to agree to. For example, Sc2, Sc3, I1, and R4 appear to be more difficult for females to agree with than males. Although informative, results presented in Figures 1 and 2 are descriptive in that no statistical comparisons are made regarding item difficulties between groups.

Table 2 shows results from a DIF analysis of Grasmick et al.'s 24 items. Importantly, results presented here allow us to determine if any scale items are biased or show differential functioning across our male and female samples. The first column lists items, the second and third columns show the estimated item difficulties for male and female samples, and the fourth column shows *z* values. Similar to a *t* test, *z* values that are above 1.96 in absolute value are statistically significant ($p < .05$), indicating that an item significantly differs in difficulty across samples. Table 2 shows that 8 (33%) of Grasmick et al.'s 24 items exhibit bias, and these items are from each of the scale's six domains. Thus, one third of the items significantly vary in their difficulty levels or agreeability for males and females when controlling for overall levels of self-control.

Females find five of the eight items more difficult to endorse than males. Specifically, women are more likely than men to devote time to preparing for the future (I1), women are less likely to agree that excitement and adventure are more important than security (R4), women are less likely to have a greater need for physical activity (P4), women are more likely to be sympathetic to others when others are having problems (Sc2), and women are more likely to care if the things they do upset people (Sc3). Males find three items significantly more difficult to endorse (or less agreeable) than females. Specifically, males are less likely than females to avoid projects that are difficult (S1), they are less likely to quit or withdraw when things get complicated (S2), and males are less likely than females to agree that, when in a serious disagreement, it is hard to talk calmly without getting upset (T4). Importantly, these differences exist after controlling for males and females overall levels of self-control. Ultimately, results from Table 2 show that bias across gender groups exists for several of the items making up Grasmick et al.'s self-control instrument.

Table 3 shows independent samples *t* tests comparing male and female scores on Grasmick et al.'s scale and subscales. These analyses exclude the eight items that exhibited DIF and are noteworthy for several reasons. First, statistically significant differences across gender

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

TABLE 2: Differential Item Function Analysis for the Grasmick et al. Scale Items Across Gender

	Female Sample		*Male Sample*		
Item	*Measure*		*Measure*		*z Test*
I1	1.35	(0.13)	0.83	(0.14)	2.76*
I2	–0.16	(0.10)	0.15	(0.13)	–1.91
I3	–0.12	(0.10)	–0.15	(0.13)	0.21
I4	–0.30	(0.11)	0.23	(0.13)	0.43
S1	–0.28	(0.10)	0.12	(0.13)	–2.41*
S2	0.55	(0.11)	1.11	(0.14)	–3.09*
S3	–0.35	(0.10)	–0.12	(0.13)	–1.39
S4	0.23	(0.11)	0.47	(0.13)	–0.13
R1	–1.25	(0.10)	–1.23	(0.12)	1.78
R2	0.39	(0.10)	–0.65	(0.12)	1.63
R3	0.34	(0.11)	0.30	(0.13)	0.23
R4	0.61	(0.11)	0.00	(0.13)	3.62*
P1	–0.84	(0.10)	–1.06	(0.12)	1.40
P2	–1.61	(0.10)	–1.38	(0.12)	–1.45
P3	–1.62	(0.10)	–1.52	(0.12)	–0.62
P4	–0.52	(0.10)	–0.89	(0.12)	2.35*
Sc1	0.14	(0.10)	0.18	(0.13)	–0.28
Sc2	1.79	(0.14)	1.33	(0.15)	2.29*
Sc3	1.72	(0.14)	0.91	(0.14)	4.10*
Sc4	1.17	(0.12)	0.87	(0.14)	1.59
T1	–0.16	(0.10)	–0.06	(0.13)	–0.63
T2	0.63	(0.11)	0.45	(0.13)	1.00
T3	–0.23	(0.10)	0.08	(0.13)	–1.96
T4	–0.89	(0.10)	–0.17	(0.13)	–4.45*

Note. Standard errors of item estimates are in parentheses.
$^{*}p < .05$.

TABLE 3: Independent Samples *t* Tests Assessing Gender Differences in Self-Control: Examination After Exclusion of Biased items

	Males		*Females*	
Variable	M	SD	M	SD
Self-control (16-item scale)	36.28	6.17	33.95	6.43*
Impulsivity	6.34	1.63	5.96	1.75*
Simple tasks	4.14	1.25	4.04	1.18
Risk taking	7.27	1.82	6.62	1.88*
Physical activities	8.48	2.07	8.00	2.03*
Self-centered	3.82	1.26	3.45	1.24*
Temper	6.23	2.06	5.88	2.11

$^{*}p < .05$.

remain after excluding the eight items that are biased. Similar to our earlier estimates, males are more likely to have lower self-control than females. In addition, males are more likely than females to be impulsive, to be risk takers, to be self-centered, and to prefer physical activities. These results show that gender differences in self-control scores are not a function of biased items, as significant differences are still found when the biased items are excluded. Second, although these results show that gender differences remain after excluding

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

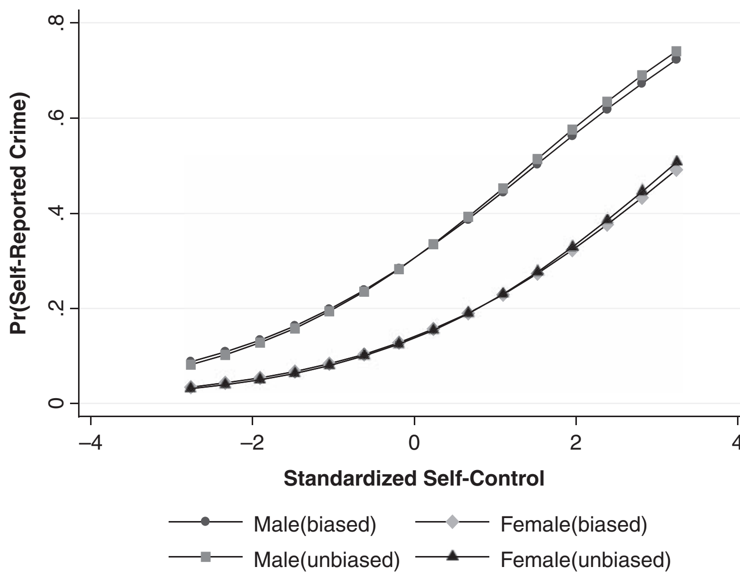


Figure 3: Predicted Probabilities of Self-Reported Crime: Effects of Unbiased and Biased Measures of Self-Control Across Gender

biased items, the gender difference on overall scale scores has been modestly reduced. Specifically, the gender difference in self-control as measured by the full 24-item scale is 3.74 (0.156 per item), and the gender difference as measured by the 16-item modified scale with no DIF items is 2.33 (0.146 per item). Thus, slightly more than 6% of the observed gender difference in self-control is because of biased items.

Finally, Figure 3 shows a graph representing the probability of self-reported crime as a function of standardized self-control scores, with more positive scores indicating lower self-control. For comparison purposes, and after making adjustments for race and age, we plotted the probability of self-reported crime for both males and females across levels of self-control. Moreover, the effects of both the full Grasmick et al. self-control measure (including all 24 items) and the reduced measure (excluding biased items and retaining unbiased items) are shown.

Earlier we discussed the possibility that a gender-biased self-control measure may result in drawing misleading conclusions about the effect of self-control on criminal outcomes across gender groups. According to the results in Figure 3, this does not appear to be the case. Instead, for both males and females, and regardless of the self-control measure used (biased or unbiased), the probability of self-report crime increased for those with lower self-control. The predicted probability curves for males are practically the same when using the biased or unbiased self-control measure, and the same holds true for females. It is worth noting that the effect of gender on the probability of committing a criminal act is relatively small for individuals with high self-control (i.e., very low scores on the low self-control continuum) but gets progressively larger as self-control decreases (i.e., as scores get higher), irrespective of the measure used. The same pattern of results (not shown) holds true for two other outcome measures: self-reported binge drinking and test cheating. Similar to past studies (Piquero et al., 2002), binge drinking was coded 0 (*average of four or fewer drinks*

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

at one sitting) or 1 (*average of five or more drinks at one sitting*). Test cheating was measured using a dichotomous item asking participants to report if they have cheated on a school test in the past year.

DISCUSSION AND CONCLUSION

This study investigated gender differences in self-control using one of the most common measures of Gottfredson and Hirschi's concept of self-control. Our study is the first to thoroughly assess how gender differences in self-control are possibly a function of measurement bias and what consequences, if any, are associated with using a gender-biased self-control measure to predict theoretically derived outcomes such as crime. Several findings from this study are noteworthy. First, we showed that Grasmick et al.'s scale and subscales are quite reliable for both males and females, a finding that is not completely consistent with those of past studies (Piquero & Rosay, 1998). Second, our findings are consistent with the hypothesis that males have lower self-control than females. Similar to the findings of Higgins (2007) and Piquero and colleagues (2000), results from a Rasch model showed that a number of items were functioning quite differently for males and females. When the eight biased items were excluded from the scale, the gender difference in self-control was reduced by approximately 6%, but significant differences remained. Moreover, the two measures (biased vs. unbiased) did not have differential effects on males' and females' self-reported crime, binge drinking, or cheating behavior.

Our analyses shed light on self-control differences for males and females that prior studies have not considered. We showed that one third of Grasmick et al.'s scale items function differently across gender groups when controlling for underlying levels of self-control, which is five items more than Piquero and colleagues (2000) found to be biased and three more than Higgins (2007) found in his analysis. Four of the eight items that were identified as exhibiting DIF in our analysis also exhibited DIF in one or both of the previous analyses. The good news is that gender differences, as indicated by average scale scores, are not because of measurement bias and most likely reflect important substantive differences. With such a large percentage of items exhibiting DIF (33%), it may be somewhat surprising that only a small percentage of the gender difference in self-control was attributed to measurement bias (6%). This is explained by the fact that because the items failed to be systematically harder for one group to endorse than the other, measurement bias did not accrue across the items and, therefore, did not substantially inflate or deflate observed gender differences in self-control. One should be cognizant of the fact that DIF can indeed contribute to nontrivial biases in the estimation of group differences, albeit this was not found to be the case in the present study. The bad news, however, is that several of Grasmick et al.'s items are not well suited for both males and females, as eight of them exhibited significant variation in difficulty between males and females after accounting for overall self-control levels. These findings naturally beg the question, why do a sizeable proportion of the Grasmick et al. items exhibit DIF across genders?

According to Piquero et al. (2000), gender differences in item functioning could reflect gender role expectations net of levels of self-control. For instance, women were found to be less likely to agree that excitement and adventure are more important than security. Differences between the social expectations of females and males across the life course could potentially

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

explain why women respond differently on this item despite having comparable levels of self-control. As an alternative explanation, gender differences in cognition could also play a role in explaining DIF, and the source could be environmental, biological, or both. Research suggests that females and males have some fundamental differences in brain processes and styles of thinking (Wright, Tibbetts, & Daigle, 2008), which could potentially influence item meaning and the resulting responses. It is critical that future research attempts to identify the sources of DIF in addition to isolating the prevalence and implications for hypotheses regarding gender differences in self-control and offending—the latter of which was the focus on the current study. There is great need for both qualitative investigations on gender role expectations and research on gender differences in brain processing as it relates to DIF and measurement. The items that have been shown to be biased in our study and prior work should serve as a useful starting point. Ultimately, a better understanding of the specific causes of DIF across genders will result in measures that can be developed or altered to explicitly deal with biased measurement up front.

Although we feel that our findings raise important issues regarding the measurement of self-control across genders, our study is not without limitations. One limitation is that a sample of male and female college students attending a university in the southern United States during the late 1990s was used to investigate gender differences in self-control. Male and female college student samples have often been frowned on when used to test criminological theories. Although we acknowledge this concern, our research intentions emphasized a methodological issue. As such, some would argue that using college student samples are suitable for exploring the type of questions proposed in this study (Piquero et al., 2000).

Another limitation is that our findings are possibly a function of the unique characteristics of our sample and, therefore, caution should be taken when attempting to make generalizations about gender differences and self-control. For instance, the significant differences found in item difficulties across gender could be smaller or more pronounced in male and female samples that possess different characteristics. Perhaps a period effect could be present, given that the data used for this study were collected several years ago. Current male and female college students may vary in ways that are different from past college students, and this could influence how Grasmick et al.'s items are perceived or interpreted across gender. Furthermore, females and males that are older or younger than those in our sample may also perceive Grasmick et al.'s items differently, resulting in more or fewer items showing statistically significant item bias or larger or smaller differences in item difficulties across groups. Consequently, researchers who continue to explore gender and self-control or gender issues related to Grasmick et al.'s scale should obtain data on males and females that can address the concerns raised above.

Although researchers have used a variety of self-control measures, the current study analyzed data that were collected using the most common self-control instrument for testing self-control theory (see Pratt & Cullen, 2000). On this note, researchers may want to assess gender differences and item bias for other known measures of self-control (e.g., see Marcus, 2003). Marcus's (2003) self-control instrument is a relatively new self-control measure that has several attractive qualities but has been subjected to limited psychometric scrutiny (Ward et al., 2010). In addition, we hope our findings spark more research interest in the measurement of self-control, with a particular focus on whether self-control measures can be equally valid across groups of participants that are different with respect to demographic and personal characteristics.

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

In conclusion, we have laid out several limitations of our study and a modest research agenda for those interested in exploring measurement issues related to self-control. Beyond the current study, we believe measurement is one of the most fundamental concerns in criminology as well as in other social sciences. Although having extended previous empirical work on gender and self-control as well as what we know about one of the most widely used measures of self-control, our work opens a larger question that is often neglected in criminology. That is, do we need measures that are gender sensitive to test criminological theories, or can we use the same measures for men and women? Theory and research both question the need for gender-specific versus general explanations of male and female antisocial and offending behaviors. Perhaps time has also come to be more sensitive to the measures we use in criminology when examining differences between males and females. For some time now creators of standardized tests have gone to lengths to create questions and items that are free from racial and gender bias so that average differences between groups are closer to true differences and cannot be easily dismissed as a function of a biased test. We feel that similar investigations are important for understanding bias in criminological measures, and we encourage researchers to explore such issues.

APPENDIX A
SELF-CONTROL ITEMS

Impulsivity component
- I1: I do not devote time and effort to preparing for the future
- I2: I act on the spur of the moment without stopping to think
- I3: I do things that bring me pleasure here and now, even at the cost of future goals
- I4: I base my decisions on what will happen to me in the short run rather than the long run

Simple Tasks component
- S1: I try to avoid projects that I know will be difficult
- S2: When things get complicated, I quit or withdraw
- S3: I do things in life which are the easiest and bring me the most pleasure
- S4: I avoid difficult tasks that stretch me abilities to the limit

Risk Taking component
- R1: I test myself by doing things that are a little risky
- R2: I take risks just for the fun of it
- R3: I find it exciting to do thing for which I might get in trouble
- R4: Excitement and adventure are more important to me than security

Physical Activities component
- P1: If I have a choice, I will do something physical rather than something mental
- P2: I feel better when I am on the move than when I am sitting and thinking
- P3: I'd rather go out and do things than read or contemplate ideas
- P4: Compared to other people my age, I have a greater need for physical activity

Self-Centered component
- Sc1: I look out for myself first, even if it makes things difficult for other people
- Sc2: I am not sympathetic to other people when they are having problems
- Sc3: I don't care if things I do upset people
- Sc4: I will try to get things I want even when I know it's causing problems for other people

Temper component
- T1: I lose my temper easily
- T2: When I am angry at people I feel more like hurting them than talking to them about why I am angry
- T3: When I am really angry, other people better stay away from me
- T4: When I have a serious disagreement with someone, it's usually hard for me to talk calmly about it without getting upset

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

APPENDIX B

DESCRIPTIVE STATISTICS FOR LOW SELF-CONTROL ITEMS ACROSS MALE AND FEMALE SAMPLES

	Full Sample (N = 333)		*Males (n = 128)*		*Females (n = 205)*	
Item[a]	M	SD	M	SD	M	SD
Impulsivity						
I1	1.59	0.70	1.77	0.74	1.47	0.64
I2	2.07	0.72	2.08	0.74	2.06	0.71
I3	2.11	0.77	2.23	0.72	2.04	0.79
I4	1.92	0.72	2.04	0.69	1.85	0.73
Simple Tasks						
S1	2.11	0.69	2.09	0.70	2.12	0.69
S2	1.72	0.62	1.66	0.63	1.75	0.61
S3	2.17	0.74	2.21	0.73	2.15	0.74
S4	1.90	0.70	1.93	0.73	1.89	0.69
Risk Taking						
R1	2.68	0.67	2.78	0.65	2.61	0.68
R2	2.29	0.81	2.48	0.80	2.17	0.80
R3	1.90	0.82	2.01	0.85	1.84	0.80
R4	1.89	0.80	2.15	0.86	1.73	0.71
Physical Activities						
P1	2.51	0.86	2.70	0.86	2.40	0.85
P2	2.82	0.76	2.86	0.78	2.80	0.75
P3	2.85	0.83	2.93	0.80	2.80	0.84
P4	2.38	0.76	2.60	0.79	2.23	0.72
Self-Centered						
Sc1	1.98	0.79	2.06	0.79	1.93	0.79
Sc2	1.44	0.66	1.59	0.66	1.35	0.64
Sc3	1.51	0.70	1.74	0.78	1.36	0.61
Sc4	1.62	0.69	1.76	0.68	1.53	0.68
Temper						
T1	2.10	0.91	2.18	0.90	2.06	0.91
T2	1.80	0.80	1.94	0.76	1.72	0.81
T3	2.10	0.93	2.11	0.91	2.10	0.95
T4	2.35	0.88	2.23	0.86	2.42	0.89

a. All items have Likert-type response sets on a 4-point scale of 1 (*strongly disagree*), 2 (*disagree*), 3 (*agree*), and 4 (*strongly agree*).

REFERENCES

Arneklev, B. J., Grasmick, H. G., & Bursik, R. J., Jr. (1999). Evaluating the dimensionality and invariance of “low self-control.” J*ournal of Quantitative Criminology, 15,* 307-331.

Arneklev, B. J., Grasmick, H. G., Tittle, C. R., & Bursik, R. J. (1993). Low self-control and imprudent behaviors. *Journal of Quantitative Criminology, 9*, 225-247.

Badia, X., Prieto, L., & Linacre, J. M. (2002). Differential item and test functioning. *Rasch Measurement Transactions, 16*, 889.

Blackwell, B. S., & Piquero, A. R. (2005). On the relationships between gender, power control, self-control, and crime. *Journal of Criminal Justice, 33*, 1-17.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences.* Mahwah, NJ: Lawrence Erlbaum.

Burton, V. S., Cullen, F. T., Evans, D. T., Alarid, L. F., & Dunaway, G. R. (1998). Gender, self-control, and crime. *Journal of Research in Crime and Delinquency, 35*, 123-147.

DeLisi, M., Hochstetler, A., & Murphy, D. S. (2003). Self-control behind bars: A validation of the Grasmick et al. scale. *Justice Quarterly, 20*, 241-263.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum.

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

Gibbs, J. J., Giever, D., & Martin, J. S. (1998). Parental management and self-control: An empirical test of Gottfredson and Hirschi's general theory. *Journal of Research in Crime and Delinquency, 35*, 40-70.

Gibson, C. L. (2005). *A psychometric investigation of a self-control scale: The reliability and validity of Grasmick et al.'s scale for a sample of incarcerated male offenders* (Unpublished doctoral dissertation). University of Nebraska at Omaha.

Gottfredson, M. R. (2006). The empirical status of control theory in criminology. In F. T. Cullen, J. P. Wright, & K. R. Blevins (Eds.), *Taking stock: The status of criminological theory (*pp. 77-100). New Brunswick, NJ: Transaction Publishing.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.

Grasmick, H. G., Tittle, C. R., Bursik, R. J., & Arneklev, B. J. (1993). Testing the core empirical implications of Gottfredson and Hirschi's general theory of crime. *Journal of Research in Crime and Delinquency, 30*, 5-29.

Hickman, M. J., Piquero, N. L., & Piquero, A. R. (2004). The validity of Niederhoffer's cynicism scale. *Journal of Criminal Justice, 32*, 1-13.

Higgins, G. (2007). Examining the original Grasmick scale: A Rasch model approach. *Criminal Justice and Behavior, 34*, 157-178.

Keane, C., Maxim, P. S., & Teevan, J. J. (1993). Drinking and driving, self-control, and gender: Testing the general theory of crime. *Journal of Research in Crime and Delinquency, 30*, 30-46.

LaGrange, T. C., & Silverman, R. A. (1999). Low self-control and opportunity: Testing the general theory of crime as an explanation for gender differences in delinquency. *Criminology, 37*, 41-72.

Longshore, D., Turner, S., & Stein, J. A. (1996). Self-control in a criminal sample: An examination of construct validity. *Criminology, 34*, 209-228.

Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new kind of fundamental measurement. *Journal of Mathematical Psychology, 1*, 1-27.

Marcus, B. (2003). An empirical examination of the construct validity of two alternative self-control measures. *Educational and Psychological Measurement, 63*, 674-706.

Nagin, D. S., & Paternoster, R. (1993). Enduring individual differences and rational choice theories of crime. *Law and Society Review, 27*, 467-496.

Piquero, A. R., Gibson, C., & Tibbetts, S. (2002). Does low self-control account for the relationship between binge-drinking and alcohol-related behaviors? Comparison across gender. *Criminal Behavior and Mental Health, 12*, 135-154.

Piquero, A. R., Gomez-Smith, Z., & Langton, L. (2004). Discerning unfairness where others may not: Low self-control and unfair sanction perceptions. *Criminology, 42,* 699-744.

Piquero, A. R., MacIntosh, R., & Hickman, M. (2000). Does self-control affect survey response? Applying exploratory, confirmatory, and item response theory analysis to Grasmick et al.'s self-control scale. *Criminology, 38*, 897-929.

Piquero, A. R., MacIntosh, R., & Hickman, M. (2001). Applying Rasch modeling to the validity of a control balance scale. *Journal of Criminal Justice, 29*, 493-505.

Piquero, A. R., MacIntosh, R., & Hickman, M. (2002). The validity of a self-report delinquency scale: Comparison across gender, age, race. *Sociological Methods and Research, 30*, 492-529.

Piquero, A. R., & Rosay, A. B. (1998). The reliability and validity of Grasmick et al.'s self-control scale: A comment on Longshore et al. *Criminology, 36*, 157-173.

Piquero, A. R., & Tibbetts, S. (1996). Specifying the direct and indirect effects of low self-control and situational factors in offender's decision making. *Justice Quarterly, 13*, 481-510.

Pratt, T. C., & Cullen, F. T. (2000). The empirical status of Gottfredson and Hirschi's general theory of crime: A meta-analysis. *Criminology, 38*, 931-964.

Raudenbush, S., Johnson, C., & Sampson. R. J. (2003). A multivariate, multilevel Rasch model with application to self-report criminal behavior. *Sociological Methodology, 33*, 169-211.

Tittle, C. R., Ward, D. A., & Grasmick, H. G. (2003). Self-control and crime/deviance: Cognitive vs. behavioral measures. *Journal of Quantitative Criminology, 19*, 333-365.

Vazsonyi, A. T., Pickering, L. E., Junger, M., & Hessing, D. (2001). An empirical test of a general theory of crime: A four-nation comparative study of self-control and the prediction of deviance. *Journal of Research in Crime and Delinquency, 38*, 91-131.

Ward, J. T., Gibson, C. L., Boman, J., & Leite, W. (2010). Assessing the validity of the Retrospective Behavioral Self-control Scale: Is the general theory of crime stronger than the evidence suggests? *Criminal Justice and Behavior*, *37*, 336-357.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis.* Chicago, IL: MESA Press.

Wright, J. P., Tibbetts, S. G., & Daigle, L. E. (2008). *Criminal in the making: Criminality across the life course.* Thousand Oaks, CA: Sage.

Chris L. Gibson is an assistant professor in the Department of Sociology and Criminology & Law at the University of Florida and a W. E. B. Du Bois Fellow of the National Institute of Justice. His research interests include neighborhood influences, life-course and biosocial criminology, and applied quantitative methods.

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010

Jeffrey T. Ward is an Assistant Professor in the Department of Criminal Justice at the University of Texas at San Antonio and a Harry Frank Guggenheim Dissertation Fellow. His research interests include quantitative methodology, sanction effects, development and life course, and space, place, and crime.

John Paul Wright is an associate professor of criminal justice in the School of Criminal Justice at the University of Cincinnati. His work focuses on the biological linkages to violence. He has published widely in leading scientific journals.

Kevin M. Beaver is an assistant professor in the College of Criminology and Criminal Justice at Florida State University. His research examines the biosocial correlates to antisocial outcomes.

Matt DeLisi is coordinator of criminal justice studies, faculty affiliate with the Center for the Study of Violence, and associate professor in the Department of Sociology at Iowa State University. His primary research interests include criminal careers and career criminals, self-control theory, corrections, psychopathy, and the molecular and behavioral genetics of antisocial behavior.

Downloaded from cjb.sagepub.com at UNIV OF FLORIDA Smathers Libraries on July 7, 2010