Social Desirability Bias among Prejudice Instruments
An Integrated Threat

Stephen M Croucher, Elvis Nshom, Djayko Rahmani & Cheng Zeng
Massey University, California State University, North Dakota State University

Abstract
A considerable amount of research has examined the extent to which members of dominant cultures perceive minority groups as threatening their way of life. While various instruments measure these perceptions of threat, few researchers have empirically analysed the statistical properties of these scales. Specifically, studies have not adequately explored social desirability of threat scales. The current study investigates the extent to which one set of threat scales is internally consistent or reliable (González et al., 2008), and explores social desirability within the González et al. (2008) integrated threat instruments by comparing self-reports to other reports (intimate other and friend reports). Results indicate the instruments are internally consistent and that self-reports and other reports of threat do not differ on most indices of threat.

Keywords: Integrated threat theory, prejudice, measurement, reliability analysis

Introduction
There is considerable research conceptualizing and operationalizing prejudice. Multiple conceptual definitions exist for prejudice. For example, Ashmore (1970) defined prejudice as “a negative attitude toward a socially defined group and toward any person perceived to be a member of that group” (p. 253). Allport (1954) conceptualized prejudice as “an aversive or hostile attitude towards a person who belongs to a group simply because he belongs to that group, and is therefore presumed to have objectionable qualities ascribed to the group” (p. 7). Aside from these two definitions many other definitions are commonly used by researchers (see Oskamp 2000; Rus & Madrid 1998; Scheibner & Morrison 2009 for reviews). However, in this study, prejudice is conceptualized according to Allport (1954, p. 7).

The study of prejudice has been, and remains, a sociocultural, economic, political, and scholarly concern for most of the 20th and 21st centuries (Burkard, Medler, & Boticki, 2001) especially in Finnish society where this study was carried out. As the world has witnessed the rise of prejudice in many forms, designing reliable and valid measures to understand and assess these negative attitudes has become an imperative for researchers, policy makers, clinicians, and other stakeholders (Burkard, Boticki, & Madison 2002; Robinson, Shaver, & Wrightsman 1991). Researchers studying prejudice have a plethora of measures and methods to assess prejudicial attitudes. The methodological approach taken in measuring prejudice depends largely on how prejudice is conceptualized and the level/type of measurement researchers want in a study. No matter what approach is taken, one common critique of prejudice measures is social desirability, or the tendency of respondents to give answers to make themselves look good (Burkhard et al. 2001; Paulhus 1991).

In the current study one conceptualization and operationalization of prejudice is analyzed, integrated threat theory (ITT) (Stephan & Stephan 1996; Stephan, Ybarra, Riess Morrison 2015). According to the integrated threat theory, four types of perceived threat may predict prejudice towards outgroups. These include realistic threat, symbolic threat, negative stereotype and intergroup anxiety. Realistic threats are threats to the economic wellbeing and political power of the ingroup. Symbolic threats are fears due to differences between the ingroup and the outgroup. These differences could be differences in religion, culture, values, morals, worldview just to name a few. Negative stereotypes are those fears that arise because of negative stereotypical beliefs the ingroup has about an outgroup. In the process of interaction, they may fear that those negative consequences will befall them. Lastly, intergroup anxiety refers to the anxiety ingroup members may experience when interacting with outgroup members especially if both groups have a history of conflict. Since intergroup anxiety is not considered a group level threat, it is not included in this study. Thus, the purpose of this study is to evaluate the reliability or stability of self-reports of prejudice/threat by comparing self-reported prejudice/threat with two other reports of the same target: evaluations by friends and evaluations by intimates. By intimate other, we are referring to someone the main respondent (self-reporter) considers their close or best friend.

In particular, one set of prejudice/threat scales assessing realistic and symbolic threat and negative stereotyping developed by González, Verkuyten, Weesie, and Poppe (2008) is evaluated. As with Croucher, DeMaris, Diers-Lawson and Roper (2017), we assert all sources of attributions have systematic biases (e.g., actor/observer discrepancies or social desirability bias). However, the triangulation of three different perspectives on an individual’s prejudice/threat predisposition should provide a strong assessment of these instruments.

Measuring Threat
Since the inception of Stephan and Stephan’s (1996) integrated threat theory, there have been several studies measuring the perception of different types of threat from several outgroups within different contexts (e.g. Berenberg, Friñay, Stephan, & Stephan 2002; Croucher 2013; Nshom 2016). The instruments for realistic threat and symbolic threat can be traced back to several studies carried out by Stephan and colleagues (1996, 1998, 1999, 2000, 2002). The majority of studies on threat have often adapted or modified these instruments to the context of each particular study (e.g. Berenberg et al. 2002; Croucher 2013, 2016; Croucher et al 2017). Concerning instrument adaptation, Scheibner and Morrison (2009) however acknowledged that the degree to which such modifications or alterations may compromise the validity of the instrument or increase measurement error is unclear. One of the main differences between the González et al. (2008) instruments used in this current study and the ones referred to above has to do with the number of items. Particularly, González et al. (2008) used three items to measure realistic threat and three to measure symbolic threat, while initial studies by Stephan and colleagues (1996, 1998, 1999, 2000, 2002) used a range of 9 to 14 items to measure realistic and symbolic threat. Several studies have followed in the footsteps of González et al. (2008) (e.g. Croucher 2013; Croucher, Aalto, Hirvonen, & Sommier 2013; Nshom 2016, Nshom & Croucher 2014, 2017).

Researchers using ITT measures have scarcely analysed these instruments for construct validity. Studies that have conducted factor analyses are rare. The few that have been done (e.g. Aberson & Gaffney 2008; Stephan, Ybarra, Martinez, Schwarzwal, & Tur-Kaspa 1998), reported that the items in the threat variable were rated as conceptually distinct. Factor analyses have also suggested the threat variables loaded on different factors. In another study, Renfro, Duran, Stephan, and Clason (2006), found the items for realistic threat loaded on two factors and most items for the symbolic threat measure loaded on the same two factors as realistic threat. Renfro et al. (2006) argued that even though realistic and symbolic threats are conceptually distinct, they are highly related.

Social Desirability
In prejudice research, Robinson et al. (1991) identified various criteria through which to evaluate instruments: writing of the items, avoidance of response set, and statistical and psychometric procedures involved in instrument construction. Response set is the tendency of individuals to respond to questions in ways other than the content of the statements/questions. Two key sources of response set have been shown in prejudice and attitudinal measures. First, participants may acquiesce, or change their attitudes in accordance with the situation (Paulhus 1991). The second is social desirability, or the tendency for participants to answer statements/questions in ways that make themselves appear more culturally positive or acceptable. Social desirability bias within self-report measures represents a concern to validity of prejudice measures (Burkard et al. 2002). In fact, various methods over the past 50 years have been developed in prejudice research to attempt to control for social desirability bias in self-reports, demonstrating the concerns over the potential effects of social desirability bias (DeMaio 1984; Lanning 1989; Zerbe & Paulhus 1987).

Regarding ITT instruments in particular, González et al. (2008) argued using self-report measures could be a potential limitation in measuring prejudicial feeling/attitudes. Similarly, Rick, Mania, and Gaertner (2006) observed that prejudice and ITT instruments overly rely on self-reports. Moreover, Nshom and Croucher (2017) also noted there is a possibility for responses to ITT measures (particularly negative stereotypes) to be affected by social desirability bias. However, even though the potential effects of social desirability in self-reports of prejudice, like ITT, have been asserted, researchers have yet to empirically explore or measure these effects. Numerous statistical and procedural remedies have been proposed for self-reports and social desirability, each with strengths and weaknesses. Podesakk and Organ (1986) for example proposed: escalating the unit of analysis, separation of measurement and instrument reordering. Spitzberg and Hecht (1984) recommended the use of other reports to reduce the level of bias inherent in self-reports.

As research using integrated threat theory increases (see Croucher 2013; Nshom & Croucher 2017; Stephan et al. 2015 for reviews) it is critical to understand how such instruments function. Other reports of prejudice could provide valuable perceptual data on measurements of prejudice (Meyer 2003; Paulhus 1991). For example, Nshom and Croucher (2017) called for future work exploring social desirability, reliability, and validity of ITT instruments; the current study is such an attempt. To assess the potential social desirability bias in ITT instruments, with the González et al. (2008) instruments serving as an example, we offer the following research questions.

**RQ1:** How internally consistent are the González et al. (2008) instruments as other-report measures?

**RQ2:** To what extent do self, friend, and intimate-other reports of threat vary?

**Method**

**Sample**

Two hundred thirty-four (N = 234) people participated in the study. An equal proportion of participants (n = 78) completed either self-reports of threat, a friend-report of threat, or an intimate-report of threat. Those who agreed to participate located a friend to complete the friend-report and an intimate (sexual) other to complete the intimate-other report. All three groups were directed to an online survey questionnaire through a cover letter that detailed the procedures and purpose of the survey. Initial participants were recruited through a snowball sample and as part of an extra credit assignment in a communication course. Of the 175 initial invitations sent out, 78 full sets of completed questionnaires were returned for a response rate of 44.57%.

As there was a disproportionate number of female self and friend-reports, and a disproportionate number of male intimate-reports, independent sample *t*-tests were conducted to assess differences between the sexes on the measures. For the self-reports, the *t*-tests revealed no significant differences between men\_M = 4.12; SD = 0.76) and women\_M = 4.40; SD = 0.86) on realistic threat \( t(76) = -1.16, p = .25 \), on symbolic threat: men\_M = 3.65; SD = 1.35), women\_M = 3.95; SD = 1.07) \( t(76) = -0.96, p = .36 \), on symbolic threat: men\_M = 2.34; SD = 0.75), women\_M = 2.21; SD = 0.51) \( t(76) = 0.94, p = .37 \). For the friend-reports, the *t*-tests revealed no significant differences between men\_M = 4.29; SD = 1.17 and women\_M = 4.16; SD = 0.92) on realistic threat \( t(76) = 0.51, p = .60 \), on symbolic threat: men\_M = 4.10; SD = 1.07), women\_M = 3.92; SD = 0.96) \( t(76) = 1.02, p = .31 \), on negative stereotypes: men\_M = 2.30; SD = 0.51), women\_M = 2.24; SD = 0.41, \( t(76) = .72, p = .47 \). For the intimate-reports, the *t*-tests revealed no significant differences between men\_M = 4.06; SD = 1.02, and women\_M = 3.95; SD = 0.90) on realistic threat \( t(76) = 2.24, SD = 0.43; SD = 0.68) \( t(76) = -1.02, p = .30 \). For the intimate-reports, the *t*-tests revealed no significant differences between men\_M = 2.31; SD = 0.54, women\_M = 2.36; SD = 0.88) \( t(76) = -1.04, p = .29 \). See Table 1 for demographic information about all three sets of participants.

**Table 1: Participant Demographics**

<table>
<thead>
<tr>
<th>Sex</th>
<th>Self</th>
<th>Friend</th>
<th>Intimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>17</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>Female</td>
<td>61</td>
<td>62</td>
<td>19</td>
</tr>
<tr>
<td>Age</td>
<td>Range</td>
<td>19-69</td>
<td>19-73</td>
</tr>
<tr>
<td>M</td>
<td>34.31</td>
<td>34.21</td>
<td>35.06</td>
</tr>
<tr>
<td>SD</td>
<td>13.48</td>
<td>13.45</td>
<td>13.93</td>
</tr>
</tbody>
</table>

**Instrumentation**

The questionnaire contained measures of realistic threat, symbolic threat, negative stereotypes from González et al. (2008) and demographic measures. These instruments in the original González et al. (2008) study showed an alpha reliability of .80, .89, and .83 for realistic threat, symbolic threat, and negative stereotypes respectively.

**Realistic threat**

The perception of realistic threat was measured with three statements (González et al., 2008). These included “Because of the presence of refugees, Finns have more difficulty finding a job,” “Because of the presence of refugees, Finns have more difficulty finding a house,” “Because of the presence of refugees, unemployment in Finland is increasing”. The same statements were slightly modified for the friend report and intimate-report simply by adding, “Your friend believes that” (for friend report) and “your partner believes that” (for intimate-report) at the beginning of each statement. Responses ranged from (1) strongly disagree to (5) strongly agree. Higher scores indicate more threat.
Symbolic threat

The perception of symbolic threat was measured with three statements (González et al., 2008). These statements included “Finnish identity/culture is threatened because there are too many refugees today,” “Finnish norms and values are threatened because of the presence of refugees today,” and “Refugees are a threat to Finnish culture.” The same statements were slightly modified for the friend report and intimate-report by simply adding, “your friend believes” (for friend report) and “your partner believes” (for intimate-report) at the beginning of each statement. Responses ranged from (1) strongly disagree to (5) strongly agree. Higher scores indicate more threat.

Negative stereotypes

For negative stereotypes, eight trait adjectives were given, and the main participant (self-report) was asked to indicate the extent to which these traits described refugees. The traits included violent, dishonest, unintelligent, friendly (reverse-scored), arrogant, kind (reverse-scored), greedy, and inferior (González et al., 2008). Responses ranged from (1) no, absolutely not, to (5) yes, certainly. For friend report and intimate report, participants where asked the extent to which they thought their friend or intimate-other believed the above stereotypes appropriately described refugees. The question started as follows: “Your friend believes the following stereotype describes refugees...” and “Your partner believes the following stereotype describes refugees...” for friend and intimate report respectively. Responses ranged from (1) no, absolutely not, to (5) yes, certainly.

Results

Regarding the first research question, alpha reliability analysis revealed the González et al. (2008) instruments of integrated threat produced consistently high reliabilities in both the self-report version and in the other-report versions (see Table 2). This suggests the measures are acceptable as other-reports.

Table 2: Reliability Coefficients for Self-and Other-Reports of Threat

<table>
<thead>
<tr>
<th></th>
<th>Realistic Threat</th>
<th>Symbolic Threat</th>
<th>Stereotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-Report</strong></td>
<td>.86</td>
<td>.94</td>
<td>.73</td>
</tr>
<tr>
<td><strong>Friend-Report</strong></td>
<td>.92</td>
<td>.94</td>
<td>.74</td>
</tr>
<tr>
<td><strong>Intimate-Report</strong></td>
<td>89</td>
<td>.95</td>
<td>.73</td>
</tr>
</tbody>
</table>

Note: n = 78 triads.

The second research question assessed the degree to which self-and-other reports of threat varied. To determine this, scores on realistic threat, symbolic threat, and negative stereotypes self-report measures were compared to the scores on realistic threat, symbolic threat, and negative stereotypes on the friend-reports and intimate other-reports. Independent samples t-tests indicated there were no significant differences between self-and-friend reports of symbolic threat \( t(151.33) = -.45, p = .65 \), realistic threat \( t(154) = 1.01, p = .32 \), and negative stereotypes \( t(154) = -19, p = .84 \). Independent samples t-tests also indicated there were no significant differences between self-and-intimate other-reports of symbolic threat \( t(154) = -.59, p = .56 \) and negative stereotypes \( t(154) = -1.15, p = .25 \). There was a statistically significant difference between self-reports of realistic threat and intimate other-reports of realistic threat \( t(154) = 2.05, p = .04 \) (see Table 3 for group means and standard deviations).

Table 3: t-test Comparison of Self/Other Reports of Threat

<table>
<thead>
<tr>
<th></th>
<th>Realistic Threat</th>
<th>Symbolic Threat</th>
<th>Stereotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-Report</strong></td>
<td>4.34a</td>
<td>.89</td>
<td>3.88</td>
</tr>
<tr>
<td><strong>Friend-Report</strong></td>
<td>4.19</td>
<td>.97</td>
<td>3.96</td>
</tr>
<tr>
<td><strong>Intimate-Report</strong></td>
<td>4.03a</td>
<td>.99</td>
<td>3.98</td>
</tr>
</tbody>
</table>

Note: Subscript indicates statistically significant mean difference based on independent sample t-test.

Discussion

This study had two main objectives: to investigate the extent to which the González et al. (2008) instrument is internally consistent as other-report measures and to examine the extent to which self, friend, and intimate-other reports of threat vary. First, the analysis revealed that the González et al. (2008) instrument was consistent as other (friend and intimate) report measures. This is supported by the fact that the instrument produced consistently high reliabilities for both the self-report version and the other-report versions. The implication of this is that the González et al. (2008) instrument is reliable not only as a self-report measure but also as an other-report measure (friend and intimate other). Even though this instrument has often obtained high reliability scores for self-reports (e.g., Nshom, 2016; Nshom & Croucher, 2017), this is the first time its reliability and use as an other report has been investigated. Second, this study showed there were no significant differences between self and friend report-measures for the different type of threats (realistic threat, symbolic threat, and negative stereotype). This means self-reports were consistent with friend reports. This study also indicated there were no significant differences between self-report and intimate report for symbolic threat and negative stereotype but regarding realistic threat, it showed there was a significant difference.

While these findings demonstrate the reliability of the González et al. (2008) ITT instrument as reliable measures of prejudice, the study is not without its limitations. The sample for this study was a convenience sample, which does limit its generalizability to the broader Finnish population. Even with this limitation, future research should be considered. Future research should continue to explore the extent to which other reports may be affected by social desirability. The tendency and possibility for friends and intimate others to answer statements/questions in ways that make their friend or intimate other appear more culturally positive or acceptable cannot be ignored. Ultimately, social desirability bias does not only represent a concern within self-report measures (Burkard et al., 2002) but also within other reports. Second, we cannot ignore the limitations of an online survey. Some groups are automatically restricted, and a large sample size becomes difficult to attain. Moreover, an important question emerging from this study is why there is a significant difference between self and intimate-other reports for realistic threat. This difference could most likely be attributed to social desirability bias on intimate other-reports, not on self-reports. This is more likely because there was no significant difference between self-report and friend report for realistic threat. In essence, intimate others may have a slightly more positive view of how their partners perceive refugees as realistically threatening Finnish society. Intimate others might tend to view their partners more positively, because they do not want to see their partner as being a prejudicial person, as having such negative behaviours/emotions. Future research should further examine this attribution. These results point to the necessity for more innovative and scientific ways of testing for social desirability among prejudice scales. Prejudice is
very sensitive but important issue in society and understanding how people really feel and think about a particular group is important to improving intergroup relations.

Future research should also consider the extent to which social desirability affects different age groups and ethnic groups. It can be expected that different age groups and different cultures may vary in the way they respond to prejudice. Future research should consider variables that account for such differences on how people respond to prejudice. It is also recommended that future studies using self and other reports, should consider the possibility of analysing the extent to which they differ on every item in the González et al. (2008) ITT scale. In this way, there will be more precision ad micro level understanding on how self-reports and other reports vary.

The González et al. (2008) ITT scale has gained popularity among researchers for the past couple of years. Yet, until now, no studies have considered social desirability bias on this scale. This study analysed the González et al. (2008) integrated threat scales, and explored the extent to which the different versions of the scale (self and other) differed. The analysis revealed that the scale was reliable and consistent. Moreover, there were no differences between self-reports and friend reports for the three types of threat. There was however a difference between self-reports and intimate other reports on realistic threat. This study advances our understanding of the prejudice/threat scales. There have been numerous studies on threat, but this is the first time one threat/prejudice scale is being analysed for social desirability by comparing self-reported measures to intimate other and friend reports.

References


### About the Authors

Stephen M Croucher (PhD, University of Oklahoma) is a Professor and Head of the School of Communication, Journalism, and Marketing at Massey University, New Zealand.

Elvis Nshom (PhD, University of Jyväskylä) is an Assistant Professor in the Department of Communication, California State University, San Marcos, USA.

Cheng Zeng (PhD, University of Jyväskylä) is an Assistant Professor in the Department of Communication at North Dakota State University.

Diyako Rahmani (PhD, University of Jyväskylä) is a Lecturer in School of Communication, Journalism, and Marketing at Massey University, New Zealand.

### Authors' Address

Stephen Croucher  
Massey Business School  
Massey University  
Private Bag 756  
Wellington 6140  
New Zealand  
T: 04 801 5799, ext: 63578

*Journal of Intercultural Communication, ISSN 1404-1634, issue 50, July 2019.*  
URL: [http://immi.se/intercultural](http://immi.se/intercultural)