Author's response to reviews

Title: Measuring bias in self-reported survey research: validating the Crowne-Marlowe Social Desirability Scale in Ethiopia, Kenya, Mozambique, and Uganda

Authors:

Alexander Vu (avu@jhsph.edu)
Nhan Tran (nhtran@jhsph.edu)
Kiemanh Pham (kpham4@jhmi.edu)
Saifuddin Ahmed (sahmed@jhsph.edu)

Version: 2 Date: 31 August 2011

Author's response to reviews: see over
Dear Editor,

Thank you for your assistance. We are grateful for the reviewers’ comments, which helped us to clarify our findings and stimulated our thinking.

We have extensively revised the manuscript now reflect major changes to the aim of the paper, based on the reviewers’ comments. The aim of the revised manuscript now focuses on measuring the reliability of the Marlowe Crowne social desirability scale in four countries in sub-Saharan Africa.

Don’t hesitate to contact me should you have any question.

Thank you again for your consideration.

Best regards,

Alexander Vu

Reviewer #Andrea Petroczi (Reviewer Comments to the Author):

General Comments:

The aim of the paper is to provide a social desirability scale to complement HIV risk behaviour research relying on self-reports among African populations. Whilst this aim is well justified, the paper – as it stands – falls short of this promise. The comments below intend to help the authors to enhance their work to deliver on this worthy aim and produce a paper that warrants publication. On a positive note, the paper is clearly written and well structured. The translation process appears to be adequate and the sampling is sufficiently rigorous.

Major compulsory revisions:
1. Marlowe-Crowne SD scale is misinterpreted in the text. SD is a tendency of individuals to reply in a way that will be, presumably, viewed favorably by others (emphasis on the ‘individuals’). In other word, Marlowe-Crowne SD taps into a personality trait coined ‘approval seeking’ (Crowne & Marlowe 1964) or later, more appropriately ‘avoidance of disapproval’ (Crowne 1979). As such, SD effect individuals’ responses on other self-reports to a varying degree. Consequently, SD effect should be discussed in relation to a potential bias it may cause in other psychological/behavioural measures. Marlowe-Crowne SD scale measures

We appreciate the Reviewer’s thoughtful comments and reference to the literatures on SD. We have revised the Background section accordingly (page 3, 2nd paragraph):

“A pervasive concern with self-reported surveys used to evaluate these programs is how the collected data is affected by social desirability (SD) [3]. Social desirability is a tendency of respondents to reply in a manner that would be viewed positively by their social peers or those that are consistent with social norms and expectations [4]. It has been found that when answering questions that concern sensitive topics, such as questions about sexual behaviors, it is common for respondents to describe their behaviors in a more favorable manner [5]. This effect of SD may reflect a deliberate misrepresentation of their true behavior. Alternatively, SD may reflect individuals’ need to seek approval by peers and as such, respondents may provide responses in order to gain approval by peers. Regardless of the reason, the possibility that respondents misrepresent socially undesirable behavior is of great concern to researchers as it puts into question the validity of self-reported survey data. The effect on self-reports remains the same: individuals with high SD who will score higher on measures of “agreeableness” and other socially desirable traits [6].

Moreover, cultural norms and attitudes regarding sexual practices differ greatly across countries and cultures. The tendency for the effect of SD to be observed in self-reported sexual practices and behaviors varies according to the setting in which the surveys are implemented [7]. The measure of the SD score of the individual should be considered in the context of and compared to the scores of the cultural norm of the target population. It is only when the SD score of the individual differs from that of the normed test scores would there be concern for SD affecting the individual’s responses to sensitive questions [6].

Marlowe and Crowne proposed and developed a social desirability scale (MC-SDS) to measure socially desirable responding [8]. The original MC-SDS contains 33 strong true or false statements that would influence the individual to respond in a manner that conforms to social expectations regarding behaviors, attitudes, and beliefs. Using this scale, the effect of SD can be observed in individuals who have tendencies to seek approval from their social peers or who may misrepresent their true behavior so as to make a good impression. These individuals would tend to score high on the social desirability scale.”

2. Evidence for reliability (internal consistency) is provided as Cronbach alpha coefficients. The authors should justify this choice (Cronbach alpha is an internal consistency estimate
for Likert (or Likert-type) scales and how it might affect the reliability estimates. Kuder-Richardson formulae (KR-20 or KR-21) are better reliability estimator for scales with dichotomous response options (e.g. true/false).

In response to the suggested analysis for the KR-20, we have re-analysed the data. The following table shows that the KR-20 is almost identical to the Cronbach alpha.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Ethiopia</th>
<th>Kenya</th>
<th>Mozambique</th>
<th>Uganda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha</td>
<td>0.800</td>
<td>0.629</td>
<td>0.657</td>
<td>0.700</td>
</tr>
<tr>
<td>KR-20</td>
<td>0.794</td>
<td>0.629</td>
<td>0.660</td>
<td>0.703</td>
</tr>
</tbody>
</table>

As most public health readers are more familiar with Cronbach alpha when reporting reliability, we have decided to use the Cronbach alpha for this manuscript. It is also important to note that, with dichotomous data, coefficient alpha is mathematically equivalent to KR-20, and Cronbach alpha may be used for both dichotomous and categorical (Likert-type) scales.

3. Further to evidence for reliability, I would have liked to see some attempt to explain the less than desirable Cronbach alpha values. Have the authors conducted any analysis (e.g. a simple item to total correlation or confirmatory factor analysis) that might shed the light on this? If so, please include. If not, please do and include.

In the process of analysing this dataset, we have already performed the factor analysis and the item to total correlation. The factor analysis results were not included because the objective of the study was not to examine the underlying different constructs of SDS, but to examine the internal consistency for which Cronbach alpha is the appropriate statistics, which we presented.

We are presenting the detailed results analysis of interests, below:

**Ethiopia:**

alpha sds_1-sds_28
- Average interitem covariance: 0.0246532
- Number of items in the scale: 28
- Scale reliability coefficient: 0.7998

kr20 sds_1-sds_28
- Kuder-Richarson coefficient of reliability (KR-20)
- Number of items in the scale = 28
- Number of complete observations = 114

<table>
<thead>
<tr>
<th>Item</th>
<th>Obs</th>
<th>difficulty</th>
<th>variance</th>
<th>correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>sds_1</td>
<td>114</td>
<td>0.9298</td>
<td>0.0653</td>
<td>0.0390</td>
</tr>
<tr>
<td>sds_2</td>
<td>114</td>
<td>0.5965</td>
<td>0.2407</td>
<td>0.3905</td>
</tr>
<tr>
<td>sds_3</td>
<td>114</td>
<td>0.6842</td>
<td>0.2161</td>
<td>0.3010</td>
</tr>
<tr>
<td>sds_4</td>
<td>114</td>
<td>0.5965</td>
<td>0.2407</td>
<td>0.3623</td>
</tr>
<tr>
<td>sds_5</td>
<td>114</td>
<td>0.3684</td>
<td>0.2327</td>
<td>0.2790</td>
</tr>
<tr>
<td>sds_6</td>
<td>114</td>
<td>0.7281</td>
<td>0.1980</td>
<td>0.3623</td>
</tr>
<tr>
<td>sds_7</td>
<td>114</td>
<td>0.6667</td>
<td>0.2222</td>
<td>0.3279</td>
</tr>
<tr>
<td>Item</td>
<td>Obs</td>
<td>difficulty</td>
<td>variance</td>
<td>correlation</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>------------</td>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>sds_1</td>
<td>113</td>
<td>0.8333</td>
<td>0.1389</td>
<td>0.1453</td>
</tr>
<tr>
<td>sds_2</td>
<td>113</td>
<td>0.2632</td>
<td>0.1939</td>
<td>0.1504</td>
</tr>
<tr>
<td>sds_3</td>
<td>113</td>
<td>0.7193</td>
<td>0.2019</td>
<td>0.1156</td>
</tr>
<tr>
<td>sds_4</td>
<td>113</td>
<td>0.3509</td>
<td>0.2278</td>
<td>0.2855</td>
</tr>
<tr>
<td>sds_5</td>
<td>113</td>
<td>0.1316</td>
<td>0.1143</td>
<td>0.0864</td>
</tr>
<tr>
<td>sds_6</td>
<td>113</td>
<td>0.8772</td>
<td>0.1077</td>
<td>0.0552</td>
</tr>
<tr>
<td>sds_7</td>
<td>113</td>
<td>0.2281</td>
<td>0.1761</td>
<td>0.1338</td>
</tr>
<tr>
<td>sds_8</td>
<td>113</td>
<td>0.4386</td>
<td>0.2462</td>
<td>0.1726</td>
</tr>
<tr>
<td>sds_9</td>
<td>113</td>
<td>0.4298</td>
<td>0.2451</td>
<td>0.3462</td>
</tr>
<tr>
<td>sds_10</td>
<td>113</td>
<td>0.8772</td>
<td>0.1077</td>
<td>0.4059</td>
</tr>
<tr>
<td>sds_11</td>
<td>113</td>
<td>0.4737</td>
<td>0.2493</td>
<td>0.2915</td>
</tr>
</tbody>
</table>

KR20 coefficient is 0.7936

Kenya

calculate alpha for sds_1-sds_28
Average interitem covariance: 0.0113569
Number of items in the scale: 28
Scale reliability coefficient: 0.6292

Kuder-Richarson coefficient of reliability (KR-20)
Number of items in the scale = 28
Number of complete observations = 113
sds_12 | 113  | 0.6228 | 0.2349 | 0.3243  
sds_13 | 113  | 0.8158 | 0.1503 | 0.2802  
sds_14 | 113  | 0.8246 | 0.1447 | 0.1619  
sds_15 | 113  | 0.6404 | 0.2303 | 0.0489  
sds_16 | 113  | 0.3684 | 0.2327 | 0.1074  
sds_17 | 113  | 0.7632 | 0.1807 | 0.2135  
sds_18 | 113  | 0.8158 | 0.1503 | 0.3844  
sds_19 | 113  | 0.2018 | 0.1610 | 0.0787  
sds_20 | 113  | 0.3509 | 0.2278 | 0.3230  
sds_21 | 113  | 0.6228 | 0.2349 | 0.1076  
sds_22 | 113  | 0.5351 | 0.2488 | 0.0780  
sds_23 | 113  | 0.6053 | 0.2389 | 0.1566  
sds_24 | 113  | 0.6404 | 0.2303 | 0.3875  
sds_25 | 113  | 0.2544 | 0.1897 | -0.0228 
sds_26 | 113  | 0.7193 | 0.2019 | 0.1700  
sds_27 | 113  | 0.6930 | 0.2128 | 0.1758  
sds_28 | 113  | 0.6140 | 0.2370 | 0.1045  

Test | 0.5611 | 0.1881  

KR20 coefficient is 0.6293

Mozambique

alpha sds_1-sds_28
Average interitem covariance: 0.0102943
Number of items in the scale: 28
Scale reliability coefficient: 0.6571

kr20 sds_1-sds_28
Kuder-Richarson coefficient of reliability (KR-20)
Number of items in the scale = 28
Number of complete observations = 114

<table>
<thead>
<tr>
<th>Item</th>
<th>Obs</th>
<th>Item difficulty</th>
<th>Item variance</th>
<th>Item-rest correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>sds_1</td>
<td>114</td>
<td>0.9737</td>
<td>0.0256</td>
<td>0.1068</td>
</tr>
<tr>
<td>sds_2</td>
<td>114</td>
<td>0.6842</td>
<td>0.2161</td>
<td>0.2004</td>
</tr>
<tr>
<td>sds_3</td>
<td>114</td>
<td>0.8158</td>
<td>0.1503</td>
<td>0.2805</td>
</tr>
<tr>
<td>sds_4</td>
<td>114</td>
<td>0.5614</td>
<td>0.2462</td>
<td>0.2948</td>
</tr>
<tr>
<td>sds_5</td>
<td>114</td>
<td>0.1842</td>
<td>0.1503</td>
<td>0.0188</td>
</tr>
<tr>
<td>sds_6</td>
<td>114</td>
<td>0.9561</td>
<td>0.0419</td>
<td>0.2873</td>
</tr>
<tr>
<td>sds_7</td>
<td>114</td>
<td>0.6228</td>
<td>0.2349</td>
<td>0.2920</td>
</tr>
<tr>
<td>sds_8</td>
<td>114</td>
<td>0.8772</td>
<td>0.1077</td>
<td>0.1531</td>
</tr>
<tr>
<td>sds_9</td>
<td>114</td>
<td>0.6140</td>
<td>0.2370</td>
<td>0.1434</td>
</tr>
<tr>
<td>sds_10</td>
<td>114</td>
<td>0.8860</td>
<td>0.1010</td>
<td>0.2417</td>
</tr>
<tr>
<td>sds_11</td>
<td>114</td>
<td>0.6228</td>
<td>0.2349</td>
<td>0.2135</td>
</tr>
<tr>
<td>sds_12</td>
<td>114</td>
<td>0.7544</td>
<td>0.1853</td>
<td>0.1240</td>
</tr>
<tr>
<td>sds_13</td>
<td>114</td>
<td>0.9298</td>
<td>0.0653</td>
<td>0.2209</td>
</tr>
<tr>
<td>sds_14</td>
<td>114</td>
<td>0.9298</td>
<td>0.0653</td>
<td>0.1502</td>
</tr>
</tbody>
</table>
sds_15  |  114     0.4561     0.2481     0.0515
sds_16  |  114     0.6228     0.2349     0.2863
sds_17  |  114     0.8421     0.1330     0.0493
sds_18  |  114     0.6930     0.2128     0.2181
sds_19  |  114     0.3158     0.2161     0.2231
sds_20  |  114     0.6140     0.2370     0.2255
sds_21  |  114     0.9386     0.0576     0.0991
sds_22  |  114     0.7632     0.1807     0.2685
sds_23  |  114     0.7807     0.1712     0.1598
sds_24  |  114     0.7807     0.1712     0.4212
sds_25  |  114     0.7632     0.1807     0.3778
sds_26  |  114     0.8596     0.1207     0.1348
sds_27  |  114     0.8596     0.1207     0.3555
sds_28  |  114     0.8684     0.1143     0.1781
---------+------------------------------------------
Test     | 0.7346                     0.2063

KR20 coefficient is 0.6604

Uganda

alpha sds_1-sds_28
  Average interitem covariance:  .0143478
  Number of items in the scale:  28
  Scale reliability coefficient:  0.6998

kr20 sds_1-sds_28
  Kuder-Richarson coefficient of reliability (KR-20)
  Number of items in the scale = 28
  Number of complete observations = 114

<table>
<thead>
<tr>
<th>Item</th>
<th>Obs</th>
<th>Item difficulty</th>
<th>Item variance</th>
<th>Item-rest correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>sds_1</td>
<td>114</td>
<td>0.8421</td>
<td>0.1330</td>
<td>0.2300</td>
</tr>
<tr>
<td>sds_2</td>
<td>114</td>
<td>0.3070</td>
<td>0.2128</td>
<td>0.1713</td>
</tr>
<tr>
<td>sds_3</td>
<td>114</td>
<td>0.7281</td>
<td>0.1980</td>
<td>0.3613</td>
</tr>
<tr>
<td>sds_4</td>
<td>114</td>
<td>0.2281</td>
<td>0.1761</td>
<td>0.1633</td>
</tr>
<tr>
<td>sds_5</td>
<td>114</td>
<td>0.1842</td>
<td>0.1503</td>
<td>0.1902</td>
</tr>
<tr>
<td>sds_6</td>
<td>114</td>
<td>0.8947</td>
<td>0.0942</td>
<td>0.3372</td>
</tr>
<tr>
<td>sds_7</td>
<td>114</td>
<td>0.4298</td>
<td>0.2451</td>
<td>0.1260</td>
</tr>
<tr>
<td>sds_8</td>
<td>114</td>
<td>0.8684</td>
<td>0.1143</td>
<td>0.4881</td>
</tr>
<tr>
<td>sds_9</td>
<td>114</td>
<td>0.5439</td>
<td>0.2481</td>
<td>0.2061</td>
</tr>
<tr>
<td>sds_10</td>
<td>114</td>
<td>0.8860</td>
<td>0.1010</td>
<td>0.4633</td>
</tr>
<tr>
<td>sds_11</td>
<td>114</td>
<td>0.4386</td>
<td>0.2462</td>
<td>0.2724</td>
</tr>
<tr>
<td>sds_12</td>
<td>114</td>
<td>0.4035</td>
<td>0.2407</td>
<td>0.2974</td>
</tr>
<tr>
<td>sds_13</td>
<td>114</td>
<td>0.8860</td>
<td>0.1010</td>
<td>0.4262</td>
</tr>
<tr>
<td>sds_14</td>
<td>114</td>
<td>0.8860</td>
<td>0.1010</td>
<td>0.3746</td>
</tr>
<tr>
<td>sds_15</td>
<td>114</td>
<td>0.3421</td>
<td>0.2251</td>
<td>0.1797</td>
</tr>
<tr>
<td>sds_16</td>
<td>114</td>
<td>0.7368</td>
<td>0.1939</td>
<td>0.1738</td>
</tr>
<tr>
<td>sds_17</td>
<td>114</td>
<td>0.5965</td>
<td>0.2407</td>
<td>-0.0253</td>
</tr>
<tr>
<td>sds_18</td>
<td>114</td>
<td>0.7895</td>
<td>0.1662</td>
<td>0.3840</td>
</tr>
</tbody>
</table>
4. Despite the title and specific aim, the paper does not provide any evidence for validity. I suggest that the authors consult the APA’ Standards for Psychological Testing before addressing this point.

The intent of the manuscript is to investigate whether or not the Marlowe Crowne Social Desirability Scale can be applied to the African context. Thus, we have clarified this aim of this study in the revised title and in the Background (page 5, last paragraph).

Title: “Reliability of the Marlowe-Crowne Social Desirability Scale in Ethiopia, Kenya, Mozambique, and Uganda”

“The overall aim of this study is to measure the reliability of the MC-SDS in four African countries: Ethiopia, Kenya, Mozambique, and Uganda. We accomplished this aim by: 1) adapt and pilot test the MC-SDS in sub-Saharan Africa; 2) measure the internal consistency of the responses to the MC-SDS in these four countries in sub-Saharan Africa; and 3) explore opportunities for the application of the MC-SDS as a tool for assessing the potential impact of social desireability in self-reported surveys of HIV related risk factors in sub-Saharan Africa.”

5. Discussion and conclusion must be re-written once the above points are addressed.

The Discussion and Conclusion sections have been entirely revised (starting page 10, last paragraph).

Minor essential revisions:
6. Justification should be provided for selecting the Marlowe-Crowne SD scale over other SD scales.

In response to the reviewer’s comment, we have added the following paragraph to the Methods section, under Formative Research (page 6, 2nd paragraph):

“In addition, a comprehensive review of the literature on HIV-related survey research in Africa and the effect of SD on self-reported surveys was performed. The literature review identified several scales and techniques that were developed to assess for the effect of social desirability on
self-reported data and to correct for this problem (the SD scale developed by Edwards [15], the Lie scale from the Eysenck Personality Inventory [16], the bogus pipeline technique developed by Jones and Sigall [17], and the SD scale developed by Marlowe and Crowne [8] are a few examples). The decision to use the Marlowe Crowne SD scale for this study was based on the finding that the MC-SDS is one of the most commonly referenced and used measure of social desirability. The formative work also contributed to the identification of strategies for the pilot testing of the MC-SDS tool as well as the location and potential study populations for the study.

7. The short forms of the Marlowe-Crowne SD scale are notoriously unreliable. Why did the authors decide to translate and test them?

We agree with the reviewer and have omitted the paragraphs describing the short forms from the Methods section.

8. I would have liked to see a slightly more substantial discussion on social desirability.

We have re-written the Background to reflect this request:

“The United States committed $18.8 billion from 2003 to 2008 as part of the President’s Emergency Plan for AIDS Relief (PEPFAR) [1]. In its original formulation, more than $3 billion was allocated for the prevention of HIV [1]. The majority of these investments are evaluated through the use of self-reported surveys of HIV knowledge, attitudes, and sexual practices (KAP) to assess a population’s understanding about and behavior towards HIV [2].

A pervasive concern with self-reported surveys used to evaluate these programs is how the collected data is affected by social desirability (SD) [3]. Social desirability is a tendency of respondents to reply in a manner that would be viewed positively by their social peers or those that are consistent with social norms and expectations [4]. It has been found that when answering questions that concern sensitive topics, such as questions about sexual behaviors, it is common for respondents to self describe their behaviors to be more flattering [5]. This effect of SD may reflect a deliberate misrepresentation of their true behavior. The possibility that respondents may lie about socially undesirable behavior is of great concern to researchers as it raises the question of the validity of self-reported survey data. Another possibility is that the SD effect may represent individuals’ need to seek approval. Such individuals may be expected to represent themselves as having socially acceptable behavior and they may in fact behave in manners that are socially approved. Regardless, the effect on self-reports remains the same: individuals with high SD who will score higher on measures of “agreeableness” and other socially desirable traits [6].

Moreover, cultural norms and attitudes regarding sexual practices differ greatly across countries and cultures. The tendency for the effect of SD to be observed in self-reported sexual practices and behaviors varies according to the setting in which the surveys are implemented [7]. The measure of the SD score of the individual should be considered in the context of and compared to the scores of the cultural norm of the target population. It is only when the SD score of the individual differs from that of the normed test scores would there be concern for SD affecting the individual’s responses to sensitive questions [6].”
9. Abstract: first line of results are not results. These are sample characteristics (thus belong to the method).

We revised the Abstract and moved the characteristic statement (page 2, 2nd paragraph).

“The MC-SDS was adapted based on consultations with local stakeholders and pilot tested in Ethiopia, Kenya, Mozambique and Uganda. A total of 455 men and women (ages 15-24 years) participated in the survey which was administered by trained interviewers. The scores for the social desirability scales were calculated for all participants. An analysis of the internal consistency of responses was conducted using Cronbach’s α coefficient. An α coefficient of 0.7 or greater would suggest high internal consistency and reliability.”

10. Methods: how were the data collected? Self-administered questionnaire or interview? A discussion on how the method might affect the responses should be included.

Please see the addition under Methods section, under Data Collection & Analysis.

“The questionnaire was administered by data collectors through face to face interviews.”

Also, we have made added the following statement in the last paragraph of the Discussion section:

“Additionally, since these questionnaires were administered by an interviewer, it is possible that in this face-to-face interview format may have enhanced the SD effect and influenced the respondents to reply in a more positive manner to conform to peer expectations.”

11. Excluded items: why was No 29 excluded?

In response to this inquiry, we have added the following statement to the Methods section, under Description, Adaptation and Development of MC-SDS:

“This was revealed after the initial piloting of the translated MC-SDS questionnaire; in particular five statements were deemed inappropriate for use in the African context. A decision was made to remove these statements from the original list of 33 resulting in a questionnaire with 28 statements. For example, statement 29 which states “I have almost never felt the urge to tell someone off” was removed because the concept of “telling someone off” could not be translated in many of the local languages and it was believed that the essence of this statement was already captured in other statements.”

Reviewer: Carol Bova (Reviewer Comments to the Author):

Summary: This manuscript describes a revision of the original Marlowe-Crowne Social Desirability Scale for us in Ethiopia, Kenya, Mozambique and Uganda. This is an important paper with implications for future work that uses self-report measures to
evaluate outcomes that have inherent social desirability bias. The paper is brief and organized. The following are some issues that the author(s) will need to address before this is ready for publication:

Major Compulsory Revisions
1. Please clarify –it is more appropriate to calculate a Kuder-Richardson 20 for this scale (not a cronbach’s alpha because the response options are dichotomous (true/false)

   Please see the response to Reviewer #1’s comment #2.

2. Need citations in background section – there are multiple assertions presented without citations.

   We have corrected this.

3. There are no validity data provided (so to assert that this scale has not been validated in Africa on p 4 leads the reader to believe that you will also present validity data – so this should be re-worded). This scale is not ready for use until validity data can also be presented.

   Please see the response to Reviewer #1’s comment #4.

4. Please clarify whether the scale was administered as an interview or a self –report. If it was administered in an interview – need to indicate how this could have influenced the responses. If this was administered as a self report measure please indicate the literacy level of the population.

   Please see the response to Reviewer #1’s comment #10.

5. Was there missing data or problems with certain items in specific populations? How were missing data handled?

   There were no missing responses for the items in the SDS responses.

6. Please provide more detail about the back-translation procedures used for the short form and full scale used in this study. How did you evaluate cultural congruence of the measure (for example did items such as “practice what I preach” translate adequately to each of the languages? )

   The tool was translated by local translators; back translations were then reviewed by the research team. Additional edits were made through an iterative process to ensure that the translations captured the essence of the statements. The statements were translated into local languages and then piloted tested with convenience samples to determine the level of understanding. This was carried out by the local research assistants under the supervision of investigators at Johns Hopkins University (JHU). The results of the pilot study were
reviewed jointly by the JHU investigators and local research team and final edits were made based on the feedback of the interviewers on the capacity of the general population to understand and respond to the statements in the questionnaire.

7. Please provide data that demonstrated that the short versions were inconsistent.

*Please see the response to Reviewer #1’s comment #7.*

8. Need a section on study limitations

*We have added the limitations of the study to the last 2 paragraphs of the Discussion section (page 12)*

Minor Essential Revisions

1. In the abstract specify the four countries that you are referring to

*We have specified this in the Abstract, page 2.*

2. The title should indicate the most common name used for the scale – Marlowe-Crowne Social Desirability Scale and the abbreviations throughout should be changed to MC-SDS

*We have modified the title per the reviewer’s suggestion.*

3. Test-retest reliability procedures would have added strength to this study – the lack of this should be mentioned as a limitation

*We acknowledged this limitation in the text (page 13, last sentence of the Discussion section).*

4. Please provide a reference for the LQAS sampling method used

*We have provided this reference in the text.*

5. Reporting medians would be helpful to evaluate the distributions

*In Ethiopia, the SDB scores ranged from 6 to 28, with a mean of 19.5 and median of 20. In Kenya the SDB scores ranged from 5 to 25, with a mean of 15.7 and median of 16, which was the lowest among the four country sites. The SDB scores in Mozambique ranged from 12 to 28 with a mean of 20.6 and median of 21, the highest of the four countries. In Uganda, the SDB scores ranged from 4 to 24 with a mean of 17.5 and median of 18.*

*Since the means and medians were comparable, we chose to present means in the manuscript.*