An expanded empirical evaluation of the cross-national comparability of survey measures of political interest using anchoring vignettes: A research note

Seonghui Lee\textsuperscript{a}, Nick Lin\textsuperscript{b}, Randolph T. Stevenson\textsuperscript{c,}\textsuperscript{*}

\textsuperscript{a} Aarhus University, Denmark
\textsuperscript{b} University of Mannheim, Germany
\textsuperscript{c} Rice University, United States

\textbf{A R T I C L E  I N F O}

\textbf{Article history:}
Received 16 June 2016
Accepted 13 August 2016
Available online xxx

\textbf{1. Introduction}

In a previous paper (Lee et al., 2015 — hereafter “LLS”), we used King et al.’s (2004) “anchoring vignette” technique on samples from three countries (France, the United Kingdom, and the Netherlands) to show that individuals across these different contexts understand questions about political interest in the same way. This was an important result, we argued, both because it opened up the possibility of tapping the vast amount of data that has been collected on political interest across many countries and over many years and because it suggested that observed cross-national differences in levels of political interest are real and deserve a great deal more scholarly attention.

A drawback of that study was the limited number of (relatively homogeneous) countries we were able to study. In this short note we update the results of our earlier analysis with a much larger and diverse set of 12 countries drawn from Western and Eastern Europe, North and South America, and Asia. To preview, replicating the analysis from the previous paper on this greatly expanded sample produces the same result: respondents in countries as different as South Korea, Chile, the U.S., and Denmark all seem to understand the typical political interest question (and its answer categories) in the same way, suggesting that the measure of political interest is broadly comparable across these countries. Further, the scope and diversity of the sample gives us confidence that this conclusion will generalize to the many other countries (and time periods) in which this question has been asked.

In the rest of this paper, we present the empirical results that support this conclusion. However, to facilitate comparison with the results in LLS, we closely follow the empirical roadmap of that paper by presenting similar results and cross-referencing the figures and tables in this article with the ones in LLS. Further, rather than rehash here the lengthy motivational and methodological explanations provided in the previous paper, we simply refer readers to those discussions. That said, it will be useful to briefly summarize the main points made there:

1. Political interest questions have been asked in hundreds of surveys across most the world’s democracies (and many non-democracies) over a very long time, but have seldom been used to compare voters across contexts. Fairly large differences in the distribution of political interest across populations are apparent in the raw data. Thus, it is important to understand if these differences reflect real differences in political phenomena or simply cross-context differences in measurement.

2. Political interest questions tend to have a simple semantic structure and are asked in remarkably similar ways in different surveys. Thus, they are perhaps more likely than other common political questions to be interpreted in the same way across contexts.

3. Differential Item Functioning (DIF) is a threat to the cross-contextual compatibility of survey questions, even when questions are asked in exactly the same way across contexts. DIF happens when characteristics of the survey item and/or survey situation systematically cause individuals in different groups or contexts to understand the same survey questions differently or to have varied interpretations of the end-points and the cut-points of an associated measurement scale.

4. “Anchoring vignettes” is a technique, developed by Gary King and his colleagues (King et al., 2004; King and Wand, 2007), that identifies and ameliorates DIF caused by differing interpretations of the “cut-points” defining answer categories. It
does so by utilizing respondent assessments of one or more vignettes, which are descriptions of the kinds of hypothetical individuals that exemplify some level of the underlying concept to be measured.

5. When we developed appropriate anchoring vignettes for the typical political interest item and applied these to samples from France, the UK, and the Netherlands, we found few differences in the response patterns between the adjusted and unadjusted responses and, more generally, found little evidence that DIF was a problem for cross-national comparability of the raw political interest responses.

2. The expanded sample

In order to use anchoring vignettes to examine the extent of DIF in the typical political interest question across a large and diverse set of countries, we commissioned surveys in China, Denmark, Germany, South Korea, Japan, Poland, Argentina, Mexico, Chile, Canada, USA, and New Zealand.1 The interest question was:

Below we provide descriptions of three different people. We would like your opinion about how interested each of them is in politics.

Using a scale where 1 means “Not interested at all” and 6 means “Very interested,” to what extent, if at all, do you think each of the following person is interested in politics?

Please also place yourself on the same scale.

Fig. 1 provides the raw distributions of political interest on our six-point scale for each of the twelve countries in our sample.

The main question we address in this note is the extent to which these distributions change when one accounts for the possibility of DIF by using anchoring vignettes.

3. Constructing vignettes for political interest

Following Hopkins and King (2010), we asked respondents to evaluate the level of political interest of three hypothetical individuals before asking them to assess their own interest on the same 6-point scale. As discussed in Section 4 of LLS, for an inherently multi-dimensional concept like political interest, our constructed vignettes need to invoke the same, important sub-dimensions of interest in each vignette and make sure the levels of each of these aspects move monotonically and in the same direction from the low interest vignette to the high interest vignette. In addition, in order to make sure the set of vignettes discriminates individuals along the whole range of the scale, we need to try to describe individuals with various plausible levels of these variables. We attempted to achieve these goals with the following vignettes:2

[Vignette A] This person enjoys keeping up with politics and seeks out news about government and politics every day. This person likes to talk about politics and is eager to exchange political views with others.

[Vignette B] This person sometimes enjoys keeping up with politics and sometimes does not. This person will seek out news about government and politics if something particularly exciting or sensational is happening, but not otherwise. This person only occasionally participates in conversation about politics with other people.

[Vignette C] This person is turned off by politics. This person avoids news about government and politics and does not like to talk about politics with other people.

These three vignettes were presented in randomized order and the respondent was asked for their own level of interest after they had answered for each hypothetical individual.

In our view, these anchoring vignettes can be arranged on an ordered scale, ranging from the most interested in politics (Vignette A), to some interest (Vignette B), and then to the least interest (Vignette C). We will be able to test whether our respondents ranked them as we expected and if these rankings differed across countries.

Notice that each vignette taps the same two dimensions of interest (e.g., attending to political news and talking about politics with others) and these vignettes vary from one another only at the levels of these activities. This kind of monotonicity in the construction of the vignettes is essential if we want to encourage the respondent to think of the scale as unidimensional.

4. Do the vignettes fall on a unidimensional scale?

If respondents are placing the vignettes on the same scale, which they also perceive to be a unidimensional scale, and if we have written the vignettes to clearly capture relative positions on that scale, then the primary evidence of that should be found in the percentage of respondents who actually rate the vignettes in the order we expect.

1 These were conducted in eight different languages (Chinese, Danish, German, Korean, Japanese, Polish, Spanish, and English). The surveys were conducted by Toluna international utilizing their broadly representative internet panels. The decision to use internet panels reflected a necessary trade-off between available resources and the necessity, for this study, of expensive random sampling. Since our purpose was to look for evidence of DIF, which relies on comparisons of within sample differences in adjusted and unadjusted measures, we had less need for random samples than might be the case in studies designed for other purposes. The first survey was fielded in December 2014 and the last survey was completed in May 2015. Each survey (except Japan) was in their field at least 5 months after the previous national election and at least five months before the upcoming national elections (thus avoiding active campaign periods). The one exception was Japan, where the survey was conducted two weeks after the House of Representative election in 2014. The surveys were completed online using computers or mobile devices (16% used mobile devices). Across all countries, there were a total of 11,064 respondents (roughly 1000 respondents per country), excluding respondents identified as “shirkers” because they “straight-lined” — giving the same answer across many questions. Each survey was translated by paid translators with an MA or higher degree in social scientific disciplines. At least two translators for each language worked on the translation from the English version. After completion of individual translation work, each translator crosschecked another translator’s work and discussed the points of disagreement and different nuances with the researchers. In the end, the translated versions of vignettes were used which had high level of agreement among translators and researchers. Questionnaires in different languages are available upon request.

2 Given the much less homogenous societies in our new twelve country sample, it was necessary to use a different set of vignettes than we used in LLS — which referred to such things as the nightly news on television that might not be a relevant experience in some less developed countries. In addition, our previous effort made it clear that three vignettes would likely be sufficient to anchor our samples and so we economized on survey time by only providing three vignettes. We took the opportunity to use these samples to test the efficacy of the technique (advocated by many using anchoring vignettes) of matching names of individuals described by the vignette to the gender and culture of the respondent (e.g., choosing a typical female or male name). Specifically, in most of our countries, a randomly selected half-sample of respondents were shown specific names for each vignette instead of the phrase “this person.” These names were always matched to the gender and culture of the respondent. In this case, however, we found no significant differences across these treatment groups — a result that may reflect the overall lack of DIF effects for political interest questions rather than more general evidence that gender and culture matching is not consequential in situations in which DIF is more of a problem.
Table 1 gives the top 10 orderings of the vignettes by all the respondents in all twelve surveys (91% of the whole sample). Over 50% of respondents placed the three vignettes in exactly the order we expected and 75% percent of the respondents had no order violations—they placed all three vignettes in the order we expect but had some ties. Further, when we replicate this table country by country, we find similar results within each country. This strongly suggests that individuals across our sample interpret political interest (as asked in the typical question) as a unidimensional concept and our vignettes as describing easily differential positions on that dimension.

5. Do respondents in different countries differ systematically in their placement of the vignettes on the political interest scale?

Even if respondents see political interest as a unidimensional scale and our vignettes as occupying different positions on that scale, they might still interpret numerical positions on that scale quite differently (e.g., in one country respondents may think of a certain kind of person as low interest, while in another country the same person may usually be identified as very low interest). If DIF across countries is an important problem for a given measure (and if anchoring vignettes have been constructed properly) then we should see evidence of this DIF in differences in the distribution of vignette placements across countries. For example, King et al. (2004) found huge differences in average levels of political efficacy assigned to the same vignettes between Chinese and Mexican respondents—a result consistent with levels of DIF large enough to switch the order of Mexico and China (on average levels of political efficacy) between unadjusted and DIF corrected measures.

Fig. 2 gives the mean responses for each vignette for each country in our sample. Clearly, just as in LLS, this result is dramatically different from King et al.’s (2004) political efficacy study. Substantively, the mean responses across countries are very close together for each vignette and quite far apart across vignettes: that is, the total variance in responses pictured in Fig. 2 is composed of a great deal more between vignette variance than within vignette variance (which would mean variation across countries within the

Table 1

<table>
<thead>
<tr>
<th>Ordering</th>
<th>Frequency</th>
<th>Proportion</th>
<th>Cumulative proportion</th>
<th>Order violations</th>
</tr>
</thead>
<tbody>
<tr>
<td>A &gt; B &gt; C</td>
<td>5929</td>
<td>0.54</td>
<td>0.54</td>
<td>0</td>
</tr>
<tr>
<td>A ≈ B &gt; C</td>
<td>1150</td>
<td>0.10</td>
<td>0.64</td>
<td>0</td>
</tr>
<tr>
<td>C &gt; A ≈ B</td>
<td>525</td>
<td>0.05</td>
<td>0.69</td>
<td>2</td>
</tr>
<tr>
<td>B &gt; C &gt; A</td>
<td>494</td>
<td>0.04</td>
<td>0.73</td>
<td>2</td>
</tr>
<tr>
<td>B &gt; A &gt; C</td>
<td>482</td>
<td>0.04</td>
<td>0.74</td>
<td>1</td>
</tr>
<tr>
<td>A ≈ B &gt; C</td>
<td>468</td>
<td>0.04</td>
<td>0.77</td>
<td>0</td>
</tr>
<tr>
<td>B &gt; A &gt; C</td>
<td>468</td>
<td>0.04</td>
<td>0.81</td>
<td>1</td>
</tr>
<tr>
<td>A ≈ B &gt; C</td>
<td>415</td>
<td>0.04</td>
<td>0.85</td>
<td>0</td>
</tr>
<tr>
<td>A &gt; C &gt; B</td>
<td>366</td>
<td>0.03</td>
<td>0.88</td>
<td>1</td>
</tr>
<tr>
<td>C &gt; B &gt; A</td>
<td>286</td>
<td>0.03</td>
<td>0.91</td>
<td>3</td>
</tr>
</tbody>
</table>

Note: A > B > C is the expected ordering; only top 10 orderings are reported. Total proportion of cases with no order violation is 75% (some of the orders contributing this total are not included in the top 10 list above). This table corresponds to Table 2 in LLS.

4 In our sample countries, the proportion of respondents who placed the vignettes in the order we expected (i.e., A > B > C) ranges from 41% to 65%. Also, the proportion of no order violation cases ranges from 65% to 79%.
Indeed, when we formally estimate these quantities, we find that 97% of the variance in Fig. 2 is explained by variation between different vignettes and only 3% by between countries for the same vignette.

We can also compare histograms for ratings of each vignette across countries. These are depicted in Fig. 3, where we plot the empirical distribution (summarized with a kernel density) for each vignette for each country. Clearly, like in our previous study, respondents in all our countries largely agree about where to place each vignette on the 6 point political interest scale they were given. This is strong evidence against the idea that DIF (rather than other substantive causes) is responsible for observed patterns of self-reported levels of political interest across countries.

6. How much do cross-national differences in the distribution of self-ratings of political interest change when we correct them for DIF?

The final test of the cross-national comparability of our measure of political interest simply asks whether our conclusions about levels of political interest across countries change significantly after adjusting individual responses using anchoring vignettes. If there are few differences, we can conclude that the uncorrected measure does not suffer from cross-nationally consequential DIF.

The specific techniques used to adjust individual self-reported levels of political interest based on responses to the vignettes are explained in detail in LLS. However, the basic idea is easy to understand: Responses to the vignettes are used to equate the scores for each vignette across all respondents (who ordered the vignettes correctly) and then self-placements are adjusted to be on this new scale. Adjustments for individuals who mis-ordered the vignettes or gave two or more vignettes the same score can be handled in a several different ways.

---

5 The one hint of a difference in these data are the relatively high scores for the low interest case for China, Korea, and Japan (a result that is repeated for the mid-interest case for China). As we show below, however, this is not enough DIF to substantially change DIF adjusted scores or to undermine our general conclusions. That said, the fact that all three of the East Asian countries in our sample evidence this same effect is intriguing. Likewise, the direction of the effect (i.e., that the less interested are perceived as higher on the scale than in other contexts) is consistent with King et al.'s (2004) conclusion about the somewhat related concept of political efficacy in China — where the typical Chinese respondent thought that relatively efficacious people had higher levels of political efficacy than in other contexts.

6 In LLS, we also evaluate the extent to which the vignettes discriminate well among respondents and whether each contributes unique information (i.e., whether it is necessary). Since this question is tangential to the main point of this note, we simply summarize here the corresponding evidence for our 12 country sample: each vignette adds significantly to our ability to discriminate between cases as revealed by the kinds of entropy plots provided in, for example, Fig. 9 in LLS. The corresponding plots for our 12 country sample are available from the authors.

7 Each of these techniques is explained in Section 5 of LLS: The first method simply omits cases with tied and mis-ordered vignettes (Omitting ties); the second method distributes tied and mis-ordered cases uniformly in all interest categories (Uniform allocation); and the final method applies a censored ordered probit model to estimate the probability of allocation of those incorrectly ordered cases to one of the possible interest categories (Ordered probit).
Fig. 3. Density plots of vignette assessments. This figure corresponds to Fig. 8 in LLS.
Consistent with the scant evidence of cross-country DIF we have found so far, using these techniques to adjust our respondents’ self-reported levels of political interest reveals no evidence of significant cross-national differences in how respondents are answering these questions. Table 2 gives the rank order of the average level of self-reported political interest for each of the countries in our sample for the unadjusted raw scores and three different ways of adjusting responses using the anchoring vignette technique. Clearly, differences are minimal.

### Table 2
Relative ranking of each country on average self-reported political interest.

<table>
<thead>
<tr>
<th>Country</th>
<th>Raw score</th>
<th>Omitting ties</th>
<th>Uniform allocation</th>
<th>Ordered probit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Canada</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Chile</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>China</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Denmark</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Germany</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Japan</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Korea</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Mexico</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>New Zealand</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>US</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Note: This table corresponds to Table 4 in LLS. Columns 2–4 use different methods for using responses to our anchoring to adjust self-reported political interest for cross-national DIF.

7. Conclusion

In this short note, we update our previous paper exploring the cross-national comparability of the usual political interest questions that have been used in many hundreds of surveys and asked to millions of respondents over more than 50 years. That paper found little evidence of differential item functioning for this question; but, the analysis was based on only three countries that were relatively similar. The current effort expands the evidentiary basis of this conclusion to a set of twelve diverse countries from around the world, including countries in Eastern and Western Europe, North and South America, and Asia.

The evidence we present here significantly strengthens the case that political interest does not suffer from significant, cross-nationally relevant DIF and implies that the wealth of survey data on the concept should be used more often to explore the institutional, social, and cultural sources of cross-national differences in the distribution of political interest.

### References


