Adjusting self-reported attitudinal data for mischievous respondents

Michael R. Hyman  
New Mexico State University  
Jeremy J. Sierra  
Texas State University–San Marcos

For various reasons, survey participants may submit phoney attitudinal self-reports meant to bypass researcher scepticism. After suggesting reasons for this new category of problematic survey participant – the *mischievous respondent* (MR) – and reviewing related response bias, faking, inattentive respondent and outlier literatures, an initial algorithm for removing such respondents from polychotomous attitudinal data sets is posited. Applied to four data sets, this algorithm marginally reduced EFA cross loadings and improved CFA model fit. Although purging subtly suspicious cases is not standard practice, the extant literature indicates that such algorithms can reduce artifactual statistical findings substantially.

Self-report biases threaten the reliability of attitudinal data. For example, social desirability bias, triggered by self-deception enhancement and/or impression management, can cause misreports or non-responses to sensitive questions and uninformed answers to knowledge-based questions (Goldsmith 1989; Fisher 2000; King & Bruner 2000; Paulhus *et al.* 2003). Other problematic response-style biases, such as acquiescence, courtesy, extreme response, personality-trait and cultural (Heide & Grønhaug 1982; Bachman & O’Malley 1984; Greenleaf 1992a, 1992b; Mathews & Diamantopoulos 1995; Yates *et al.* 1997; Hurley 1998; Baumgartner & Steenkamp 2001; Dolnicar & Grün 2007), may cause an increase in outlier response patterns.

Careless responses also threaten the reliability of attitudinal data. Barnette (1999) assessed the effect of systematic and random responses on coefficient alpha. In a Monte Carlo simulation that mimicked responses to multiple seven-point Likert-type items, he replaced a fraction of *good data* with *bad data* that followed one of eight patterns: mono-extreme (all 1s or 7s), mono-middle (all 4s), big-step (1234567654321), small-step (1234567654321).

Received (in revised form): 20 January 2011
Adjusting self-reported attitudinal data for mischievous respondents

(1232123212321), checker-extreme (1717171717), checker-middle (3535353535), random, and a mixture of these patterns. Replacing only 5% of good data with mono-extreme or small-step data inflated coefficient alphas of 0.7 to 0.89 and 0.83, respectively. In contrast, the other patterns deflated coefficient alphas of 0.7 modestly to meaningfully, with the reduction worsening as the percentage of bad data increased. Thus, ‘even relatively few occurrences of some [bad data] patterns may highly influence alpha’ (Barnette 1999, p. 45).

Response biases and carelessness are reducible through research and questionnaire designs that boost participation benefits and trust in the researcher yet lower participation cost (Cohen & Carlson 1995; Berrens 2000; Vance et al. 2003; Dillman et al. 2009). Nonetheless, good study design alone cannot eliminate phoney self-reports meant to bypass researcher scepticism. Although the preponderance of phoney self-reporters, or mischievous respondents (MRs), is unknown, the many reasons for their behaviour ensure they comprise a potentially problematic subset of any survey sample. These reasons include: suspicion that findings will be applied nefariously; distrust of researchers and their work; general hostility towards business; thrill of being disrespectful and sabotaging a study; opportunity to play a prank; transference of anger from non-survey-related circumstance(s); retaliation against the survey sponsor; retaliation for previously encountered privacy invasions; aggressive solicitations to participate in surveys; and unethical survey practices (Walonick 1993; Maddox 1995; Barnette 1999; Nancarrow et al. 2004; Dillman et al. 2009). In addition, the oft-queried university student respondent (Peterson 2001) may retaliate for a poor expected grade or onerous course requirements.

Like outliers, data from several MRs may bias statistical results meaningfully. Previous studies probed the effects of outliers on estimates of means, correlations, regression parameters, t tests, F tests and \( \alpha \) (Liu & Zumbo 2007). Outliers can distort all least-squares statistics; for example, one outlier in a sample of 29 observations can reduce \( r \) from 0.99 to 0.00 (Lind & Zumbo 1993). Non-normally distributed outliers can inflate \( \alpha \) substantially (e.g. from 0.4 to 0.95) (Liu & Zumbo 2007). Seemingly meaningful factor structure can be an artifact of one or two outliers; even one outlier can create an extra factor (Huber 1981; Bollen & Arminger 1991). Several outliers can inflate fit indices and bias parameter estimates of ‘correct’ structural equation models (Yuan & Bentler 2001). Removing outliers can correct for SEM results with negative variances and correlations with latent variables greater than one (Bollen 1987). When
data represent many dimensions, ‘a small fraction of outliers can result in very bad estimates’ (Rocke & Woodruff 1996, p. 1047).

MRs would answer attitudinal questions in one of four ways: ‘as if’ another person (e.g. child who guesses a technophobic parent’s responses about Twitter), outrageously (e.g. extreme responses), systematically (e.g. 123212321) and indiscriminately. Responses of the first type are difficult to detect because they mimic legitimate data. Responses of the second type are detectable in data collected with well-designed questionnaires. Responses of the latter two types are detectable by visual inspection or algorithms of the type introduced later.

Because MRs – like careless respondents – can undermine surveys (Goldsmith 1989; Piferi & Jobe 2003), and published estimates of MR incidence run as high as 14% (Cooke & Regan 2008), researchers should remove MRs’ answers from data sets prior to substantive analyses. (Note: estimates of MR incidence by several informally queried survey researchers ran from 5% to 15%.) If MR incidence is substantial, then MRs need not be extreme outliers to threaten attitude research. Even if they fail to bias mean and variance estimates, MRs provide responses that exaggerate the power of statistical tests based on sample size.

Although MRs will tend to provide longitudinally inconsistent answers, retesting remains an unlikely option because data collection budgets often are constrained and most attitude surveys ensure recontact-precluding respondent anonymity. Instead we offer a novel, outlier-focused, data-cleansing approach.

We proceed as follows. After reviewing related response bias, faking, inattentive respondent and outlier literatures, we posit an algorithm – demonstrated with SPSS and Microsoft Excel but generalisable to any statistical and spreadsheet software – for purging likely MR cases from polychotomous attitudinal data sets. Applied to four data sets, this algorithm marginally reduced EFA cross-loadings and improved CFA model fit. Subsequently, we suggest a programme for future research.

**Possibilities inspired by solutions to parallel problems**

**Response bias**

Dedicated scales can detect response bias (e.g. Greenleaf 1992b). However, such scales add non-substantive questions that may offend some respondents, boost field service costs and promote respondent fatigue (Fisher 2000). For example, using a 30- to 40-item social-desirability-bias
scale or parallel (in)direct items to create methods factors in SEM would lengthen a questionnaire meaningfully (Fisher 2000; Jo 2000). Similarly, dedicated scales to detect MRs are non-preferred.

Reverse-coded items may reduce extreme response bias (Baumgartner & Steenkamp 2001) and detect inattentive respondents. Because the cognitive complexity of negatively worded items can create a method factor (Herche & Engelland 1996; Swain et al. 2008), researchers should use directly worded stems with bidirectional response options (i.e. all statements worded positively but response categories run in one direction for some statements and the reverse for other statements) (Barnette 2000). Other study-design-based techniques to reduce response bias include assuring confidentiality or anonymity, using indirect questions (e.g. answer ‘as if’ your best friend), avoiding personal data collection methods, using face-saving questions which imply that problematic behaviours are excusable, and using a randomised response method (Gordon & Petty 1971; Nancarrow et al. 2001). Although these methods might reduce mischievous responding, they cannot detect it.

**Faking**

The MR problem parallels the faking problem on personality and integrity tests (Zickar & Robie 1999; Peeters & Lievens 2005). Faking on such tests – to secure employment or university admission (Dalen et al. 2001), to conceal problematic behaviour (Johnson et al. 1998) or to malinger (Dannenbaum & Lanyon 1993) – means fakers answer in ways they believe will depict them favourably. Successful faking is possible; respondents instructed to fake good or fake bad on personality inventories produce profiles consistent with those instructions (Scandell & Wlazelek 1996; Bagby et al. 2002; Peeters & Lievens 2005). Unfortunately, faking reduces these tests’ predictive validity and creates increased common variance unrelated to substantive construct variance (Zickar & Robie 1999; Peeters & Lievens 2005).

Generally, researchers detect faking with outlier analysis (e.g. excessive scales or factors with scores at least one standard deviation from the mean) or (in)direct self-reported faking scales (Scandell & Wlazelek 1996; Dwight & Donovan 2003). Although faking scales suffer the same limitations as dedicated response bias scales, outlier analysis suggests the algorithm introduced later. For computer administration, response latency may indicate faking because fakers may take longer to complete questionnaires (Hsu et al. 1989).

Study-design ways to reduce faking include ensuring anonymity, warning that the questionnaire enables faking detection, and using subtle items
(which are less fakeable because they are less face valid to respondents) (Worthington & Schlottmann 1986; Dannenbaum & Lanyon 1993; Bornstein et al. 1994; Dwight & Donovan 2003). Although useful, such methods can neither eliminate nor detect MRs.

**Careless and random responding**

Incompatible answers to similar but reverse-worded reliability check items suggest careless or random responding. Odd replies to bogus items, like *I was born on February 30th* and/or infrequently endorsed items like *I was on the front cover of several magazines last year*, are clear indicators (Beach 1988; Retzlaff et al. 1991; Bagby et al. 2002; Dwight & Donovan 2003). When questionnaires include a separate answer sheet, respondents who answer non-existent questions (i.e. ghost answer sets) tend to answer reverse-scored items more inconsistently (Piferi & Jobe 2003). For repeatedly fielded surveys, over-reporting relative to standard baselines/cut-offs implies careless or random responding (Stein et al. 1995). Unfortunately, none of these methods would spot detection-shy MRs.

Asking the same or similar question repeatedly, or warning respondents about fictitious items, may reduce random responding but also may annoy respondents and imply they are distrusted (Goldsmith 1989; Calsyn et al. 2001; Cooke & Regan 2008). The MR-prone may interpret detection threats as challenges; hence such warnings may boost, rather than depress, MR incidence. Validity scales to detect random responders to personality assessment tools are reliable for identifying completely but not partially random responses (Archer et al. 2002). However, such scales suffer the same limitations as dedicated response bias and faking scales.

Other detection mechanisms include visual inspection for odd response patterns (Barnette 1999) and counting excessive use/non-use of one or more response categories. For online surveys, overly quick responses suggest inattentiveness (Bailey 1994; Barnette 1999; Cooke & Regan 2008). As not all surveys are administered online, a more widely applicable mechanism for detecting MRs is preferred. However, odd response patterns relate to the proposed algorithm.

**Addressing the mischievous respondent problem**

As with outliers, the post hoc ways to address the MR problem are robust statistical methods and data cleansing. Robust statistics, like maximum likelihood estimators, are less sensitive to outliers (Lind & Zumbo 1993).
Matrices with robust correlation/covariance estimates allow outlier-resistant factor analyses and structural equation models (Yuan & Bentler 1998, 2001; Pison et al. 2003). If MRs pose problems similar to outliers, then such statistics represent a viable tool for mitigating MR-induced bias.

Outlier analyses rely on measures that consider the distance of each case from the centroid given the covariance of the distribution, like the Mahalanobis $D^2$ or Comrey $D_k$ (Comrey 1985; Rasmussen 1988; Bacon 1995). Outliers that differ from the centroid at some arbitrary level of significance can be deleted. Although outliers and influential points can be detected with graphical methods, these methods become unwieldy for multidimensional data; hence, analytical methods are needed (Chatterjee et al. 1991).

Data cleansing methods like outlier analyses have shortcomings. In particular, outlier analyses rely on mean and variance estimates that include the data to be deleted, can become unwieldy with large multivariate data sets, rely on arbitrary rules that may exclude valid cases or include invalid cases, and can compromise randomness and reduce sample size (Allen 1966; Lind & Zumbo 1993).

**Proposed algorithm**

Inattentive respondents who are unconcerned about detection may answer in obvious systematic ways, like a Christmas-tree pattern on a mark sense form or the ways simulated by Barnette (1999). To preclude detection, MRs would avoid such distinct patterns; thus, visual inspection would be insufficient to detect many MRs. Nonetheless, the patterns simulated by Barnette (1999) and MRs’ answers should have similar exaggerated mean and variance profiles. The MR algorithm introduced here recognises this similarity.

Outlier analyses compare observations to centroids; in other words, an inter-case analysis comparing each case to all other cases. In contrast, the posited MR algorithm begins with an intra-case assessment of each respondent’s answers. It assumes that MRs, in trying to sabotage studies yet remain undetected, will answer in ways that produce high/low intra-case means or variances relative to other survey participants. Then, like outlier analyses, the algorithm identifies candidate cases for removal based on a threshold criterion. Hence, the MR algorithm differs markedly from outlier analyses and avoids their first two aforementioned limitations.

The Appendix summarises the distribution-free, sample-size-unconstraining, backward-stepping MR algorithm applied here. After
transposing a data set in SPSS, its mean and variance values are entered into a spreadsheet program. Next, one of four criteria is used to remove high- and low-extreme respondents from the data set: 0.50% highest and 0.50% lowest mean response; 0.50% highest and 0.50% lowest variance in responses; 3.0% highest and 3.0% lowest mean response, and 3.0% highest and 3.0% lowest variance in responses.

EFA, CFA and SEM examples

To examine the effects of removing cases with extremely high/low intra-case means and variances, factor analyses were performed on four data sets originally collected for SEM-type testing. The mean sample size of the four data sets is 197.25. The average number of factors and variables studied is 4.75 and 21.25, respectively. All data are based on seven-point Likert-type or semantic differential scales.

Exploratory and confirmatory factor analyses (EFAs and CFAs), with different numbers and types of suspicious cases deleted, were run on each data set. Specifically, EFAs were evaluated on variance explained, factor loadings and cross-loadings; CFAs were evaluated on factor loadings and select model fit indices.

Table 1 summarises the first EFA set, which relied on maximum likelihood estimation, oblique rotation and pairwise deletion. Variance

Table 1  Exploratory factor analysis (1)

<table>
<thead>
<tr>
<th></th>
<th>Data set analysed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>Data set 1</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>174</td>
</tr>
<tr>
<td># of factors and variables</td>
<td>5, 23</td>
</tr>
<tr>
<td>Variance explained</td>
<td>66.73%</td>
</tr>
<tr>
<td># and % of factor loadings &gt;0.7</td>
<td>16, 13.9%</td>
</tr>
<tr>
<td># and % of cross loadings &gt;0.2</td>
<td>20, 17.4%</td>
</tr>
<tr>
<td>Data set 2</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>207</td>
</tr>
<tr>
<td># of factors and variables</td>
<td>4, 19</td>
</tr>
<tr>
<td>Variance explained</td>
<td>73.76%</td>
</tr>
<tr>
<td># and % of factor loadings &gt;0.7</td>
<td>16, 21.1%</td>
</tr>
<tr>
<td># and % of cross loadings &gt;0.2</td>
<td>57, 75%</td>
</tr>
</tbody>
</table>
explained generally decreased as cases were removed; for data sets #2 and #4, variance explained improved minimally when 3.0% of highest/lowest variance cases were removed (i.e. 0.18% and 0.17%, respectively). Factor loadings greater than 0.7 only increased for data set #4 (i.e. from 14 to 15) when 3.0% of highest/lowest variance cases were removed. All other factor loading results showed no improvement. Cross loadings greater than 0.2 decreased meaningfully for data sets #2, #3 and #4 when 3.0% of highest/lowest mean cases were removed (i.e. 9, 22 and 33 fewer, respectively). All other cross loadings results showed no improvement.

Table 2 summarises the second EFA set, which relied on principal components estimation, varimax rotation and pairwise deletion. For data set #1, both highest/lowest mean and the 3.0% highest/lowest variance deletion rules increased factor loadings greater than 0.7 by one (i.e. from 19 to 20). For data set #2, the 0.5% highest/lowest mean and 3.0% highest/lowest variance deletion rules increased factor loadings greater than 0.7 by one and two, respectively (i.e. from 14 to 15 and 16). Cross loadings greater than 0.2 decreased for data sets #2 through #4; specifically, they decreased from 23 to 18 when the 3.0% highest/lowest mean deletion rule was applied to data set #2; from 4 to 3 when both highest/lowest mean deletion rules were applied to data set #3; and from 27 to 25, 27 to 20, and 27 to 26 for both highest/lowest mean and 3.0% highest/lowest variance
deletion rules, respectively, for data set #4. All other cross loadings results showed no improvement.

Table 3 summarises the CFAs. The 3.0% highest/lowest variance deletion rule improved the number of factor loadings greater than 0.70 from 14 to 15 for data set #4; all other results showed no improvement. Hence, case deletion generally did not improve average variance extracted, which parallels the previous variance results. In contrast, case deletion tended to improve model fit; for example, $\chi^2/df$ improved for all data sets under the 3.0% highest/lowest mean deletion rule. Under the 0.5% highest/lowest mean delete rule, the $\chi^2/df$ value also improved slightly from 1.66 to 1.64 for data set #2. The 0.5% highest/lowest variance and 3.0% high/low mean deletion rules improved NNFI, RMSEA and SRMR for data sets #1 and #4, and the 3.0% highest/lowest variance deletion rule improved RMSEA for data set #2. Of greatest interest is that both 0.5% deletion rules meaningfully improved the fit of the poor-fitting model (Hu & Bentler 1999) from data set #1. Perhaps the fit of poor-fitting models can be improved substantially by removing data from the most likely MRs.

**Table 2** Exploratory factor analysis (2)

<table>
<thead>
<tr>
<th>Data set analysed</th>
<th>Total</th>
<th>0.5% H/L mean deleted</th>
<th>3.0% H/L mean deleted</th>
<th>0.5% H/L variance deleted</th>
<th>3.0% H/L variance deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data set 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>174</td>
<td>172</td>
<td>162</td>
<td>172</td>
<td>162</td>
</tr>
<tr>
<td># and % of factor loadings &gt;0.7</td>
<td>19, 16.5%</td>
<td>20, 17.39%</td>
<td>20, 17.39%</td>
<td>19, 16.5%</td>
<td>20, 17.39%</td>
</tr>
<tr>
<td># and % of cross loadings &gt;0.2</td>
<td>9, 7.8%</td>
<td>9, 7.8%</td>
<td>12, 10.43%</td>
<td>9, 7.8%</td>
<td>9, 7.8%</td>
</tr>
<tr>
<td><strong>Data set 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>207</td>
<td>205</td>
<td>195</td>
<td>205</td>
<td>195</td>
</tr>
<tr>
<td># and % of factor loadings &gt;0.7</td>
<td>14, 18.4%</td>
<td>15, 19.7%</td>
<td>14, 18.4%</td>
<td>14, 18.4%</td>
<td>16, 21.05%</td>
</tr>
<tr>
<td># and % of cross loadings &gt;0.2</td>
<td>23, 30.3%</td>
<td>23, 30.3%</td>
<td>18, 23.68%</td>
<td>26, 34.21%</td>
<td>24, 31.57%</td>
</tr>
<tr>
<td><strong>Data set 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>191</td>
<td>189</td>
<td>179</td>
<td>189</td>
<td>179</td>
</tr>
<tr>
<td># and % of factor loadings &gt;0.7</td>
<td>22, 15.27%</td>
<td>22, 15.27%</td>
<td>22, 15.27%</td>
<td>22, 15.27%</td>
<td>22, 15.27%</td>
</tr>
<tr>
<td># and % of cross loadings &gt;0.2</td>
<td>4, 2.7%</td>
<td>3, 2.08%</td>
<td>3, 2.08%</td>
<td>6, 4.16%</td>
<td>7, 4.86%</td>
</tr>
<tr>
<td><strong>Data set 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>217</td>
<td>215</td>
<td>205</td>
<td>215</td>
<td>205</td>
</tr>
<tr>
<td># and % of factor loadings &gt;0.7</td>
<td>16, 21.05%</td>
<td>15, 19.73%</td>
<td>14, 18.42%</td>
<td>16, 21.05%</td>
<td>16, 21.05%</td>
</tr>
<tr>
<td># and % of cross loadings &gt;0.2</td>
<td>27, 35.52%</td>
<td>25, 32.89%</td>
<td>20, 26.31%</td>
<td>28, 36.84%</td>
<td>26, 34.21%</td>
</tr>
</tbody>
</table>

Note: Principal components estimation, varimax rotation and pairwise deletion were used. Data were taken from the rotated component matrix.
Table 3  Confirmatory factor analysis

<table>
<thead>
<tr>
<th>Data set analysed</th>
<th>Total</th>
<th>0.5% H/L mean deleted</th>
<th>3.0% H/L mean deleted</th>
<th>0.5% H/L variance deleted</th>
<th>3.0% H/L variance deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data set 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>174</td>
<td>172</td>
<td>162</td>
<td>172</td>
<td>162</td>
</tr>
<tr>
<td>NNFI</td>
<td>0.87</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>CFI</td>
<td>0.89</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.103</td>
<td>0.063</td>
<td>0.062</td>
<td>0.064</td>
<td>0.066</td>
</tr>
<tr>
<td>SRMR</td>
<td>0.078</td>
<td>0.064</td>
<td>0.067</td>
<td>0.064</td>
<td>0.065</td>
</tr>
<tr>
<td>$\chi^2/df$</td>
<td>2.82</td>
<td>1.68</td>
<td>1.61</td>
<td>1.70</td>
<td>1.72</td>
</tr>
<tr>
<td># and % of factor loadings &gt;0.7</td>
<td>17, 73.91%</td>
<td>15, 65.22%</td>
<td>15, 65.22%</td>
<td>15, 65.22%</td>
<td>15, 65.22%</td>
</tr>
<tr>
<td><strong>Data set 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>207</td>
<td>205</td>
<td>195</td>
<td>205</td>
<td>195</td>
</tr>
<tr>
<td>NNFI</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>CFI</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.056</td>
<td>0.056</td>
<td>0.057</td>
<td>0.058</td>
<td>0.051</td>
</tr>
<tr>
<td>SRMR</td>
<td>0.047</td>
<td>0.050</td>
<td>0.056</td>
<td>0.049</td>
<td>0.048</td>
</tr>
<tr>
<td>$\chi^2/df$</td>
<td>1.66</td>
<td>1.64</td>
<td>1.63</td>
<td>1.68</td>
<td>1.51</td>
</tr>
<tr>
<td># and % of factor loadings &gt;0.7</td>
<td>15, 78.94%</td>
<td>15, 78.94%</td>
<td>14, 73.68%</td>
<td>15, 78.94%</td>
<td>15, 78.94%</td>
</tr>
<tr>
<td><strong>Data set 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>191</td>
<td>189</td>
<td>179</td>
<td>189</td>
<td>179</td>
</tr>
<tr>
<td>NNFI</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>CFI</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.053</td>
<td>0.053</td>
<td>0.051</td>
<td>0.055</td>
<td>0.054</td>
</tr>
<tr>
<td>SRMR</td>
<td>0.056</td>
<td>0.058</td>
<td>0.059</td>
<td>0.057</td>
<td>0.058</td>
</tr>
<tr>
<td>$\chi^2/df$</td>
<td>1.53</td>
<td>1.53</td>
<td>1.47</td>
<td>1.56</td>
<td>1.52</td>
</tr>
<tr>
<td># and % of factor loadings &gt;0.7</td>
<td>18, 75%</td>
<td>17, 70.83%</td>
<td>17, 70.83%</td>
<td>17, 70.83%</td>
<td>18, 75%</td>
</tr>
<tr>
<td><strong>Data set 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>217</td>
<td>215</td>
<td>205</td>
<td>215</td>
<td>205</td>
</tr>
<tr>
<td>NNFI</td>
<td>0.97</td>
<td>0.97</td>
<td>0.96</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>CFI</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.074</td>
<td>0.075</td>
<td>0.076</td>
<td>0.064</td>
<td>0.058</td>
</tr>
<tr>
<td>SRMR</td>
<td>0.049</td>
<td>0.051</td>
<td>0.059</td>
<td>0.048</td>
<td>0.045</td>
</tr>
<tr>
<td>$\chi^2/df$</td>
<td>2.19</td>
<td>2.19</td>
<td>2.18</td>
<td>1.87</td>
<td>1.69</td>
</tr>
<tr>
<td># and % of factor loadings &gt;0.7</td>
<td>14, 73.68%</td>
<td>14, 73.68%</td>
<td>11, 57.89%</td>
<td>14, 73.68%</td>
<td>15, 78.95%</td>
</tr>
</tbody>
</table>

**Conclusion**

Because survey participants may be motivated to respond mischievously for many reasons, and estimates of MR incidence differ meaningfully from zero, attitude researchers should ensure their statistical results are
not an artifact of MR data. If a small proportion of outliers can distort scale reliabilities, correlation coefficients, factor structures and structural equation models, then a similar proportion of MR cases may prove equally problematic. Outlier detection relies on inter-case analyses; in contrast, the posited algorithm relies on intra-case computations followed by a casewise deletion rule.

As Barnette (1999) notes, ‘It clearly is desirable to remove surveys that add nothing but error into the data set’ (p. 45). The issue is how best to accomplish this task. Our results suggest that deleting likely MR cases from attitude data may improve EFA and CFA results. Although variance explained did not improve, cross loadings greater than 0.2 decreased meaningfully for three of four data sets. Hence, deleting likely MR cases may create more stable and readily interpreted factors. In addition, our results suggest that eliminating these cases improves CFA model fit; for both good- and poor-fitting models, fit statistics improved. In essence, ridding MR cases from data sets may not always lead to substantial empirical improvement for EFA, CFA and SEM techniques. Nonetheless, the proposed algorithm can identify likely outlier MRs, cleanse data files of these cases, and – if our results plus the extant literature are indicative (e.g. Huber 1981; Bollen 1987; Lind & Zumbo 1993; Barnette 1999) – improve statistical results.

Limitations and future research

The generalisability of our conclusions is constrained by the number and nature of the evaluated data sets. The four data sets included roughly 200 often-student respondents; thus, similar analyses on larger non-student data sets are needed (Winer 1999). Evaluating data sets with more than five factors and 25 variables would ensure our findings pertain to larger models. As only one of the four data sets originally yielded a poor fitting model, additional research would ensure that the fit of poor-fitting models improves with casewise deletion of likely MRs.

Additional research could address the following issues.

• Creating improved algorithms for post hoc identification of MRs. The basic patterns delineated by Barnette (1999) may not include all detectable patterns. For example, MRs could answer like other respondents, only more extremely. Because ‘outliers with the same shape as the main data are in some sense the hardest to find’ (Rocke & Woodruff 1996, p. 1048), creating an algorithm sensitive to this
pattern may prove challenging. Simple algorithms based on item-based outlier scores, in which data from respondents who provide lesser popular answers or answer pairs are removed (Zijlstra et al. 2007), may suffice. Conversely, more sophisticated algorithms based on item response theory may be required (Zickar & Robie 1999; De Jong et al. 2008). Algorithms based on clustering or Q-type factor analysis may mitigate the swamping (i.e. false positives) and masking (i.e. larger outliers obscure smaller outliers) problems common to multiple outlier detection schemes (Atkinson 1994; Becker & Gather 1999). Rather than the backward-stepping deletion algorithm demonstrated here, forward-stepping algorithms, which add cases by proximity to the fitted model and use plots to identify aberrant observations, may prove superior (Atkinson 1994; Poon & Wong 2004; Mavridis & Moustaki 2008). Computationally tractable graphical methods also may be possible (Atkinson 1994; Poon & Wong 2004; Peña & Prieto 2007).

- **Identifying deletion rules** – by population of interest, data collection method and other study characteristics – that best balance bad data elimination with sample representativeness. For MR classification, the optimal ratio of Type I to Type II errors depends on the relative bias introduced by phoney responses versus smaller and less representative samples. As the deletion thresholds illustrated here were conservative, the efficacy of thresholds that remove more than 3% of questionable cases should be tested (Barnette 1999). More sophisticated deletion rules could consider differences in MR incidence by subgroup and profile variables that distinguish MRs from other respondents. As a result, the threshold for deleting suspicious cases may drop for higher-propensity groups and rise for lower-propensity groups. Alternatively, suspicious cases could be down-weighted by MR-likelihood or influence rather than deleted outright (Huber 1981; Yuan & Bentler 1998, 2001).

- **Identifying questionnaire designs** – especially instructions – that minimise MR incidence. The literatures on response bias, faking and inattentive respondents provide strong candidates for evaluation. For example, researchers might minimise MR incidence through subtle warnings that the questionnaire permits MR detection.
• **Determining relative and joint efficacies of purging algorithms and robust statistics in mitigating MR-induced bias.** Both the posited algorithm and the use of robust statistics are post hoc methods for reducing bias. Using one does not preclude using the other. Future research can determine if one approach is preferred or both approaches are complementary.

### Appendix: Data transposition steps

In SPSS 16.0:

1. Select Data in the drop down menu
2. Select Transpose
3. Select variables of interest; click right arrow; click OK
   (Note: A new Data View spreadsheet, which will need to be renamed and saved, will open with the transposed variables; use this spreadsheet for subsequent analyses in SPSS)
4. Select Analyze in the drop down menu; select Descriptive Statistics; select Descriptives
5. Select variables; click the right arrow; select Options; select Variance and Mean; click Continue; click OK
6. From the SPSS Viewer Output, Copy and Paste the Table, including the Variables, and the Mean, and Variance values into an Excel spreadsheet.

In Excel (Microsoft Office 2007):

1. Delete all columns and rows except for Variables, Mean, and Variance
2. Place the cursor on Var001
3. Select Data from the drop down menu; select Sort
4. In the Sort By drop down menu, select Mean or Variance; click OK
5. Examine High/Low Mean or Variance values.

Finally:

1. Delete Case Numbers in the SPSS file
2. Perform desired factor analysis or other techniques.
References


**About the authors**

Michael R. Hyman (PhD, Purdue University) is Stan Fulton Chair of Marketing at New Mexico State University. He has (co)authored more than 75 journal articles (in outlets such as *Journal of Marketing, Journal of the Academy of Marketing Science, Journal of Business Research, Journal of Retailing, Journal of Advertising, Journal of Marketing Theory & Practice* and *Journal of Business Ethics*), 45 conference proceedings papers (including nine ‘best paper award’ winners), three books (including *Marketing Research Kit for Dummies*), and six book chapters. His current research interests include consumers’ responses to advertising, ethics in marketing, survey research methods, knowledge acquisition in academia, and philosophical analyses in marketing.

Jeremy J. Sierra (PhD, New Mexico State University) is Associate Professor of Marketing at Texas State University-San Marcos. He has published 16 journal articles (in outlets such as *Journal of Advertising, Journal of Current Issues & Research in Advertising, Journal of Marketing Education, Journal of Marketing Theory and Practice* and *Journal of Services Marketing*) and 19 conference proceedings papers (including two ‘best paper award’ winners). His research interests include advertising effects, consumer behaviour, marketing ethics, and services marketing.

Address correspondence to: Michael R. Hyman, Stan Fulton Professor of Marketing, College of Business, New Mexico State University, Las Cruces, NM 88003-8001.

Email: mhyman@nmsu.edu