Small sample sizes, overextraction, and unrealistic expectations
A commentary on M. Mattsson

de Winter, J.C.F.

DOI
10.1016/j.aap.2012.07.003

Publication date
2013

Document Version
Final published version

Published in
Accident Analysis & Prevention

Citation (APA)
de Winter, J. C. F. (2013). Small sample sizes, overextraction, and unrealistic expectations: A commentary on M. Mattsson. *Accident Analysis & Prevention*, *50*, 776-777. https://doi.org/10.1016/j.aap.2012.07.003

Accident Analysis and Prevention 50 (2013) 776–777

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Short communication

Small sample sizes, overextraction, and unrealistic expectations: A commentary on M. Mattsson

J.C.F. de Winter

Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands

ARTICLE INFO

Article history:
Received 9 May 2012
Received in revised form 28 June 2012
Accepted 1 July 2012

Keywords:
Driver Behaviour Questionnaire
Commentary
Factorial invariance
Errors and violations

ABSTRACT

In a recent article about the Manchester Driver Behaviour Questionnaire (DBQ), Mattsson (2012) concluded that the factor structure was not invariant across subgroups of respondents. This commentary contests this conclusion.

Following a meticulous analysis of the Manchester Driver Behaviour Questionnaire (DBQ) across subgroups of various ages and both genders, Mattsson (2012) concluded that "the factor structure was not invariant across subgroups of respondents and should thus not be used as a basis for multiple-group comparisons (such as comparing the mean values of factors among the subgroups)" (p. 395). Mattsson further reported that item loading patterns showed differences between men and women, and that, therefore, "some of the results of the pooled sample of respondents could even be considered artifacts of the analysis procedure" (p. 391).

1. This commentary contests these findings

1.1. Small sample sizes

Mattsson's sample sizes are too small to draw reliable conclusions on factorial patterns. A simulation study by De Winter et al. (2009) showed that for an *ideal* simple structure of 24 variables, four factors, and loadings of 0.4 using multinormal data, a sample size of at least 349 is needed to obtain adequate congruence with the population solution. De Winter et al. (2009) also showed that congruence drastically deteriorates when one or more pairs of factors are substantially correlated (≥.5) or when model error is present. Mattsson reported factor correlations as high as .61 (e.g., between slips and lapses) but used sample sizes as low as 233. Thus, it is likely that Mattsson's conclusions are themselves artifacts, caused by failing to recover stable factors.

1.2. Overextraction

Mattsson extracted four factors and argued against a two-factor solution: "The errors/violations distinction found to be stable across DBQ studies by De Winter and Dodou (2010) is too general to be of interest in this sense" (p. 391). However, there is overwhelming theoretical support for a higher-order distinction between errors and violations. Reason et al. (1990) have proposed a two-factor solution in their original DBQ research: "In considering the human contribution to accidents, it seems necessary to make a distinction between errors and violations; two forms of aberration which may have different psychological origins" (p. 1315). Even before the first DBQ was published, the distinction between errors and violations was a research theme at Manchester University. For example, in 1988, Reason published a paper entitled "Errors and violations: The lessons of Chernobyl" (Reason, 1988). The distinction between errors and violations resembles other well-established taxonomies in the traffic psychology literature, such as driving skill versus driving style (Elander et al., 1993) and driver performance versus driver behavior (Evans, 2004). Using the same data as Mattsson, Lajunen et al. (2004) also extracted errors and violations as two higher-order factors.

1.3. Unrealistic expectations

Mattsson set unrealistic expectations by demanding that "factor means of subgroups of respondents should not be compared unless strong factorial invariance can be demonstrated" (p. 391). Meredith and Teresi (2006) explained that "strong and strict invariance may be less important in the context of basic research in which group differences in specific factors are indicative of individual differences

E-mail address: j.c.f.dewinter@tudelft.nl

0001-4575/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2012.07.003

Table 1
Eigenvalues of correlation matrix of female sample (n = 5769) and male sample (n = 3329).

Eigenvalue	Female sample	Male sample
1	7.16	7.18
2	2.62	2.73
3	1.32	1.49
4	1.25	1.36
5	1.21	1.25
6	1.12	1.14
7	1.10	1.09

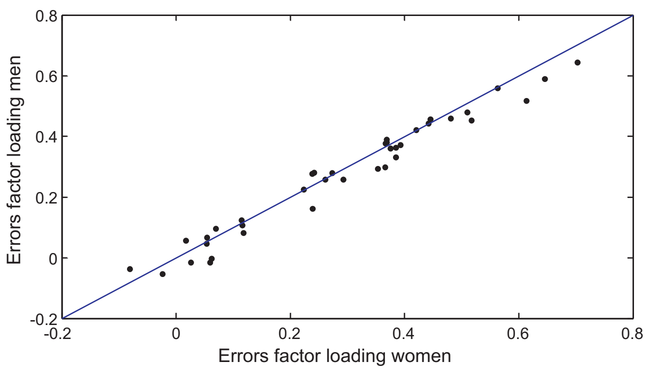

Fig. 1. Errors factor loading based on a sample of women (n = 5769) versus errors factor loading based on a sample of men (n = 3329). The highest loading for the errors factor was found for the item: "Drive in either too high or too low a gear for the conditions".

that are important for scientific exploration" (p. 69). Looking at the totality of the evidence, versions of the DBQ have been used in female populations (Dobson et al., 1999), male populations (e.g., Shahar, 2009; Sümer et al., 2005), adolescents (Brookland et al., 2008), and old drivers (Schwebel et al., 2007), all yielding interpretable results. Furthermore, objective gender and age effects in driver behaviors (e.g., accident statistics, speeding infringements) are well established. Therefore, arguing against comparing means of these subgroups of respondents is an unnecessarily skeptical standpoint.

1.4. Similarity of factor loadings for a two-factor solution

Here I provide evidence that factor loadings of males and females are markedly similar. I calculated factor loadings for a female sample (n = 5769) and male sample (n = 3329), by extracting two factors from a 38-item DBQ, using exploratory factor analysis with principal axis factoring and direct oblimin rotation (data from the cohort of novice drivers of TRL (2008), see also Wells et al., 2008, selecting those items asking "When driving, how often do you do each of the following?" on a scale from 1 = *never* to 6 = *nearly all the time*). The factor loadings for men and women were highly correlated (r = .98 for both factors), see Fig. 1 for an illustration. The inter-factor correlation was .33 for the female sample, as well as for the male sample. Eigenvalues of the correlation matrices (Table 1) clearly support a two-factor solution, while providing no support for a four-factor solution.

Bartlett factor scores were virtually identical to the Bartlett factor scores based on the loadings of the opposite sex (r above .99 for both factors and for both genders). The similarity of factor scores is in accordance with Wilks' theorem, which states that the correlation between two linear composites approaches unity as the number of positively correlated variables is large, irrespective of the weights that are used (Wilks, 1938; see also Jensen and Weng, 1994; Ree et al., 1998). It is also known that the results of principal components analysis and factor analysis are highly similar, especially with a large number of variables, so Mattsson's criticism of the use of PCA is invalid from a practical perspective (cf., Velicer and Jackson, 1990).

2. Conclusion

I argued that Mattsson's results are contaminated by a serious degree of sampling error, that he extracted too many factors, and that his analyses do not undermine the validity of the DBQ. Mattsson stated that the DBQ factor loadings are substantially different between both genders, but the present results indicate that loadings for male and female novice drivers are highly similar in a parsimonious two-factor solution. Despite the fact that Mattsson's work has some important limitations, it has to be acknowledged that his work has opened up avenues for investigating factorial invariance, and for improving and standardizing the DBQ.

References

De Winter, J.C.F., Dodou, D., 2010. The Driver Behaviour Questionnaire as a predictor of accidents: a meta-analysis. Journal of Safety Research 41, 463–470.

De Winter, J.C.F., Dodou, D., Wieringa, P.A., 2009. Exploratory factor analysis with small sample sizes. Multivariate Behavioral Research 44, 147–181.

Dobson, A., Brown, W., Ball, J., Powers, J., McFadden, M., 1999. Women drivers' behaviour, socio-demographic characteristics and accidents. Accident Analysis and Prevention 31, 525–535.

Elander, J., West, R., French, D., 1993. Behavioral correlates of individual differences in road-traffic crash risk: an examination of methods and findings. Psychological Bulletin 113, 279–294.

Brookland, R., Begg, D., Langley, J., Ameratunga, S., 2008. Parent and adolescent risky driving behaviours: New Zealand drivers study. Australasian College of Road Safety Journal 20, 52–59.

Evans, L., 2004. Traffic Safety. Science Serving Society, Bloomfield Hills, MI.

Jensen, A.R., Weng, L.-J., 1994. What is a good g? Intelligence 18, 231–258.

Lajunen, T., Parker, D., Summula, H., 2004. The Manchester Driver Behaviour Questionnaire: a cross-cultural study. Accident Analysis and Prevention 36, 231–238.

Mattsson, M., 2012. Investigating the factorial invariance of the 28-item DBQ across genders and age groups: an Exploratory Structural Equation Modeling study. Accident Analysis and Prevention 48, 379–396.

Meredith, W., Teresi, J.A., 2006. An essay on measurement and factorial invariance. Medical Care 44, 69–77.

Reason, J., 1988. Errors and violations: the lessons of Chernobyl. In: Conference Record for 1988 IEEE Fourth Conference on Human Factors and Power Plants, pp. 537–540.

Reason, J.T., Manstead, A., Stradling, S., Baxter, J., Campbell, K., 1990. Errors and violations on the roads: a real distinction? Ergonomics 33, 1315–1332.

Ree, M.J., Carretta, T.R., Earles, J.A., 1998. In top-down decisions, weighting variables does not matter: a consequence of Wilks' theorem. Organizational Research Methods 1, 407–420.

Shahar, A., 2009. Self-reported driving behaviors as a function of trait anxiety. Accident Analysis and Prevention 41, 241–245.

Schwebel, D.C., Ball, K.K., Severson, J., Barton, B.K., Rizzo, M., Viamonte, S.M., 2007. Individual difference factors in risky driving among older adults. Journal of Safety Research 38, 501–509.

Sümer, N., Ayvasik, B.H., Er, N., 2005. Cognitive and psychomotor correlates of self-reported driving and behaviour. In: Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, pp. 96–103.

Transport Research Laboratory, Safety, Security and Investigations Division [TRL], 2008. Cohort II: A Study of Learner and Novice Drivers, 2001–2005 [Computer file]. Colchester, Essex: UK Data Archive [distributor], July 2008. SN: 5985.

Velicer, W.F., Jackson, D.N., 1990. Component analysis versus common factor analysis: some further observations. Multivariate Behavioral Research 25, 97–114.

Wells, P., Tong, S., Sexton, B., Grayson, G., Jones, E., 2008. Cohort II: A Study of Learner and New Drivers. Volume 1. Main Report (Report no. 81). London: Department for Transport.

Wilks, S.S., 1938. Weighting systems for linear functions of correlated variables when there is no dependent variable. Psychometrika 3, 23–40.