Measurement invariance of the Driver Behavior Questionnaire across samples of young drivers from Finland and Ireland

Markus Mattsson [a,*], Fearghal O'Brien [b], Timo Lajunen [c], Michael Gormley [b], Heikki Summala [a]

[a] Traffic Research Unit, Institute of Behavioural Sciences, University of Helsinki, Finland
[b] School of Psychology, Trinity College Dublin, Ireland
[c] Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

ARTICLE INFO

ABSTRACT

This article investigates the factor structure of the 27-item Driver Behavior Questionnaire (DBQ) in two samples of young drivers (18–25 years of age); one from Finland and the other from Ireland. We compare the two-, three-, and four-factor solutions using Confirmatory Factor Analysis (CFA) and show that the four-factor model (with the latent variables rule violations, aggressive violations, slips and lapses) fits the data from the two countries best. Next, we compare the fit of this model across samples by the means of a measurement invariance analysis in the CFA framework. The analysis shows that the four-factor model fails to fit both samples equally well. This is mainly because the socially-oriented latent variables (rule violations and aggressive violations) are different in nature in the two samples. The cognitively-oriented latent variables (slips and lapses) are, however, similar across countries and the mean values of slips can be compared using latent variable models. However, the common practice of calculating sum scores to represent the four latent DBQ variables and comparing them across subgroups of respondents is unfounded, at least when comparing young respondents from Finland and Ireland.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The Driver Behavior Questionnaire (DBQ) is perhaps most commonly used psychometric instrument in traffic psychology, with roughly 200 studies being published by 2010 (De Winter and Dodou, 2010). The DBQ is most commonly assumed to measure from two to four latent variables, though factor structures embodying anything from one (Hennessy and Wiesenthal, 2005) to seven (Kontogiannis et al., 2002) factors have been published. In this study, we investigate the cross-cultural equality of the three most commonly used factor structures, namely the two-, three-, and four-factor solution in two samples of young drivers, one collected in Finland and the other one in Ireland.

The two-factor model represents the fundamental distinction between *unintentional errors* and *intentional violations*.[1] The meta-analysis of De Winter and Dodou (2010) showed that these two factors can be used as common denominators for the various factor

structures encountered in the literature. This is a noteworthy finding because the instrument comes in many versions, comprising anything from 10 (Martinussen et al., 2013) to 112 (Kontogiannis et al., 2002) items. The basic distinction between voluntary and involuntary forms of traffic behavior has its roots in the theory of errors presented in Reason (1990).

The three-factor model, on the other hand, is derived from the primary study of the DBQ (Reason et al., 1990). In that study, a five-factor structure was hypothesized to underlie the individual items. The structure of the questionnaire was investigated using principal components analysis (PCA), which resulted in a three-component solution of *involuntary errors, involuntary lapses* and *intentional violations*. *Errors* were judged by the researchers as "potentially dangerous" in contrast to *lapses*, which were characterized as "not dangerous" or "silly". It is of historical interest to note that the three-factor structure of the DBQ is based on the results of the PCA carried out by Reason et al. (1990), rather than being derived from the underlying theory (Reason, 1990). In subsequent DBQ studies some of the individual items were dropped (Parker et al., 1995; Lawton et al., 1997; Åberg and Rimmö, 1998) and others added (Lawton et al., 1997). In the resulting 28-item version of the questionnaire, the two factors related to involuntary *errors* can

---

[1] In this article, we refer to latent variables using *italics*.

perhaps be interpreted as attention-related *slips* and memory-related *lapses* (Mattsson, 2012) in accordance with the theory upon which the DBQ was originally based (Reason, 1990).

The four-factor structure of the DBQ results from dividing the subscale of *violations* into *rule violations* and *aggressive violations* (Lawton et al., 1997). The resulting questionnaire, which is also used in the present study, consists of eight items that are assumed to load on a *lapses* factor, nine on a *rule violations* factor, eight on a factor variously referred to as *errors* or *slips* and three on an *aggressive violations* factor.

In this study, we use modern structural equation modeling and factor analytical methods to investigate whether the same factor structure can be used in explaining the patterns of intercorrelations among the questionnaire items in Finnish and Irish samples of young drivers. In particular, we examine whether one of the three factor solutions fits the data collected from young, inexperienced drivers in one or both of the two countries. Methodologically, the present contribution is based on the measurement invariance framework that has thus far been little used in traffic psychology. Additionally, new methods of visualizing the results are utilized.

Previous studies have investigated the cross-cultural stability of the DBQ factor structures and the four-factor solution has been found to be more or less stable across countries (Lajunen et al., 2004; Özkan et al., 2006). In these studies, the factor structures were compared by examining the factor loading matrices and calculating various indices of approximate factor similarity, such as identity, additivity, proportionality and correlation coefficients (van de Vijver and Leung, 1997). The values of these indices ranged from 0.85 to 0.98 when comparing Finnish, Dutch and British data (Lajunen et al., 2004). However, no statistical test is associated with these indices of factor similarity and there remains an element of subjective judgment on which values of the indices to consider "large" and which ones "small". In addition, it is known that Tucker's phi[2] values of over 0.9 may well be obtained even when the factor structures are actually dissimilar across groups (van de Vijver and Leung, 1997).

In addition, competing factor models (the two-, three- and four-factor solutions) were not compared in the studies of Lajunen et al. (2004) and Özkan et al. (2006). The meta-analysis by De Winter and Dodou (2010) argued for the two-factor solution while the studies by Lajunen et al. (2004) and Özkan et al. (2006) stated that the four-factor model offers a good fit across countries and traffic cultures. Then again, the original study by Reason et al. (1990) and, for instance, the more recent study Davey et al. (2007) concluded that the three-factor (or three-component) solution fits the data best. A formal evaluation of the issue across cultures is in order.

This study builds on these earlier studies and complements them by utilizing modern structural equation modeling tools in comparing the three measurement models across two countries, Finland and Ireland. In the first stage of the analysis, the 2-, 3- and 4-factor models were fit to the two samples separately in order to find the one with the best fit. In the second stage, the model chosen in the first stage was fit to the two samples simultaneously and the differences in model fit were evaluated by analyses of measurement invariance. In short, our research questions were:

1. Which of the three competing models fits the two samples of data best? Specifically, is the model at issue the same or different in the two samples? If the latter question is answered in the affirmative, we proceed to investigate research question 2, i.e.,
2. In what respects is the best-fitting model comparable across samples? The analysis proceeds in distinct stages, i.e.,

(2.1) Are the factors themselves identical?
(2.2) Are the factor loadings identical?
(2.3) Are item intercepts identical?
(2.4) Are item error variances identical?

The statistical analyses that were used to answer these questions are described in detail in Section 2.3.

2. Materials and methods

2.1. Participants and data

In the present study, Finnish and Irish data on the driving behavior of young drivers (18–25 years of age) was compared. The Finnish data set consisted of a sample of 1051 young drivers with an overall response rate of 35.3%. The sample was collected as a stratified random sample from the driving license register. The respondents were enrolled in a lottery with two 250 euro pecuniary rewards as incentives to participate. Comparison of the responders and non-responders indicated that the two groups did not differ in terms of penalties received for reckless driving or driving under the influence of alcohol. The mean age of the Finnish respondents was 20.6 years, and median age 20. Other characteristics of the sample are presented in Table 1. Cases with missing values in DBQ variables 1–9 or 11–19 were removed from the data because this pattern of missing values was likely due to the respondent not realizing that the questionnaire continued on a different page.

The patterns of missing values in the DBQ variables were investigated using the Missing Values Analysis (MVA) procedure in SPSS (IBM Corp., 2012). The analysis showed that the number of missing values varied between zero and 12, which amounts to 0–1.1% of the total number of values. Little's MCAR test showed that the values were missing completely at random $\chi^2(3438, N = 1051) = 3506.45$, $p = 0.204$ with respect to the variables gender, age, the time that the respondent had possessed a driver's license, exposure (kilometers driven per month) and whether the respondent had been involved in an accident. The missing values were not imputed because the Full Information Maximum Likelihood (FIML) estimation procedure in the R (R Development Core Team, 2013) package *lavaan* (Rosseel, 2012) was used when performing the analyses.

In contrast to the Finnish sample, the Irish sample was collected using an online questionnaire. The respondents were acquired from among college students at Trinity College, Dublin and people visiting a number of online car forums, or car sections of general interest online forums. The respondents from the college completed the questionnaire in response to an email sent around their college department by a member of administration while forum respondents were notified through a general post. Participants were entered into a lottery for a €50 gift voucher. As the online system did not allow the user to continue before answering all the items, the Irish data set contained no missing values. The data set consisted of 816 drivers with mean age of 20.3, and median age of 20. Respondents'other characteristics are presented in Table 1.

Table 1
Respondents' characteristics.

	Country	
	Finland	Ireland
n	1051	816
Sex (percent female)	62.5	53.6
Mean years license held (sd)	2.44 (1.71)	NA
License type (percent provisional/full)	NA	37.6/62.4

---

[2] One of the similarity indices.