Accident Analysis and Prevention 78 (2015) 185–200

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

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

Article history:
Received 22 December 2014
Received in revised form 9 February 2015
Accepted 26 February 2015
Available online 19 March 2015

Keywords:
Driver Behavior Questionnaire
DBQ
CFA
Confirmatory Factor Analysis
Measurement invariance

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.

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

* Corresponding author at: P.O. Box 9, Siltavuorenpenger 1 A, FI-00014 University of Helsinki, Finland. Tel.: +358 40 7689406; fax: +358 9 19129422.
E-mail address: markus.mattsson@helsinki.fi (M. Mattsson).

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

http://dx.doi.org/10.1016/j.aap.2015.02.017
0001-4575/

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] One of the similarity indices.

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

2.2. Measures and model specification

The 28-item questionnaire (Lawton et al., 1997) and its Finnish translation (Lajunen et al., 2004) served as the basis of the current study. The item related to driving under the influence of alcohol was removed as recommended by Lajunen et al. (2004). The 27-item version of the questionnaire thus obtained was used in the present study.

The English version of the questionnaire (which was used in the Irish sample) is included as Appendix A. The DBQ variables were measured on a six-point Likert scale ranging from 1 ("Never") to 6 ("Nearly all the time"). The distributions of the DBQ variables in the Finnish sample are presented in Fig. 1 and those in the Irish sample in Fig. 2. The figures make it clear that the observed variables were not distributed normally.

In the two-factor model, the latent variables *violations* and *errors* were assumed to underlie the observed variables. In the three-factor solution, the *violations* factor was assumed identical to that of the two-factor solution. However, the *errors* factor was split into *slips* and *lapses*. For the four-factor solution, the assumed factor loadings on the factors of *slips* and *lapses* were identical to those of the three-factor solution. However, the latent variable *violations* was split into *rule violations* and *aggressive violations*. The specific items that were assumed to load on each factor in each solution are reported in Appendix A.

The metric of the latent variables was set by using the method proposed by Little et al. (2006). This method involves setting the average loading of the indicator variables to unity and the average intercept term to zero. No cross-loadings or correlated indicator errors were specified in the initial models. The correlation matrices between the observed variables are reported in Appendix B so that our analyses can be replicated. The presence of missing data in the Finnish sample will, however, produce slight differences between results obtained from using raw data and correlation matrices as input.

2.3. Statistical analyses

The analysis of measurement invariance consists of fitting a sequence of models with increasingly restrictive constraints on the parameters of the model: (1) model with no constraints fit to both samples simultaneously (configural invariance), (2) constraining factor loadings to equality across groups (weak invariance), (3) constraining item intercepts to equality (strong invariance) and (4) constraining item errors to equality (strict invariance). An excellent summary of the stages of an analysis of measurement invariance can be found in Gregorich (2006) and especially Fig. 1 therein.

Passing or failing the test at each stage has direct practical consequences for the model. The first two tests are concerned with the qualitative similarity of the interpretations given to the items across groups. If configural invariance is reached in the first stage, the items load on the same factors in the two samples. Passing the second test can be interpreted as showing that respondents being compared assign similar meanings to the latent constructs (Gregorich, 2006).

The remaining two tests indicate whether factor means (stage 3) and composite means (stage 4) can be meaningfully compared across groups. The test of strong factorial invariance is related to testing the equality of the intercept terms of the items. Inequality of intercepts indicates the presence of unequal systematic, additive effects across groups. For instance, different levels of social

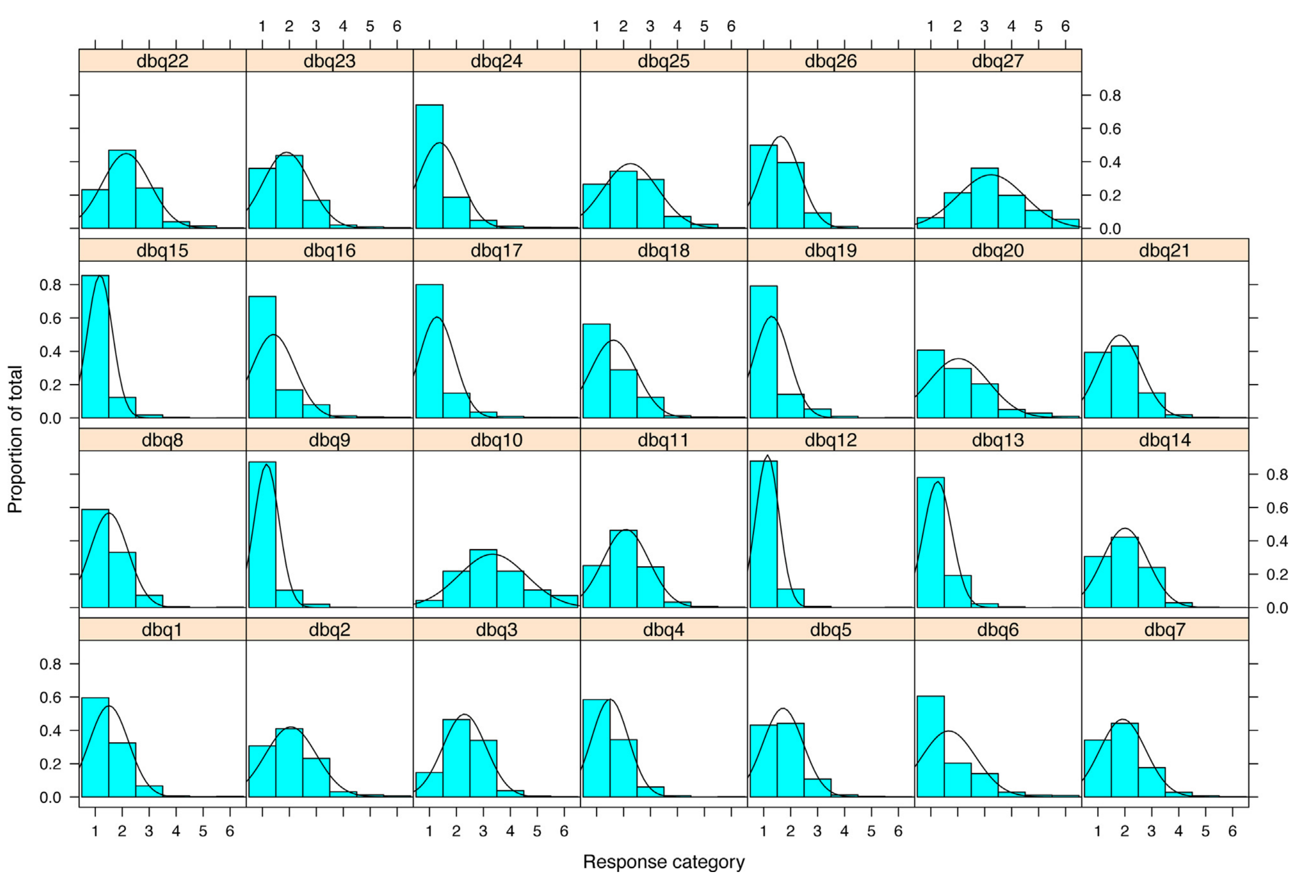

Fig. 1. Distributions of the DBQ variables in the Finnish sample.

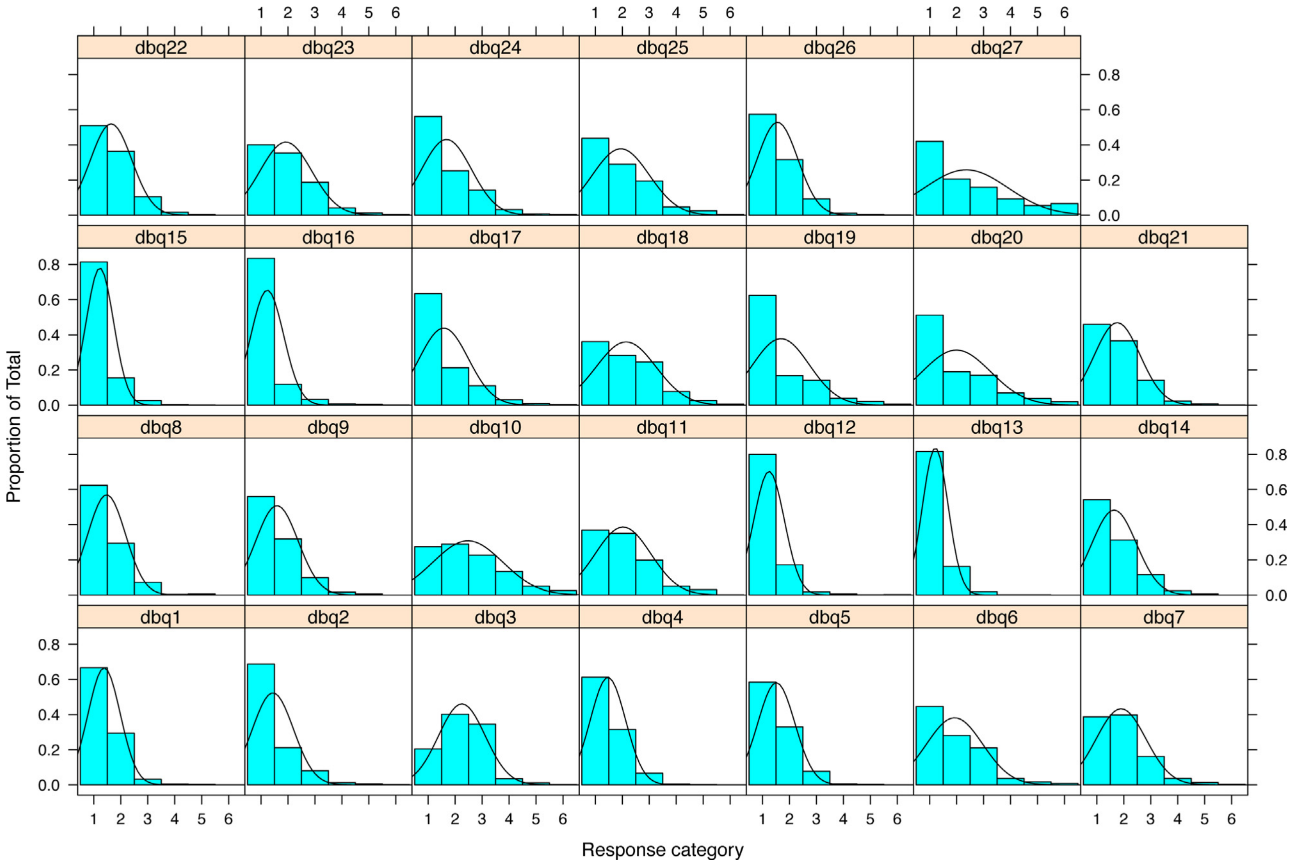

Fig. 2. Distributions of the DBQ variables in the Irish sample.

desirability of certain behaviors could lead to respondents from the two countries responding differently to an item even if they have identical values on the latent variable. This type of item functioning is also known as differential additive bias.

The last stage of the analysis (stage 4) is related to testing the equality of error variances. According to the theory of factor analysis, observed variation is composed of (a) variation due to the common factors underlying the sets of items and (b) variation due to factors specific to each item. Thus for the comparison of composite means (or sum scores) to be meaningful, the individual items should be composed of similar amounts of factor variance and specific (error) variance.

This study complements the previous studies on cross-cultural similarity of the DBQ factor structures in also testing for partial measurement invariance. Partial invariance refers to relaxing some of the invariance constraints for particular items. This type of analysis gives us practical information on which specific items actually function in an equivalent manner across cultures and which may be in need of reformulation in either one version of the questionnaire or both. Further, from a practical point of view, obtaining partial strong invariance is the minimum requirement for comparing latent factor means across groups. Still, it is important to differentiate comparing latent means from comparing sum scores of observed variables; for the latter, strict invariance is required.

The models were fit to the data using the R software (R Development Core Team, 2013) packages *lavaan* (Rosseel, 2012) and semTools (Pornprasertmanit et al., 2013). The visualizations were produced using the package qgraph (Epskamp et al., 2012). The data was analyzed using the MLR estimator that is robust to the obvious non-normality of the observed variables. Modification indices and plots of residual correlations were used to improve model fit when the modifications were deemed theoretically reasonable and/or being in accordance with previous empirical results.

The following indices of approximate model fit were utilized to describe different aspects of model fit: CFI, RMSEA, SRMR and AIC. The CFI has been noted to be especially sensitive to misspecification of factor loadings (Hu and Bentler, 1999). For CFI, the cut-off value recommended by Hu and Bentler (1999) (0.95) was used as a starting point while keeping in mind that Marsh et al. (2004) caution against over-interpreting the proposed value. The RMSEA is a parsimony-adjusted index that favors simpler models over more complex ones. For RMSEA, according to Browne and Cudeck (1993) values of <0.05 may indicate good fit, while also a cut-off value of 0.06 has been proposed (Hu and Bentler, 1999). The SRMR indicates the average absolute value of the residual correlations among the observed variables, i.e., the differences between the observed and predicted correlations. For SRMR, a cut-off value of 0.08 has been proposed (Hu and Bentler, 1999). The SRMR, however, does not take into account variation in residuals: a small number of theoretically important residuals may be high even if most of the residuals are low. For this reason, we investigated individual residuals using graphical methods in addition to reporting the SRMR. AIC was reported to facilitate comparisons between models. When assessing the fit of models that were modified based on modification indices and patterns of residual correlations, avoiding overfitting models to these particular samples of data was deemed a priority; modified models were thus especially expected to show good fit to data according to these indices.

It is still an open question which analysis method is most appropriate for assessing partial measurement invariance and for identifying items that function differently across groups.

Table 2
Fit indices for the three models fit to the two samples separately.

Model	χ^2	df	p	RMSEA	90% CI	CFI	AIC	SRMR
Finnish data								
Two factors (2F)	1572.97	323	<0.001	0.061	0.058–0.063	0.766	60412	0.066
Three factors (3F)	1492.78	321	<0.001	0.059	0.056–0.062	0.781	60307	0.065
Four factors (4F)	1346.65	318	<0.001	0.055	0.053–0.058	0.808	60094	0.062
$\Delta\chi^2$(3F - 2F)	58.66	2	<0.001			0.015*		
$\Delta\chi^2$(4F - 3F)	59.04	3	<0.001			0.027*		
Irish data								
Two factors	1262.21	323	<0.001	0.060	0.057–0.063	0.788	50268	0.071
Three factors	1179.89	321	<0.001	0.057	0.054–0.060	0.806	50166	0.070
Four factors	1003.20	318	<0.001	0.051	0.048–0.055	0.845	49945	0.068
$\Delta\chi^2$(3F - 2F)	77.88	2	<0.001			0.018*		
$\Delta\chi^2$(4F - 3F)	168.29	3	<0.001			0.039*		

The $\Delta\chi^2$-values are Satorra–Bentler scaled chi square values.
* ΔCFI.

Sometimes, modification indices are used for this purpose. This procedure is referred to as the “traditional approach” by Gregorich (2006). We, however, chose to follow the build-up strategy (Pornprasertmanit et al., 2013) when performing the analyses of partial invariance and proceeded as follows: (1) all loadings/intercepts were freely estimated, (2) each loading/intercept in turn was constrained to equality across groups and an individual χ^2-test was performed separately for each loading/intercept to assess the fit of the constrained model against the unconstrained model (a Bonferroni correction was applied to account for multiple testing), (3) the indicator associated with the lowest χ^2-test value was constrained to equality across groups if doing so did not worsen model fit in a statistically significant manner and (4) the model thus obtained was used as the baseline model for testing if further indicator loadings could be constrained to equality across groups. We only report the partial invariance models that we arrived at using this procedure in order to save space; the individual model tests can be obtained from the first author upon request.

When working within the EFA framework, the question of the number of factors to extract was evaluated using multiple means. Parallel analysis (Horn, 1965) based on the mean eigenvalue criterion and 5000 iterations was performed using the paran package (Dinno, 2012). The model selection perspective to choosing the correct number of factors (Preacher et al., 2013) was also applied. Preacher et al. (2013) encourage researchers to first define the meaning of “right number of factors” before choosing the number of factors to extract. This study focused more on finding the approximately correct number of factors rather than the most replicable number of factors. In such a case, Preacher et al. (2013) recommend using either the RMSEA or the lower limit of the confidence interval of RMSEA, and the value of the Bayesian information criterion (BIC).

3. Results

Model fits for the two-, three-, and four-factor models are reported in Table 2. The two-factor model is nested within the three-factor model and the three-factor model within the four-factor model. Because of this, it was possible to compare changes in model fit via the likelihood ratio test as factors were added to the

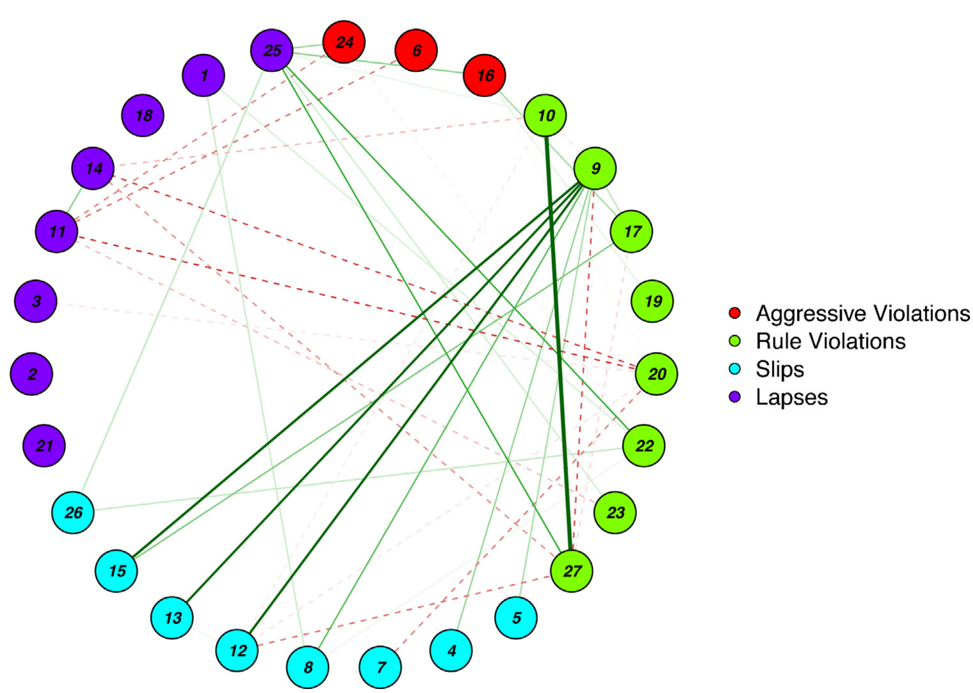

Fig. 3. Residual correlations (|r| > 0.10) among the DBQ items after fitting the four-factor model to the Finnish sample. The color and type of the lines indicates whether the correlation is positive (solid green) or negative (dashed red), while the width and the level of transparency of the line indicate the strength of the correlation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
The ten largest modification index values for the two samples.

	Finnish sample				Irish sample		
	Modification	Modification index	Expected parameter change		Modification	Modification index	Expected parameter change
1	x10 ∼∼ x27	338.04	0.622	1	Slips =∼ x9	118.87	1.064
2	Slips =∼ x9	135.60	0.723	2	Lapses =∼ x9	90.88	0.739
3	x9 ∼∼ x27	65.28	−0.113	3	x10 ∼∼ x27	77.65	0.393
4	Ruleviol =∼ x25	64.00	0.664	4	Slips =∼ x22	53.51	0.649
5	Aggr =∼ x25	47.03	0.398	5	Lapses =∼ x22	49.21	0.494
6	Lapses =∼ x20	45.74	−0.776	6	x7 ∼∼ x11	48.57	0.198
7	Lapses =∼ x9	44.13	0.336	7	x12 ∼∼ x13	46.71	0.049
8	Slips =∼ x20	43.06	−0.926	8	Lapses =∼ x27	39.61	−0.821
9	Aggr =∼ x11	41.11	−0.313	9	Slips =∼ x27	38.88	−1.023
10	Slips =∼ x27	39.70	−0.963	10	x10 ∼∼ x17	34.11	−0.160

∼∼: covariance; =∼: measured by; aggr: aggressive violations; ruleviol: rule violations.

model. These tests showed that the three-factor model fitted the data better than the two-factor model, and the four-factor model better than the three-factor model in both samples.

It is, however, not clear that the fit of the four-factor model was satisfactory. The chi-square tests rejected the assumption that the four-factor model fit well in either sample and the CFI values were extremely low in both samples. In addition to being lower than the cut-off limit of 0.95, they were much lower than what is encountered in practice for models considered as well-fitting (Jackson et al., 2009). The RMSEA value exceeded the more stringent cut-off of 0.05 in both groups while being slightly lower than the more lenient cut-off of 0.06. The SRMR values were in the acceptable range in both samples.

Because it could not be concluded that the fit of the four-factor model was satisfactory, modification indices and patterns of residual correlations were inspected to better understand the sources of the lack of model fit. Residual correlations among the 27 DBQ items are shown in Fig. 3 (Finnish sample) and 4 (Irish sample). The measurement model is represented by the differently colored nodes. Correlations of >|0.10| are shown in the two figures according to the recommendation given by Kline (2011). The ten largest modification index (MI) values are shown in Table 3 for both samples.

Overall, there were numerous strong residual correlations (RCs) among the items in spite of the SRMR values falling in the acceptable range. In both samples, item 9 ("pull out of junction") had large residual correlations with many of the items measuring *slips*. The modification index values for regressing the item on *lapses* were also high in both samples. Further, item 9 had a negative residual correlation with the speeding-related item 27 in the Finnish sample, indicating that the model over-estimated the correlation between these two items. These results are illustrated by the lines originating from item 9 in Figs. 3 and 4.

In addition, the patterns of residual correlations exhibited certain dissimilarities across samples (Fig. 5). Items 2, 14 and 9 had, in general, positive residual correlations with other items in the Irish sample, while those residuals were close to zero in the Finnish sample. The residuals between items 27 and 11 and other items were similar but different in magnitude in the two samples. The other residuals lacked a clear pattern across the samples. These results show that many of the potential modifications to improve model fit would be different in the two samples (namely, all modifications that are not related to differences in the magnitude of the residual correlations across the samples).

Overall, examination of the MIs and RCs resulted in two modifications to the four-factor model that improved model fit in

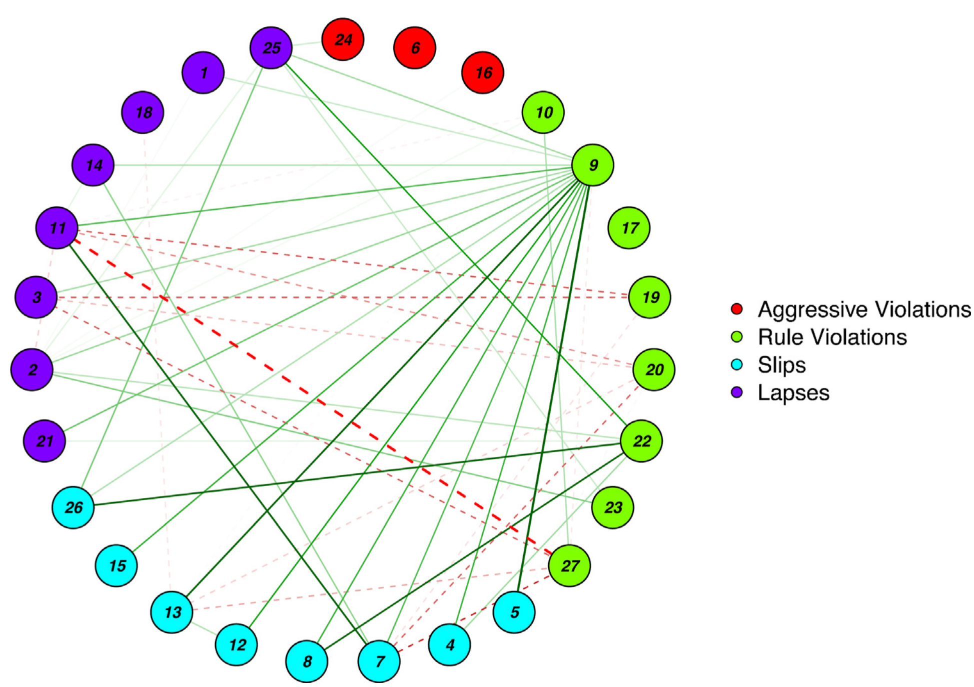

Fig. 4. Residual correlations (|*r*| > 0.10) among the DBQ items after fitting the four-factor model to the Irish sample.

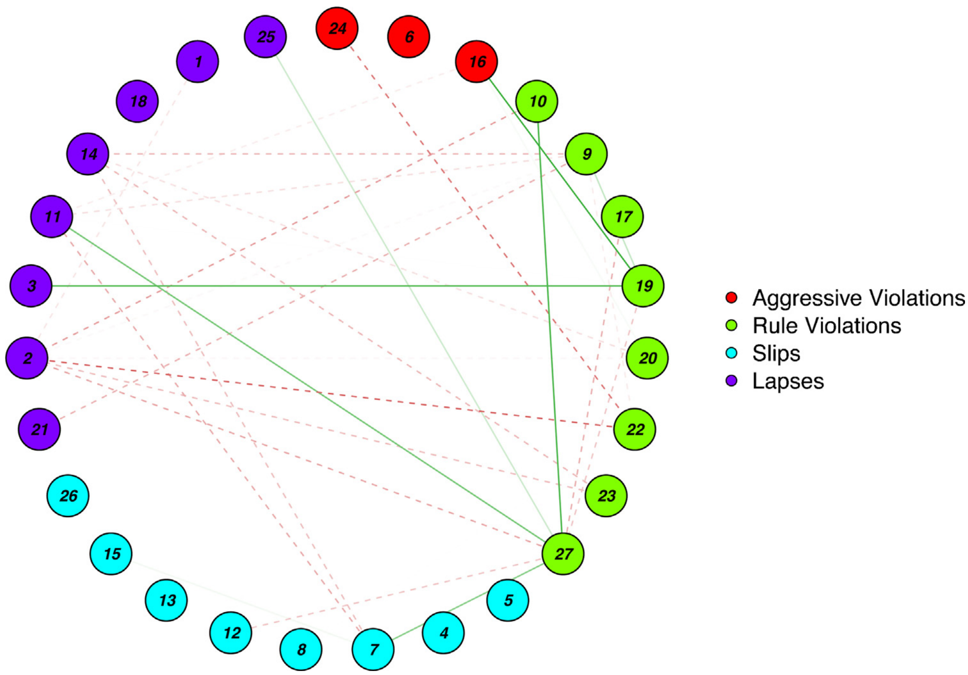

Fig. 5. Differences (Finnish residuals − Irish residuals, $|r| > 0.10$) among residual correlations between the Finnish and Irish samples.

Table 4
Fit indices for the tests of factorial invariance.

Model	χ^2	df	p	RMSEA	90% CI	CFI	SRMR
Configural inv.	1777.0	632	<0.001	0.044	0.042–0.046	0.883	0.056

both samples: (1) item 9 was re-specified to measure *slips* and (2) the error variances of speeding-related items 10 and 27 were allowed to correlate. The fit of this modified model is reported in Table 4; this model also served as the basis (configural model) of the measurement invariance analyses.

The test of configural invariance is reported in Table 4. The RMSEA indicated acceptable model fit, while the χ^2-test and the CFI indicated lack of model fit. The SRMR indicated that the average residual correlation was in the acceptable range. The low CFI value and the different patterns of residual correlations across samples (Fig. 5) were interpreted as indicating that the factor structures differed across the two samples. These considerations led us to the conservative decision to reject the configural model and to adopt the exploratory mode of analysis.

Accordingly, separate exploratory factor analyses (EFAs) were carried out in the two samples using the semTools package (Pornprasertmanit et al., 2013). The question of number of factors

Table 5
Fit indices for exploratory factor analyses with 1–5 factors.

	χ^2-test			RMSEA				BIC
	value	df	p-value	Estimate	90% C.I. LL	90% C.I. UL	p(RMSEA) ≤ 0.05	
FINLAND								
1 Factor	2291.3	324	<0.001	0.087	0.084	0.090	<0.001	61671
2 Factors	1293.3	298	<0.001	0.065	0.062	0.068	<0.001	60586
3 Factors	658.6	273	<0.001	0.044	0.041	0.048	0.996	59976
4 Factors	507.4	249	<0.001	0.038	0.034	0.041	1.000	59922
5 Factors	444.8	226	<0.001	0.035	0.031	0.039	1.000	59973
IRELAND								
1 Factor	1994.4	324	<0.001	0.091	0.088	0.095	<0.001	51559
2 Factors	940.2	298	<0.001	0.061	0.057	0.065	<0.001	50406
3 Factors	728.1	273	<0.001	0.053	0.049	0.057	0.099	50270
4 Factors	534.7	249	<0.001	0.037	0.034	0.041	0.986	50182
5 Factors	580.4	226	<0.001	0.041	0.037	0.046	0.999	50229

C.I.: confidence interval, LL: lower Limit, UL: upper limit.

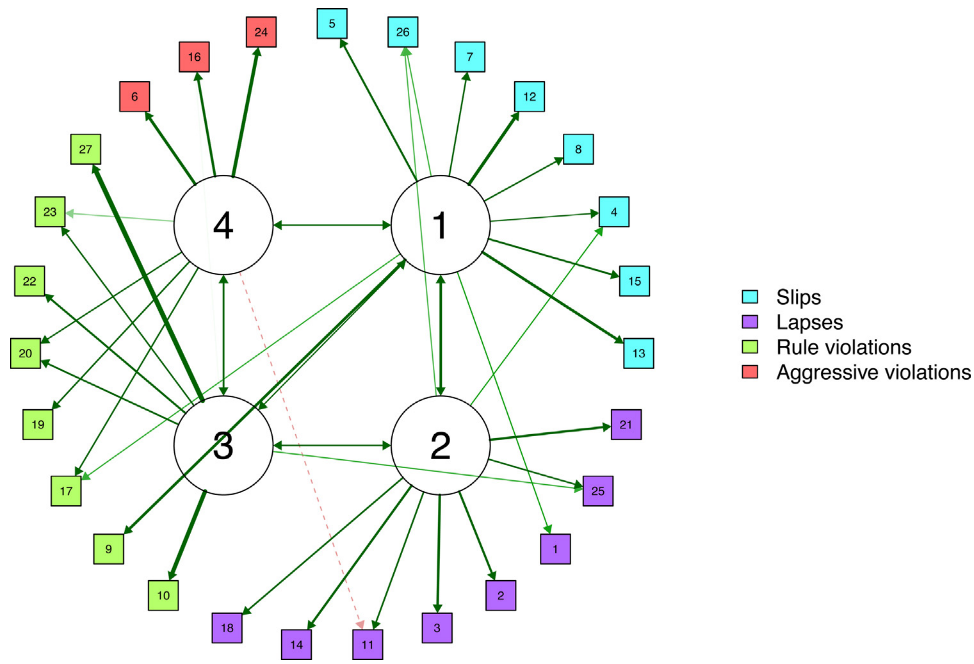

Fig. 6. The results of the exploratory factor analysis in the Finnish sample. The factor loadings implied by the original four-factor model are shown in the legend. Loadings with absolute value >0.2 are shown.

to extract was evaluated as specified in Section 2.3. Parallel analysis (Horn, 1965) suggested that four factors should be retained in both samples. The results of the model selection method of choosing the correct number of factors (Preacher et al., 2013) were in line with the parallel analysis results. In the Irish sample, the BIC and RMSEA suggested that four factors should be extracted; in the Finnish sample BIC favored extracting four factors and RMSEA three (Table 5). Based on these results together with the CFA results reported in Table 2, we decided to perform the exploratory factor analysis based on extracting four factors.

A four-factor model with oblique target rotation to the expected four-factor solution was specified in both samples. The robust maximum likelihood (MLR) estimator was used. The results of these analyses are shown in Figs. 6 and 7 and Tables 6 and 7.

Performing the exploratory factor analyses separately in the two samples helped us understand the reason for the poor fit of the configural model: the items related to various violations loaded on different factors in the two samples. In the Irish sample the factor loadings conformed to the original four-factor model (Fig. 7,

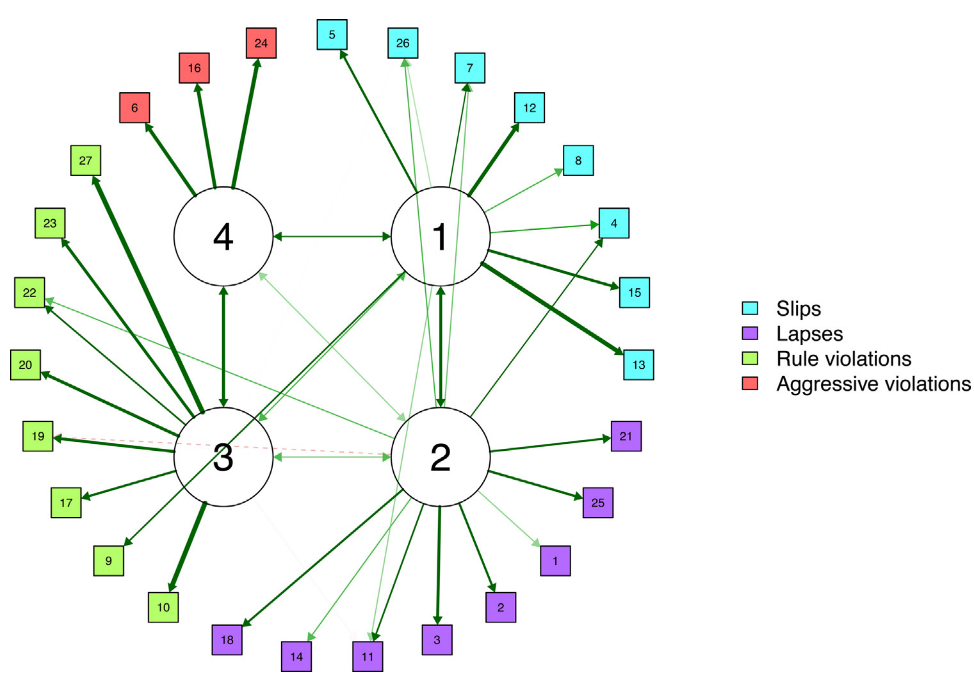

Fig. 7. The results of the exploratory factor analysis in the Irish sample. The factor loadings implied by the original four-factor model are shown in the legend. Loadings with absolute value >0.2 are shown.

Table 6
Factor loadings: EFA with target rotation for the 27 DBQ items in the Finnish sample.

Item	Factor 1 Slips	Factor 2 Lapses	Factor 3	Factor 4	λ^2
Nearly hit cyclist (12)	0.67	0.01	−0.02	0.04	0.46
Pull out of junction (9)	0.58	−0.08	0.00	0.17	0.38
Miss “Give way” signs (13)	0.58	0.01	0.10	−0.06	0.36
Fail to notice pedestrians (5)	0.45	0.12	0.06	−0.01	0.30
Attempt to overtake turning (15)	0.43	0.06	−0.01	0.13	0.27
Brake too quickly (8)	0.36	0.15	0.04	0.03	0.24
Fail to check mirror (7)	0.36	0.10	0.09	−0.13	0.18
Nearly hit car when queueing (4)	0.30	0.29	0.01	0.09	0.31
Hit something when reversing (1)	0.29	0.15	0.13	−0.07	0.19
Underestimate speed of oncoming car (26)	0.27	0.27	0.14	−0.01	0.29
Get into wrong lane (3)	0.03	0.55	−0.03	−0.01	0.31
Misread signs (21)	0.01	0.51	0.05	0.05	0.30
Take off in third gear (14)	0.05	0.50	−0.12	−0.07	0.24
“Wake up” on an unintended route (2)	−0.03	0.49	−0.05	0.08	0.21
Forget where you left car (18)	0.02	0.39	−0.01	0.10	0.27
Switch on wrong thing (11)	0.20	0.39	0.00	−0.23	0.18
No recollection of road (25)	0.00	0.37	0.28	0.04	0.31
Disregard speed limit on motorway (27)	−0.11	−0.03	0.96	−0.13	0.77
Disregard speed limit on residential road (10)	0.02	−0.07	0.85	−0.09	0.63
Drive too close (22)	0.18	0.15	0.45	0.00	0.38
Race from lights (20)	−0.07	−0.01	0.41	0.35	0.37
Cross a junction when lights red (23)	0.06	0.14	0.36	0.25	0.37
Become angered, indicate hostility (24)	−0.01	0.06	0.11	0.73	0.62
Sound horn (6)	0.01	−0.05	0.14	0.62	0.47
Become angered, give chase (16)	0.06	0.03	0.20	0.55	0.49
Overtake on the inside (19)	0.13	0.08	0.13	0.37	0.28
Force your way on a lane (17)	0.28	−0.01	0.14	0.36	0.34

Correlations among the factors

	Slips	Lapses	Factor 3	Factor 4
Slips	1.00	0.55	0.31	0.32
Lapses	0.55	1.00	0.38	0.15
Factor 3	0.31	0.38	1.00	0.42
Factor 4	0.32	0.15	0.42	1.00

The largest loading in black, other loadings in gray, crossloadings >|0.15| on gray background.

Table 7), whereas in the Finnish sample a different pattern of loadings was found (Fig. 6, Table 6). The three items related to aggressive violations loaded most strongly on the fourth factor, while four other items also had a strong loading on the factor. These items were (in descending order by the size of the loading) 19 (“overtake on the inside”), 17 (“force your way to another lane”), 20 (“race from traffic lights”), and 23 (“cross a junction after the traffic lights have turned against you”). In the Finnish sample, the speeding-related items (10 and 27) had the strongest loadings on the third factor, with items 22 (“drive close to another car”), 20 (“race from traffic lights”) and 23 (“cross a junction after the traffic lights have turned against you”) also loading on the factor.

The EFAs thus showed that the factor loadings related to *violations* were rather different in the two samples. Still, the loading patterns related to *slips* and *lapses* were similar across

Table 7
Factor loadings: EFA with target rotation for the 27 DBQ items in the Irish sample.

Item	Factor 1 Slips	Factor 2 Lapses	Factor 3 Rule violations	Factor 4 Aggressive violations	λ^2
Miss "Give way" signs (13)	0.76	−0.08	0.00	0.01	0.52
Nearly hit cyclist (12)	0.68	−0.12	0.10	−0.01	0.42
Attempt to overtake turning (15)	0.53	−0.03	0.05	0.09	0.32
Fail to notice pedestrians (5)	0.45	0.18	0.09	−0.06	0.34
Pull out of junction (9)	0.36	0.13	0.19	0.03	0.30
Fail to check mirror (7)	0.31	0.25	−0.09	0.00	0.22
Brake too quickly (8)	0.27	0.12	0.05	0.13	0.19
Get into wrong lane (3)	0.12	0.56	−0.04	−0.03	0.39
No recollection of road (25)	−0.03	0.47	0.18	0.07	0.31
"Wake up" on an unintended route (2)	−0.12	0.45	0.17	0.08	0.25
Forget where you left car (18)	0.01	0.43	−0.02	0.10	0.21
Misread signs (21)	0.19	0.43	0.06	0.01	0.33
Switch on wrong thing (11)	0.24	0.37	−0.20	0.01	0.27
Nearly hit car when queueing (4)	0.30	0.32	0.03	0.12	0.36
Underestimate speed of oncoming car (26)	0.23	0.27	0.20	−0.03	0.27
Take off in third gear (14)	0.19	0.27	0.01	−0.05	0.16
Hit something when reversing (1)	0.15	0.24	0.05	0.04	0.14
Disregard speed limit on motorway (27)	−0.05	−0.10	0.83	−0.03	0.61
Disregard speed limit on residential road (10)	−0.05	0.05	0.83	−0.14	0.56
Race from lights (20)	−0.04	−0.03	0.58	0.16	0.44
Cross a junction when lights red (23)	0.04	0.17	0.56	0.00	0.41
Overtake on the inside (19)	0.12	−0.22	0.55	0.08	0.36
Force your way on a lane (17)	0.15	−0.03	0.46	0.16	0.38
Drive too close (22)	0.06	0.27	0.37	0.11	0.37
Become angered, indicate hostility (24)	−0.08	0.10	0.13	0.73	0.66
Sound horn (6)	0.06	−0.12	0.07	0.66	0.48
Become angered, give chase (16)	0.06	0.04	−0.10	0.64	0.38

Correlations among the factors

	Slips	Lapses	Rule violations	Aggressive violations
Slips	1.00	0.55	0.27	0.31
Lapses	0.55	1.00	0.27	0.24
Rule violations	0.27	0.27	1.00	0.59
Aggressive violations	0.31	0.24	0.59	1.00

The largest loading in black, other loadings in gray, crossloadings >|0.15| on gray background.

samples. The invariance of these two factors was tested using the build-up strategy of invariance testing introduced in Section 2.3 (Model estimation and evaluation). The results of these analyses are reported in Tables 8 and 9.

The invariance analysis of *slips* indicated that the factor loadings could be treated as identical across the samples (Table 8, non-significant $\Delta\chi^2$-value of weak invariance analysis). The strong invariance assumption was, however, rejected. This prompted us to

Table 8
Fit indices for the tests of factorial invariance of slips.

Model	χ^2	df	p	RMSEA	90% CI	CFI	AIC	SRMR	$\Delta\chi^2$	Δdf	p	ΔCFI
Configural inv.	106.5	52	<0.001	0.034	0.026–0.041	0.975	29082	0.03				
Weak invariance	115.4	60	<0.001	0.031	0.024–0.039	0.974	29092	0.04	11.5	8	0.17	<0.001
Strong invariance	370.1	68	<0.001	0.069	0.063–0.075	0.859	29424	0.06	380.5	8	<0.001	0.12
Partial strong inv.	123.2	63	<0.001	0.032	0.025–0.039	0.972	29095	0.04	8.6	3	0.03*	0.002

The $\Delta\chi^2$-values are Satorra–Bentler scaled chi square values.
* Even though the p-value <0.05, the build-up strategy with Bonferroni-corrected p-values indicated that this model provided acceptable fit. See also Section 4.

Table 9
Fit indices for the tests of factorial invariance of lapses.

Model	χ^2	df	p	RMSEA	90% CI	CFI	AIC	SRMR	$\Delta\chi^2$	Δdf	p	ΔCFI
Configural inv.	149.4	40	<0.001	0.054	0.046–0.062	0.916	36226	0.04				
Weak invariance	167.5	47	<0.001	0.052	0.045–0.060	0.907	36240	0.05	18.9	7	0.009	0.009
Partial weak inv.	156.1	46	<0.001	0.051	0.043–0.058	0.915	36226	0.04	8.1	6	0.230	0.001
Strong invariance	235.2	52	<0.001	0.061	0.055–0.068	0.859	36312	0.05	93.1	6	<0.001	0.048

The $\Delta\chi^2$-values are Satorra–Bentler scaled chi square values.
The strong invariance model was compared with the partial weak invariance model.
The intercepts of the items whose loadings were freely estimated in the comparison model were also freely estimated.

investigate partial strong invariance using the build-up strategy described in Section 2.3. The items that were constrained to equality in the partial strong invariance model were items 7, 8 and 15. Items were constrained to equality in the following order: 7, 15, 8.

The invariance analysis of lapses showed that the configural model could be treated as identical across groups but the weak invariance assumption was rejected. In the partial weak invariance model that fit the data adequately, the loadings of items 2 and 18 were estimated freely and the other loadings were constrained to equality across samples. The strong invariance model with the corresponding intercepts freely estimated but all other item intercepts constrained to equality failed to fit the data as well. All the partial strong invariance models that were tested also fit significantly worse than the partial weak invariance model and so no associated results are reported in Table 9.

4. Discussion

In this study, our aim was to compare the fit of the existing two-, three-, and four-factor models to data collected from 18–25 year-old drivers in Finland and Ireland. Performing these types of analyses is important because in the DBQ tradition, it has been common practice to compare sum scores of respondents across age groups, genders or countries with little effort being put to showing that the instrument actually measures the same latent variables in each group in the same way. Indeed, it has been shown that the factor structures differ across age groups and genders for the 28-item version of the questionnaire that is standardly used (Mattsson, 2012). One way of examining the cross-cultural measurement invariance of the DBQ involves comparing samples of similar age across countries: this has the beneficial effect of ruling out the potential confounding effect of age.

In the present study, the four-factor model proved to fit the data from both countries best (research question 1). Still, two modifications needed to be made to the model: item 9 ("pull out of junction") was specified to load on *slips* and the error variances of the speeding-related items (10 and 27) were specified to correlate. The strong residual correlations between item 9 and the items measuring *slips* were interpreted as showing that at least these drivers may pull out of a junction too far out of misjudgment rather than when deliberately breaking traffic rules. The strong residual correlation between the speeding-related items (10 and 27) was thought to reflect the fact that the same drivers who speed on motorways may also plausibly speed on smaller roads.

This modified model was used as the starting point (configural model) for the invariance analyses. The configural model was, however, rejected as specified in Section 3. The conservative choice to reject the model was made for several reasons. The low value of the CFI index and the different patterns of residual correlations across samples hinted at the possibility of factor structures differing across samples. On a more theoretical note, the confirmatory analyses had actually become exploratory already when the original model was modified based on residual correlations and modification indices. It has been argued that in this situation, the discovery of misspecified loadings is more direct through rotating the factor matrix than examining modification indices (Browne, 2001, p. 113). Due to these considerations, exploratory factor analyses were performed separately for the two samples of data.

The four-factor exploratory factor analysis indicated that the Finnish sample deviated more radically from the assumed four-factor structure than the Irish one. The EFA of the Irish sample indicated that even though the largest factor loadings were as expected based on earlier research, many of the items had a secondary loading on another factor (see Table 7 and Fig. 7). Accordingly, the lack of fit of the four-factor model in the Finnish sample was due to the fact that the model was more plainly misspecified, while the lack of fit in the Irish sample was explained by the numerous cross-loadings. The practical consequence of this result for the *violations* factors is that factor means or sum scores formed on the basis of observed variables should not be compared across samples of Finnish and Irish young drivers since the very nature of the latent factors differed across the two samples. Research question 2.1. was thus answered in the negative in the case of the two *violations* factors. Further, it would be advisable to perform similar analyses before comparing the mean scores (latent or observed) across other samples of drivers.

In spite of these results, the EFAs showed that two of the four factors, *slips* and *lapses*, had similar patterns of factor loadings across samples. We decided to investigate the matter in more detail by carrying out analyses of measurement invariance separately for these two factors. The analyses showed that the factor loadings could be assumed equal for both factors, even though two loadings (those of items 2 and 18) needed to be estimated freely for the *lapses* factor to obtain adequate model fit. Research questions 2.1. and 2.2. were, then, answered in the affirmative for these factors, unlike the two *violations* factors.

We then investigated the similarity of item intercepts to obtain further information on differential item functioning across samples. Looking at *lapses*, item intercepts were clearly of different magnitudes across the samples and research question 2.3. was answered in the negative. This means that for two individuals (one from each country) with an equal standing on the latent factor, the responses to the individual items would be expected to be systematically different. Looking at *slips*, and depending on the test carried out, the item intercepts of two (7 and 15) or three (7, 15 and 8) items could be constrained to equality across samples while the intercepts of the other items needed to be estimated freely. Research question 2.3. was thus answered in the affirmative for the specified items, in the negative for most.

The practical conclusions related to these latter results are as follows. Factor loadings on *lapses* were mostly equal across samples, which suggests that the phenomenon itself was similar across the two countries. Still, the mean scores of *lapses* (latent or observed) should not be compared across samples of Finnish and Irish young drivers. The latent mean scores of *slips* could perhaps be compared with caution across samples of Irish and Finnish young drivers even though forming sum scores of the observed variables does not seem warranted. In this case, it would be advisable to treat all factor loadings and the intercepts of the items 7, 15 (and perhaps 8) as equal and estimate the other intercepts freely.

One possible interpretation of the preceding results is that the respondents from the two countries understood the items related to the more cognitively-oriented factors (*slips* and *lapses*) more similarly than the items related to the more socially-oriented aspects of traffic (different forms of *violations*).[3] This is of interest, since Finland and Ireland are similar in many respects: they are both Western, industrialized, rich, democratic countries. Still, the result seems to suggest that the social conventions in traffic differ in these two countries. Differences in implicit social norms and traffic cultures may have had an influence on which items the respondents perceived as being related to aggressive behavior on the road. If the *slips* are indeed related to the amount of attention paid to the driving task, as suggested by Reason et al. (1990), then it is at least understandable that respondents from two similar countries with similar traffic systems might commit similar attention-related errors. The same reasoning might apply to lapses as errors potentially related to absent-mindedness or lack of experience.

These results add to the on-going discussion of the correct number of factors to extract in DBQ studies (Mattsson, 2012; De Winter and Dodou, 2010; De Winter, 2013; Mattsson, 2014). In a meta-analysis of the various versions of the DBQ the two-factor solution has proved useful (De Winter and Dodou, 2010) and De Winter (2013) considers factor structures involving more than two factors as “over-extraction” of factors. Still, individual studies investigating the matter using state-of-the-art statistical methodology (Mattsson, 2012; the present contribution) seem to arrive at a different conclusion at least when basing the analysis on the 28- or 27-item version of the DBQ. The fact that a two-factor solution can be used in a meta-analysis does not mean that this would be the optimal structure in any of the individual studies included in the meta-analysis. Re-analyzing the data from these studies would be a fruitful endeavor and future studies should investigate the matter rather than taking the similarity of any one factor structure across subgroups of respondents for granted.

The ways in which the factor structure of the Finnish sample deviated from what was expected merit closer inspection. Looking at factor four, the strongest loadings were on the three aggression-related items (6, 16 and 24) while the remaining items that loaded (or strongly cross-loaded) on this factor were a subset of the items assumed to be related to *rule violations* (items 17, 19, 20 and 23). These items are related to forcing one’s way into the other lane, overtaking on the inside, racing from the traffic lights and crossing a junction after lights have changed against the driver. Perhaps the most economical explanation for the present results is that the Finnish young drivers perceive these traffic behaviors more as forms of aggressive personal interaction between two drivers than as violations of societal norms or legislation. That is, Finnish and Irish young drivers may have different conceptions of what aggressive driving consists of. Interestingly, a very much similar pattern of results was observed by Mesken et al. (2002). These authors also found that the factor of interpersonal violations included two additional items (overtaking on the inside and pushing in at the last minute) with a cross-loading on a third (racing from lights).

A further observation in the Finnish data is that the factor *rule violations* does not appear as such. Rather, factor three comprises two speeding-related items and one item related to driving close to another vehicle. Further, items 20 and 23 that are related to behavior at traffic lights cross-load on this factor. These results are markedly similar to those found in Mattsson (2012) in the group of 18–24-year-old drivers (the analyses were based on separately collected sets of data). Rather than interpreting the factor as one related to rule violations, it may be more appropriate to label it either “driving fast” or “maintaining progress”. A similar pattern of results was, again, found by Mesken et al. (2002). These authors interpreted the affective tone related to this factor as one of enjoying speed and unimpeded progress rather than as interpersonal aggression. The study by Mesken et al. (2002) was also based on a sample of Finnish drivers, so the present study replicates that finding on an independent Finnish sample.

Examining the distributions of the items in Figs. 1 and 2 may shed further light on the differences between the factor structures in the two samples. In particular, the distributions of the speeding-related items (10 and 27) appeared rather symmetric in the Finnish sample and positively skewed in the Irish sample. Could speeding actually be more common among Finnish drivers than the Irish ones? Official reports from the two countries may shed light on the issue. When comparing reports from the year 2011 (Road Safety Authority Research department, 2012; Ylönen, 2012), some differences in actual speeding across the two countries can indeed be seen. Due to different conventions in reporting the data only some of the measurements are easily comparable. On dual carriageways with a 100 km/h speed limit, 44% of the Finnish car drivers and 31% of the Irish car drivers exceeded the speed limit. On motorways with a 120 km/h speed limit, the respective figures were 24% (Finland) and 16% (Ireland). These comparisons are of course extremely tentative, but at least they are compatible with the idea that it is more common to exceed the speed limits in Finland than in Ireland. The differences in the distributions of the DBQ items might then be related to actual differences between countries.

In the Irish sample, the lack of the four-factor model fit was mainly due to numerous cross-loadings. Perhaps the most parsimonious interpretation for these findings is that there are numerous causes for any single type of traffic behavior. In particular, the inexperienced drivers that comprised the present sample may have committed the behaviors that were thought to reflect *rule violations* unintentionally because of the traffic environment “overloading” their capacity to perform the driving tasks that have not yet been automatized. Another explanation may be related to the fact that correctly estimating properties of the traffic environment (such as distances, item 22, or times, item 23) are skills that are learned and the young drivers may simply have lacked the necessary experience to perform these tasks correctly. The items cross-loading on slips and lapses may have more subtle differences in their causal origins. Perhaps the factor of *slips* is more closely related to problems of focusing attention correctly in traffic (also a skill to be learned), while the *lapses* factor may reflect inexperience with driving and the traffic environment.

Besides the actual analyses of measurement invariance and comparisons of factor structure, we introduced network modeling techniques of visualizing results to traffic psychological research. To this end, the residual correlations of the confirmatory factor analyses were presented as networks in which the items function as the nodes and the residual correlations as the edges; further, the color, width and level of transparency of the edges represented the strength of the residual correlations. Similar

[3] By choosing the terms "cognitively-oriented" and "socially-oriented" to refer to the two groups of factors we wish to emphasize that the decision to break rules in traffic requires knowledge of rules that are social in nature. We consider *slips* and *lapses* to involve cognitive processing in a more narrow sense.

mode of representation was used to communicate the results of the exploratory factor analyses. These visual representations allow us to process high-dimensional information efficiently (Epskamp et al., 2012). In the present case, it would be tedious to interpret and compare, say, the triangular matrices of residual correlations, both with 378 unique entries.

The present contribution is naturally not without limitations. For one, even though the methods used in the present contribution offer notable benefits over the similarity-based methods (van de Vijver and Leung, 1997) used in previous cross-cultural DBQ studies (Lajunen et al., 2004; Özkan et al. 2006), there is also an element of subjectivity associated with the presently utilized methods. For instance, it is not a clear-cut question of when to accept a configural model. As the value of the χ^2-test statistic is likely to be significant in a largish sample, researchers are likely to base their decision of when a model is acceptable on the values of the descriptive fit indices. This course of action was taken also in this study. The indices of approximate fit remain a hotly debated topic in the field of structural equation modeling. In this contribution we took the middle road between abandoning their use altogether (as suggested by Barrett, 2007) and using them as near-equivalents of proper test statistics: a moderate position is that the indices of approximate fit provide useful information on the ways that the model fails to fit the data (Kline, 2011).

Second, the present results do not conclusively show whether the results were due to differences between the Finnish and English language versions of the DBQ items or due to genuine cultural differences. However, it has been previously speculated that the Finnish translation of at least one of the items (item 9, pulling out of junction) could be interpreted as related to either voluntary or involuntary action (Lajunen et al., 2004; Mattsson, 2012). What the present results do show is that at least this interpretational confound was not specific to the Finnish translation, as item 9 loaded on *slips* in both samples with a cross-loading on one of the *violations* factors.

Third, different methods of data collection were used in the two countries. In particular, the representativeness of the Irish sample may be called into question. Further, it may be that systematic differences exist in how people respond to online questionnaires and traditional paper-and-pencil questionnaires. In particular, it may be that there are differences in social desirability in responses to online questionnaires and paper questionnaires. While the representativeness of the Irish sample remains an open question, the other concerns may be address based on published research findings. In the field of epidemiology, it has been shown that respondents taking an online questionnaire do not systematically differ from those surveyed by more traditional means in terms of age, gender, income, education or health status (van Gelder et al., 2010). The issue of differences in social desirability was addressed in a recent meta-analysis encompassing 16,700 participants (Dodou and de Winter, 2014). That study found no differences in social desirability scores in paper and computer surveys. Still, it would be beneficial to carry out a future study that explicitly examines the question of measurement invariance of the online and paper-and-pencil version of the DBQ.

Furthermore, it remains a possibility that the different sampling methods affected the results in some other ways that we have not taken into account. The high degree of similarity between the samples for *slips* and *lapses* suggests that the results were not simply due to different sampling methods. The two factors related to *violations* remain more open to criticism, however, as they were found to be more different in nature across samples. The differences due to genuine cultural factors and those due to sampling methods might affect the violations differently from the more cognitively-oriented factors, *slips* and *lapses*. These possibilities remain interesting targets for future investigations.

In conclusion, the present results underscore the need to take the issue of measurement invariance into account when comparing the results obtained using questionnaire instruments in traffic psychology. The factor structure of the DBQ should be further developed based on theories of cognitive ergonomics, cognitive psychology and traffic psychology. Future studies should investigate these issues also in samples of older drivers or samples of a more heterogeneous age range. The need of such studies attests, in itself, to the fact that the issue of measurement invariance can no longer be neglected in traffic psychology.

Acknowledgements

The first author would like to acknowledge funding provided by the Traffic Safety Committee of Insurance Companies (VALT) for the planning stages of the present contribution. The authors contributed to the work as follows. MM devised the idea for the publication together with FOB. MM wrote the article and performed the analyses. FOB contributed to writing in Section 2. FOB and MG collected the Irish data, TL and HS the Finnish data. HS contributed ideas to Section 4.

Appendix A.

The 27 items of the English version of the DBQ and the assumed factor loadings:

1. Hit something when reversing that you had not previously seen,
2. Intending to drive to destination A, you “wake up” to find yourself on the road to destination B,
3. Get into the wrong lane approaching a roundabout or a junction,
4. Queuing to turn left onto a main road, you pay such close attention to the main stream of traffic that you nearly hit the car in front,
5. Fail to notice that pedestrians are crossing when turning into a side street from a main road,
6. Sound your horn to indicate your annoyance to another road user,
7. Fail to check your rear-view mirror before pulling out, changing lanes, etc.,
8. Brake too quickly on a slippery road or steer the wrong way in a skid,
9. Pull out of a junction so far that the driver with right of way has to stop and let you out,
10. Disregard the speed limit on a residential road,
11. Switch on one thing, such as the headlights, when you meant to switch on something else, such as the wipers,
12. On turning left nearly hit a cyclist who has come up on your inside,
13. Miss “Give Way” signs and narrowly avoid colliding with traffic having right of way,
14. Attempt to drive away from the traffic lights in third gear,
15. Attempt to overtake someone that you had not noticed to be signalling a right turn,
16. Become angered by another driver and give chase with the intention of giving him/her a piece of your mind,
17. Stay in a motorway lane that you know will be closed ahead until the last minute before forcing your way into the other lane,
18. Forget where you left your car in a car park,
19. Overtake a slow driver on the inside,

20. Race away from traffic lights with the intention of beating the driver next to you,
21. Misread the signs and exit from a roundabout on the wrong road,
22. Drive so close to the car in front that it would be difficult to stop in an emergency,
23. Cross a junction knowing that the traffic lights have already turned against you,
24. Become angered by a certain type of a driver and indicate your hostility by whatever means you can,
25. Realise that you have no clear recollection of the road along which you have just been travelling,
26. Underestimate the speed of an oncoming vehicle when overtaking,
27. Disregard the speed limit on a motorway.

In the two-factor solution, the *violations* factor was thought to be measured by the following items: 10, 9, 17, 19, 20, 22, 23, 27, 24, 6, 16 and *errors* by the items 25, 7, 12, 5, 8, 4, 15, 26, 21, 1, 2, 3, 11, 14, 18, 13.

In the three-factor solution, the *violations* factor was assumed to be identical to that of the two-factor solution. The errors factor, however, was split in two. *Slips* were thought to be measured by the items 5, 4, 7, 8, 12, 13, 15 and 26; *lapses* by 21, 2, 3, 11, 14, 18, 1 and 25.

In the four-factor solution, the assumed factor loadings on the factors of *slips* and *lapses* were identical to those of the three-factor solution. In the four-factor solution, the *violations*-factor was split in two, items 6, 24 and 16 being detached from the previous *violations* factor to measure *aggressive violations*. Apart from these three items, the *violations* factor was identical to the other two factor solutions.

Appendix B.

Pearson correlations, means and standard deviations of the 27 DBQ variables in the two samples.

Table B1
Correlations, means and standard deviations of the 27 DBQ variables in the Finnish sample.

	dbq1	dbq2	dbq3	dbq4	dbq5	dbq6	dbq7	dbq8	dbq9	dbq10	dbq11	dbq12	dbq13	dbq14	dbq15	dbq16	dbq17	dbq18	dbq19	dbq20	dbq21	dbq22	dbq23	dbq24	dbq25	dbq26	dbq27
dbq1	1.00																										
dbq2	0.18***	1.00																									
dbq3	0.18***	0.30***	1.00																								
dbq4	0.27***	0.26***	0.28***	1.00																							
dbq5	0.20***	0.17***	0.18***	0.35***	1.00																						
dbq6	0.10**	0.07*	0.04	0.15***	0.13***	1.00																					
dbq7	0.17***	0.12***	0.17***	0.20***	0.30***	0.05	1.00																				
dbq8	0.28***	0.19***	0.18***	0.31***	0.30***	0.15***	0.21***	1.00																			
dbq9	0.18***	0.14***	0.19***	0.28***	0.27***	0.23***	0.20***	0.29***	1.00																		
dbq10	0.18***	0.12***	0.14***	0.18***	0.19***	0.27***	0.14***	0.17***	0.15***	1.00																	
dbq11	0.16***	0.16***	0.27***	0.21***	0.23***	−0.06*	0.23***	0.13***	0.16***	0.10**	1.00																
dbq12	0.27***	0.17***	0.21***	0.36***	0.36***	0.15***	0.21***	0.31***	0.38***	0.13***	0.22***	1.00															
dbq13	0.24***	0.14***	0.21***	0.25***	0.30***	0.17***	0.25***	0.25***	0.37***	0.19***	0.21***	0.45***	1.00														
dbq14	0.14***	0.17***	0.27***	0.19***	0.17***	0.00	0.15***	0.14***	0.10**	0.02	0.33***	0.15***	0.16***	1.00													
dbq15	0.21***	0.12***	0.18***	0.26***	0.23***	0.19***	0.25***	0.19***	0.36***	0.10***	0.14***	0.34***	0.26***	0.19***	1.00												
dbq16	0.12***	0.11***	0.09**	0.20***	0.22***	0.45***	0.11***	0.20***	0.28***	0.33***	0.00	0.22***	0.16***	0.03	0.20***	1.00											
dbq17	0.20***	0.13***	0.14***	0.23***	0.21***	0.32***	0.12***	0.18***	0.34***	0.26***	0.06*	0.30***	0.21***	0.06	0.31***	0.42***	1.00										
dbq18	0.16***	0.23***	0.20***	0.21***	0.13***	0.10***	0.14***	0.21***	0.16***	0.10**	0.17***	0.20***	0.15***	0.16***	0.16***	0.16***	0.18***	1.00									
dbq19	0.13***	0.11***	0.15***	0.20***	0.16***	0.31***	0.10**	0.11***	0.27***	0.25***	0.06*	0.23***	0.21***	0.08**	0.24***	0.38***	0.35***	0.16***	1.00								
dbq20	0.04	0.09**	0.05	0.17***	0.13***	0.38***	−0.01	0.11***	0.17***	0.40***	−0.03	0.11***	0.15***	−0.04	0.12***	0.41***	0.31***	0.11***	0.28***	1.00							
dbq21	0.19***	0.23***	0.35***	0.24***	0.22***	0.10**	0.12***	0.18***	0.15***	0.16***	0.25***	0.23***	0.26***	0.26***	0.22***	0.19***	0.17***	0.21***	0.21***	0.20***	1.00						
dbq22	0.25***	0.11***	0.19***	0.27***	0.24***	0.21***	0.22***	0.28***	0.22***	0.41***	0.21***	0.22***	0.29***	0.17***	0.17***	0.34***	0.29***	0.16***	0.28***	0.28***	0.24***	1.00					
dbq23	0.23***	0.18***	0.20***	0.25***	0.22***	0.30***	0.14***	0.20***	0.24***	0.40***	0.04	0.21***	0.25***	0.12***	0.19***	0.35***	0.33***	0.15***	0.31***	0.37***	0.23***	0.39***	1.00				
dbq24	0.10**	0.12***	0.08**	0.23***	0.17***	0.58***	0.08**	0.18***	0.25***	0.28***	−0.04	0.23***	0.15***	0.00	0.23***	0.54***	0.38***	0.13***	0.38***	0.38***	0.15***	0.25***	0.41***	1.00			
dbq25	0.19***	0.23***	0.20***	0.23***	0.24***	0.14***	0.16***	0.24***	0.15***	0.29***	0.16***	0.23***	0.20***	0.15***	0.13***	0.25***	0.17***	0.31***	0.18***	0.21***	0.25***	0.35***	0.30***	0.25***	1.00		
dbq26	0.17***	0.20***	0.22***	0.27***	0.33***	0.12***	0.26***	0.26***	0.23***	0.18***	0.20***	0.31***	0.28***	0.20***	0.29***	0.21***	0.19***	0.18***	0.18***	0.10**	0.28***	0.32***	0.23***	0.20***	0.34***	1.00	
dbq27	0.17***	0.07*	0.10***	0.14***	0.12***	0.25***	0.12***	0.10**	0.08**	0.69***	0.06*	0.08**	0.13***	0.00	0.11***	0.31***	0.22***	0.12***	0.22***	0.40***	0.19***	0.43***	0.40***	0.27***	0.35***	0.23***	1.00
mean	1.51	2.06	2.30	1.50	1.72	1.67	1.92	1.51	1.16	3.34	2.09	1.14	1.25	2.00	1.17	1.41	1.28	1.62	1.29	2.03	1.81	2.15	1.90	1.37	2.26	1.62	3.24
sd	0.73	0.95	0.80	0.68	0.75	1.00	0.85	0.70	0.46	1.25	0.85	0.44	0.53	0.84	0.47	0.80	0.66	0.85	0.65	1.12	0.80	0.89	0.87	0.77	1.03	0.72	1.24

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Table B2
Correlations, means and standard deviations of the 27 DBQ variables in the Irish sample.

	dbq1	dbq2	dbq3	dbq4	dbq5	dbq6	dbq7	dbq8	dbq9	dbq10	dbq11	dbq12	dbq13	dbq14	dbq15	dbq16	dbq17	dbq18	dbq19	dbq20	dbq21	dbq22	dbq23	dbq24	dbq25	dbq26	dbq27
dbq1	1.00																										
dbq2	0.27***	1.00																									
dbq3	0.24***	0.27***	1.00																								
dbq4	0.25***	0.29***	0.35***	1.00																							
dbq5	0.21***	0.19***	0.24***	0.39***	1.00																						
dbq6	0.08*	0.14***	0.06	0.17***	0.10**	1.00																					
dbq7	0.13***	0.06	0.24***	0.18***	0.31***	0.04	1.00																				
dbq8	0.16***	0.12***	0.15***	0.26***	0.29***	0.19***	0.24***	1.00																			
dbq9	0.18***	0.21***	0.24***	0.29***	0.34***	0.18***	0.24***	0.26***	1.00																		
dbq10	0.15***	0.23***	0.09**	0.17***	0.16***	0.25***	0.07	0.14***	0.25***	1.00																	
dbq11	0.12***	0.08*	0.35***	0.25***	0.21***	0.01	0.35***	0.09**	0.24***	0.02	1.00																
dbq12	0.19***	0.11**	0.22***	0.31***	0.30***	0.18***	0.20***	0.25***	0.30***	0.15***	0.22***	1.00															
dbq13	0.24***	0.13***	0.27***	0.36***	0.37***	0.14***	0.29***	0.25***	0.33***	0.11**	0.25***	0.49***	1.00														
dbq14	0.10**	0.11**	0.25***	0.16***	0.25***	0.04	0.27***	0.07*	0.20***	0.08*	0.28***	0.16***	0.20***	1.00													
dbq15	0.15***	0.11**	0.22***	0.28***	0.32***	0.22***	0.17***	0.25***	0.28***	0.12***	0.15***	0.35***	0.39***	0.23***	1.00												
dbq16	0.16***	0.19***	0.09*	0.23***	0.11**	0.39***	0.09*	0.19***	0.16***	0.18***	0.11**	0.15***	0.19***	0.07*	0.22***	1.00											
dbq17	0.16***	0.21***	0.13***	0.22***	0.18***	0.32***	0.08*	0.16***	0.31***	0.34***	0.03	0.21***	0.22***	0.07	0.23***	0.28***	1.00										
dbq18	0.13***	0.21***	0.25***	0.24***	0.18***	0.14***	0.18***	0.25***	0.15***	0.06	0.23***	0.16***	0.12***	0.20***	0.16***	0.10**	0.16***	1.00									
dbq19	0.02	0.09*	−0.02	0.12***	0.11**	0.34***	0.00	0.10**	0.17***	0.38***	−0.06	0.17***	0.12***	0.06	0.16***	0.15***	0.40***	0.08*	1.00								
dbq20	0.08*	0.18***	0.04	0.20***	0.08*	0.34***	−0.03	0.13***	0.21***	0.47***	−0.02	0.14***	0.08*	0.05	0.17***	0.25***	0.41***	0.09*	0.44***	1.00							
dbq21	0.19***	0.26***	0.43***	0.34***	0.27***	0.12***	0.19***	0.18***	0.26***	0.15***	0.27***	0.27***	0.30***	0.22***	0.28***	0.12***	0.20***	0.32***	0.12***	0.15***	1.00						
dbq22	0.19***	0.23***	0.23***	0.30***	0.24***	0.25***	0.21***	0.33***	0.32***	0.36***	0.08*	0.19***	0.20***	0.14***	0.24***	0.27***	0.33***	0.15***	0.22***	0.34***	0.26***	1.00					
dbq23	0.18***	0.27***	0.20***	0.23***	0.19***	0.26***	0.14***	0.18***	0.28***	0.48***	0.06	0.18***	0.20***	0.20***	0.20***	0.19***	0.42***	0.14***	0.36***	0.39***	0.20***	0.44***	1.00				
dbq24	0.15***	0.23***	0.12***	0.26***	0.15***	0.55***	0.10**	0.21***	0.26***	0.35***	0.04	0.17***	0.16***	0.08*	0.17***	0.49***	0.39***	0.16***	0.33***	0.42***	0.20***	0.38***	0.37***	1.00			
dbq25	0.16***	0.33***	0.26***	0.27***	0.27***	0.15***	0.19***	0.22***	0.22***	0.22***	0.16***	0.17***	0.19***	0.17***	0.18***	0.16***	0.19***	0.32***	0.13***	0.23***	0.27***	0.32***	0.26***	0.28***	1.00		
dbq26	0.16***	0.17***	0.27***	0.31***	0.29***	0.09*	0.20***	0.25***	0.23***	0.19***	0.17***	0.30***	0.26***	0.10**	0.28***	0.17***	0.23***	0.24***	0.13***	0.16***	0.28***	0.35***	0.23***	0.20***	0.36***	1.00	
dbq27	0.09*	0.19***	0.01	0.11**	0.10**	0.33***	−0.05	0.14***	0.17***	0.61***	−0.12***	0.16***	0.08*	0.03	0.12***	0.19***	0.45***	0.08*	0.42***	0.47***	0.12***	0.36***	0.39***	0.40***	0.21***	0.22***	1.00
mean	1.38	1.44	2.25	1.47	1.51	1.92	1.90	1.47	1.59	2.47	2.03	1.24	1.21	1.64	1.22	1.23	1.58	2.14	1.68	1.99	1.76	1.64	1.92	1.68	1.94	1.56	2.36
sd	0.60	0.76	0.87	0.65	0.69	1.04	0.92	0.70	0.78	1.30	1.03	0.57	0.48	0.83	0.51	0.61	0.91	1.11	1.05	1.27	0.85	0.77	0.96	0.93	1.05	0.75	1.54

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

References

Åberg, L., Rimmö, P., 1998. Dimensions of aberrant driver behaviour. Ergonomics 41 (1), 39–56. doi:http://dx.doi.org/10.1080/001401398187314.

Barrett, P., 2007. Structural equation modelling: adjudging model fit. Pers. Indiv. Differ. 42 (5), 815–824.

Browne, M.W., 2001. An overview of analytic rotation in exploratory factor analysis. Multivar. Behav. Res. 36 (1), 111–150.

Browne, M., Cudeck, R., 1993. Alternative ways of assessing model fit. In: Bollen, K.A., Long, J.S. (Eds.), Testing Structural Equation Models. SAGE, Newbury Park, CA, pp. 136–162.

Davey, J., Wishart, D., Freeman, J., Watson, B., 2007. An application of the driver behaviour questionnaire in an Australian organisational fleet setting. Transport. Res. Part F: Traffic Psychol. Behav. 10 (1), 11–21.

De Winter, J., Dodou, D., 2010. The driver behaviour questionnaire as a predictor of accidents: a meta-analysis. J. Saf. Res. 41 (6), 463–470.

De Winter, J., 2013. Small sample sizes, overextraction, and unrealistic expectations: a commentary on M. Mattsson. Accid. Anal. Prev. 50, 776–777. doi:http://dx.doi.org/10.1016/j.aap.2012.07.003.

Dinno, A., paran: Horn's Test of Principal Components/Factors [R package version 1.5.1] Retrieved from 2012; http://CRAN.R-project.org/package=paran.

Dodou, D., de Winter, J., 2014. Social desirability is the same in offline, online, and paper surveys: a meta-analysis. Comput. Hum. Behav. 36, 487–495.

Epskamp, S., Cramer, A.O., Waldorp, L.J., Schmittmann, V.D., Borsboom, D., 2012. Qgraph: network visualizations of relationships in psychometric data. J. Stat. Software 48 (4), 1–18.

Gregorich, S.E., 2006. Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. Med. Care 44, S78–94. doi:http://dx.doi.org/10.1097/01.mlr.0000245454.12228.8f (11 Suppl. 3).

Hennessy, D.A., Wiesenthal, D.L., 2005. Driving vengeance and willful violations: clustering of problem driving attitudes. J. Appl. Soc. Psychol. 35 (1), 61–79. doi:http://dx.doi.org/10.1111/j.1559-1816.2005.tb02093.x.

Horn, J., 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30, 179–185.

Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model.: A Multidiscip. J. 6 (1), 1–55. doi:http://dx.doi.org/10.1080/10705519909540118.

IBM Corp., 2012. IBM SPSS Statistics for Windows, Version 21.0.

Jackson, D.L., Gillaspy Jr, J.A., Purc-Stephenson, R., 2009. Reporting practices in confirmatory factor analysis: an overview and some recommendations. Psychol. Methods 14 (1), 6.

Kline, R.B., 2011. Principles and Practice of Structural Equation Modeling, third ed. The Guilford Press, New York

Kontogiannis, T., Kossiavelou, Z., Marmaras, N., 2002. Self-reports of aberrant behaviour on the roads: errors and violations in a sample of Greek drivers. Accid. Anal. Prev. 34 (3), 381–399. http://dx.doi.org/10.1016/S0001-4575(01)00035-5.

Lajunen, T., Parker, D., Summala, H., 2004. The Manchester Driver Behaviour Questionnaire: a cross-cultural study. Accid. Anal. Prev. 36 (2), 231–238.

Lawton, R., Parker, D., Manstead, A.S.R., Stradling, S.G., 1997. The role of affect in predicting social behaviors: the case of road traffic violations. J. Appl. Soc. Psychol. 27 (14), 1258–1276. doi:http://dx.doi.org/10.1111/j.1559-1816.1997.tb01805.x.

Little, T.D., Slegers, D.W., Card, N.A., 2006. A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. Struct. Equ. Model. 1, 59–72.

Marsh, H.W., Hau, K., Wen, Z., 2004. In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. Struct. Equ. Model.: A Multidiscip. J. 3, 320–341.

Martinussen, L.M., Lajunen, T., Møller, M., Özkan, T., 2013. Short and user-friendly: the development and validation of the Mini-DBQ. Accid. Anal. Prev. 50, 1259–1265.

Mattsson, M., 2014. On testing factorial invariance: a reply to J.C.F. de Winter. Accid. Anal. Prev. 63 (0), 89–93. http://dx.doi.org/10.1016/j.aap.2013.10.031.

Mattsson, M., 2012. Investigating the factorial invariance of the 28-item DBQ across genders and age groups: an exploratory structural equation modeling study. Accid. Anal. Prev. 48 (0), 379–396. doi:http://dx.doi.org/10.1016/j.aap.2012.02.009.

Mesken, J., Lajunen, T., Summala, H., 2002. Interpersonal violations, speeding violations and their relation to accident involvement in Finland. Ergonomics 7, 469–483.

Özkan, T., Lajunen, T., Chliaoutakis, J.E., Parker, D., Summala, H., 2006. Cross-cultural differences in driving behaviours: a comparison of six countries. Transport. Res. Part F: Traffic Psychol. Behav. 9 (3), 227–242. doi:http://dx.doi.org/10.1016/j.trf.2006.01.002.

Parker, D., Reason, J.T., Manstead, A.S.R., Stradling, S.G., 1995. Driving errors, driving violations and accident involvement. Ergonomics 38 (5), 1036–1048.

Pornprasertmanit, S., Miller, P., Schoemann, A., Rosseel, Y., 2013. Sem tools: useful tools for structural equation modeling. R Package Available on CRAN.

Preacher, K., Zhang, G., Kim, C., Mels, G., 2013. Choosing an optimal number of factors in exploratory factor analysis: a model selection perspective. Multivar. Behav. Res. 48, 28–56. doi:http://dx.doi.org/10.1080/00273171.2012.710386.

R. Development Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria

Reason, J., 1990. Human Error. Cambridge University Press, Cambridge

Reason, J., Manstead, A., Stradling, S., Baxter, J., Campbell, K., 1990. Errors and violations on the roads: a real distinction? Ergonomics 33 (10), 1315–1332.

Road Safety Authority Research department, 2012. Free Speed Survey 2011 (Urban and Rural). Road Safety Authority

Rosseel, Y., 2012. Lavaan: an R package for structural equation modeling. J. Stat. Software 2, 1–36.

van de Vijver, F., Leung, K., 1997. Methods and data analysis of comparative research. In: Berry, J.W., Poortinga, Y.H., Pandey, J. (Eds.), Handbook of Cross-Cultural Psychology: Theory and Method, Vol. 1. Allyn & Bacon, Boston, pp. 257–300.

van Gelder, M.M.H.J., Bretveld, R.W., Roeleveld, N., 2010. Web-based questionnaires: the future in epidemiology? Am. J. Epidemiol. 11, 1292–1298. doi:http://dx.doi.org/10.1093/aje/kwq291.

Ylönen, J., 2012. Autojen nopeudet vuonna 2011. Liikennevirasto, Helsinki