Accident Analysis and Prevention 48 (2012) 379–396

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

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

Investigating the factorial invariance of the 28-item DBQ across genders and age groups: An Exploratory Structural Equation Modeling Study

Markus Mattsson *

Traffic Research Unit, Institute of Behavioural Sciences, University of Helsinki, PO Box 9 (Siltavuorenpenger 1 A), FI-00014 University of Helsinki, Finland

ARTICLE INFO

Article history:
Received 28 July 2011
Received in revised form
26 November 2011
Accepted 12 February 2012

Keywords:
Driver Behavior Questionnaire
DBQ
Factorial invariance
Exploratory Structural Equation Modeling
ESEM

ABSTRACT

The Driver Behaviour Questionnaire (DBQ) is perhaps the most widely used questionnaire instrument in traffic psychology with 174 studies published by late 2010. The instrument was developed based on a plausible cognitive ergonomic theory (the Generic Error Modeling System, GEMS), but the factor structure obtained in the original study (Reason et al., 1990) did not mirror the theory's conceptual structure. This led to abandoning GEMS and adopting the obtained factor structure as a starting point for further DBQ research. This article argues that (1) certain choices in the original study, concerning statistical methodology and the wording of individual question items, may have contributed to the ways the obtained factor structure deviated from the underlying theory and (2) the analysis methods often used in DBQ studies, principal components (PC) analysis and maximum likelihood (ML) factor analysis, are not optimal choices for the non-normally distributed categorical data that is obtained using the instrument. This is because ML produces biased results when used with this type of data, while PC is by definition unable to uncover latent factors as it summarizes all variation in the measured variables. (3) Even though DBQ factor scores have been routinely compared in subgroups of men and women and respondents of different ages, DBQ's factorial invariance in these groups has not been rigorously tested. These concerns are addressed in this article by framing the results of certain previous DBQ studies as a structural equation model (SEM) and an Exploratory Structural Equation Model (ESEM) and testing measurement model fit in subgroups of respondents. The SEM analyses indicate that the model does not fit data from the whole sample of respondents as it stands, while the ESEM analyses show that a modification of the model does. However, the ESEM analyses indicate the DBQ measures different underlying latent variables in the different subgroups. Based on the analyses and a review of recent advances in attention and memory research, an update to the theory underlying the DBQ is suggested.

1. Introduction

Traffic psychology is a field that provides a fertile testing ground for the ecological validity and generalizability of psychological theories. Theories of human error are one much researched example of this – for instance Reason et al. (1990) state that: "The road environment makes an excellent natural laboratory for observing aberrant behaviors". One influential view of human error, presented in Reason (1990) differentiates the possible kinds of human error based on whether the error was due to the action sequence not proceeding as planned (*slips* and *lapses*) or to the chosen action not being appropriate for the context in which it was executed (*mistakes*). *Slips* and *lapses* are skill-based errors, related to the execution of a motor plan, the former being especially related to attention and/or execution of movements, the latter to retrieval from memory. *Mistakes* are caused by an unsuccessful choice of means to attain an end – or, in plain English, by bad planning. In addition to these types of error that Reason (1990) deems definable in relation to the cognitive processes of the individual, the category of *violations* (of social rules and/or norms) are defined as "deliberate . . . violations from those practices deemed necessary . . . to maintain the safe operation of a potentially hazardous system". The ideas presented in Reason (1988) and Reason (1990) were adapted to the traffic context by Reason et al. (1990) who developed the Driver Behavior Questionnaire (DBQ) to measure these different types of human error that may occur on the roads. The original 50-item version of the questionnaire was based on the conceptual framework of the Generic Error Modeling System developed in Reason (1990).

These categorizations of errors are not rigid: *mistakes* can be further divided based on whether they are rule- or knowledge-based while *slips* and *lapses* can be combined into one category as, for instance, Reason et al. (1990) do, calling the resulting category *silly errors*. A further, higher-level categorization combines *slips*, *lapses* and *mistakes* into a single category of *unintentional errors*,

* Tel.: +358 40 7689406; fax: +358 91 9129422.
E-mail address: markus.mattsson@helsinki.fi

0001-4575/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2012.02.009

differentiating them from *violations*. In fact, a meta-analysis of all the DBQ studies carried out so far found that the only truly stable distinction is that between *unintentional errors* and *intentional violations* (de Winter and Dodou, 2010); this distinction seems to hold true irrespective of the age and gender of the respondent, the country (or traffic culture) investigated or the type of vehicle used. Even though the *errors/violations* – distinction is of use when predicting accidents, it is of more interest to the planners of traffic safety interventions than to the traffic psychologist interested in which cognitive process malfunctions when a given type of error occurs. Grounding the proposed error categories in malfunctions of basic human cognitive processes is clearly a strong point of the GEMS in this respect.

After it was found that violations predict self-reported accident-involvement (Parker et al., 1995), Lawton et al. (1997) added to the instrument more items concerning interpersonal *aggressive violations* on the one hand and ordinary *highway code violations* on the other. In addition to the *violations* factor, also the *lapses* factor has since the original study been divided into different kinds of subfactors. Åberg and Rimmö (1998) postulated the categories of *inattention* and *inexperience errors* after collecting data using a 104-item version of the DBQ. When used in an applied setting, items measuring factors not related to accidents in previous studies have been omitted altogether; for instance Freeman et al. (2009) used a 20-item version of the DBQ including items related to *errors*, *aggressive violations* and *highway code violations*. Similarly, Özkan et al. (2006a) used a 19-item version of the DBQ measuring only the above-mentioned dimensions. Dropping the items loading on the *lapses*-factor is not, however, a feasible choice for a traffic psychologist carrying out basic research as they relate to malfunctions in (working) memory processes, which are clearly of interest for a researcher aiming to understand the cognitive processes underlying traffic behavior.

There are probably many reasons for why the factor structures found in DBQ studies have failed to mirror the theoretical constructs postulated by Reason (1990). One way of categorizing the reasons is to divide them into substantive and methodological. The substantive reasons include:

- Unrealistic expectations of what the respondents are capable of reporting, such as questions of the form “please try to remember the last time you have forgotten X” (Lajunen and Özkan, 2011).
- Differences in traffic cultures between countries – the factorial structure found in Reason et al. (1990) may not be universal even in principle as evidenced by the results of Rimmö and Hakamies-Blomqvist (2002), Eugenia Gras et al. (2006) and Xie and Parker (2002).

The methodological reasons include:

- The use of principal components analysis and thus summarizing all variation in the data even if the analysis’ objective is finding latent constructs that would explain the variation in observed variables. In this situation, factor analysis should be performed. Studies that employ this methodology include, e.g. Reason et al. (1990), Parker et al. (1995), Eugenia Gras et al. (2006) and Xie and Parker (2002).
- The use of statistical methods not suitable for non-normally distributed non-continuous data, such as ML estimation with no correction (Özkan et al., 2006a). Principal axis factoring (Lajunen et al., 2004; Özkan et al., 2006b) is not an optimal method, either, being based on product-moment correlations rather than polychoric correlations (Holgado-Tello et al., 2010).
- Possible systematic differences in factor structures between respondent groups based on age (Parker et al., 2000) and sex (Rimmö and Hakamies-Blomqvist, 2002).
- Comparing values of factor means in different groups of respondents without testing the assumption that the same factor structure can be obtained in each of the groups – that is, without testing the instrument’s *factorial invariance* and *measurement invariance*. This is the main point of departure for the empirical part of the present study: does the DBQ measure the same latent constructs in different subsamples of respondents?

Concerning the substantive points, it is an interesting and open empirical question whether the DBQ could be developed in a theory-driven manner, as a tool for measuring errors related to the basic cognitive processes that influence situation specific driving choices, such as the trajectory or speed chosen for a certain type of bend on a certain type of road and gap acceptance in overtaking. As the different versions of the DBQ include lots of applied, safety-oriented items at the expense of cognitively motivated items, in the future it may be of interest to include items that serve cognitive analysis of driver behavior (and safety) better. This way of proceeding receives support from a recent study by af Wåhlberg et al. (2011) in which factors complying with the *slips*, *lapses*, *mistakes*, *violations* – classification were not found and in which the DBQ items could not be used to predict drivers’ actual crashes, but only those that the drivers reported themselves.

Concerning the methodological points, certain further considerations are in order. When psychometric instruments, such as the DBQ, are used to compare groups of respondents, it is necessary to first ensure that the instrument measures the same latent constructs in the same way in the different groups. The process of doing so is called testing the *factorial invariance* of the instrument; for an overview see, e.g. Dimitrov (2010). One way of testing the assumption of factorial invariance is to fit structural equation models (SEMs) to the data in a series of stages. In the first stage, called the test of *configural invariance*, a model based on a theory (or the results of prior research) is fit first to the whole sample and then, after establishing satisfactory model fit, separately to the subsamples. While the same pattern of fixed and free model parameters (i.e., in practice, factor loadings) is specified in the subsamples to be compared, parameter values are not constrained to be equal. In practice passing this test means that similar but not identical latent variables are being measured in the different subsamples. In the second stage, factor loadings are constrained to be equal across groups – this stage is known as the test of *weak* (or *metric*) *measurement invariance*. If this test is passed, it can be concluded that the same latent variables are being measured in the different subsamples. In the third stage, factor loadings and indicator means (in the case of continuous variables) or indicator thresholds (in the case of categorical variables) are constrained to equality – this stage is called the test of *strong* (or *scalar*) *measurement invariance*. Muthén and Muthén (2010, pp. 433–435) recommend evaluating weak and strong measurement invariance simultaneously when operating with categorical data since factor loadings and item thresholds both determine the shape of the item characteristic curve. In the fourth stage the invariance of factor loadings, indicator means (or thresholds) and item uniquenesses are constrained to be equal (*strict measurement invariance*). This methodology is described for example in Marsh et al. (2009), Vandenberg and Lance (2000) and Dimitrov (2010). From the point of view of DBQ studies it is of interest to note that comparing factor means across groups requires that the test of strong measurement invariance be passed.

One of the challenges for the researcher opting to test the invariance assumptions using structural equation models is, however, that real-life data in the human sciences is seldom as clearly divisible into discrete categories as using the method demands. In a