Transportation Research Part F 44 (2017) 105–119

Contents lists available at ScienceDirect

Transportation Research Part F

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

Psychometric costs of retaking driving-related cognitive ability tests

Markus Sommer [a], Martin E. Arendasy [a,*], Bettina Schützhofer [a,b]

[a] University of Graz, Austria
[b] sicher unterwegs – Verkehrspsychologische Untersuchungen GmbH, Austria

ARTICLE INFO

Article history:
Received 21 January 2016
Received in revised form 6 October 2016
Accepted 6 October 2016
Available online 26 November 2016

Keywords:
Traffic psychological assessment
Practice effects
Retesting
Measurement fairness

ABSTRACT

The present study tested predictions deduced from competing models advanced in the literature to explain practice effects in driving-related cognitive ability tests due to retesting. The models differ in terms of the processes assumed to be responsible for practice effects and make competing predictions with regard to the level measurement invariance across test administration sessions. A total of $N = 239$ test-takers solved four driving-related cognitive ability tests at three time-points of measurement. Item response theory analyses indicated that practice effects can be explained in terms of an increase in test-specific abilities. The size of the practice effect varied across driving-related cognitive ability tests and across test-takers' level of general mental ability. Latent mean and covariance structure analyses indicated that the observed improvements in test performance are purely test-specific and do not generalize to broad cognitive speededness.

1. Introduction

In several European countries traffic-psychological assessment is mandatory to regain a driving license after a withdrawal (cf. Schuhfried, 2010). In addition, driving-related cognitive ability tests are used to select professional drivers. While there are strict regulations on retesting in the former field of application, retesting is common in personnel selection settings (Lievens, Buyse, & Sackett, 2005). Unfortunately, there is evidence that allowing test-takers to retake cognitive ability tests improves their test scores (cf. Calamia, Markon, & Tranel, 2012; Hausknecht, Halpert, Di Paolo, & Moriarty Gerrard, 2007; Kulik, Kulik, & Nangert, 1984). The increase in test scores due to retesting is called practice effect (cf. Calamia et al., 2012; Lievens et al., 2005). Depending on the size of the practice effect and the reasons for the improvement in test scores diagnosticians may run the risk of drawing invalid conclusions based on test-takers retest scores if practice effects are left unaccounted for (cf. Heilbronner et al., 2010). The present study was conducted to determine the size of practice effects in commonly used traffic psychological tests and to examine causes for the practice effect to evaluate whether retesting constitutes a problem in traffic psychological assessment or not.

1.1. Locating driving-related cognitive tests within models of human intelligence

Traffic psychological test batteries usually contain measures of (1) complex choice reaction time, (2) simple choice reaction time, (3) selective attention, and (4) perceptual speed. Factor-analytic studies (e.g. Ackerman & Cianciolo, 2000; Carroll,

* Corresponding author at: Psychological Methods and Assessment Group, Department of Psychology, University of Graz, Universitätsplatz 2, 8010 Graz, Austria.

E-mail address: martin.arendasy@uni-graz.at (M.E. Arendasy).

http://dx.doi.org/10.1016/j.trf.2016.10.014
1369-8478/

1993; Danthiir, Wilhelm, & Roberts, 2012; Danthiir, Wilhelm, Schulze, & Roberts, 2005; Häusler & Sturm, 2009; Johnson & Deary, 2011; Karner & Neuwirth, 2000; Krumm, Schmidt-Atzert, Schmidt, Zenses, & Stenzel, 2012; Roberts & Stankov, 1999; Rożnowski, Gorbaniuk, Biela-Warenica, & Biela, 2015) indicated that these tests measure different facets of broad cognitive speededness (G_s). This line of research also indicated that complex choice reaction time, selective attention and perceptual speed exhibit higher factor loading on broad cognitive speededness (=highest G_s-saturation), than measures of simple choice reaction. Furthermore, the four facets of broad cognitive speededness (G_s) measured with traffic psychological test batteries have been shown to exhibit differential predictive utility in predicting safe driving behavior and accident proneness (e.g. Aniţei, Chraif, Schuhfried, & Sommer, 2011; Bukasa, Christ, Ponocny-Seliger, Smuc, & Wenninger, 2003; Ferreira, Simões, & Maroco, 2013; Karner & Neuwirth, 2000; Risser et al., 2008; Sommer & Häusler, 2005; Sommer et al., 2008). In general, measures exhibiting a higher G_s-saturation also exhibited a higher prognostic utility.

1.2. Moderators of the effect size of practice effects

Meta-analytic studies (cf. Calamia et al., 2012; Hausknecht et al., 2007; Kulik et al., 1984) indicated that the effect size of the practice effect depends on (1) characteristics of the psychometric test used, (2) the time interval between test administration sessions, and (3) test-takers' intelligence.

1.2.1. Construct- and test characteristics as a moderator of practice effects

Several studies indicated that the effect size of the practice effect tends to be small to medium for measures of fluid intelligence, complex choice reaction time, and perceptual speed tests and small to negligible for simple choice reaction tests and verbal ability (e.g. Arendasy & Sommer, 2013a; Bartels, Wegrzyn, Wiedl, Ackermann, & Ehrenreich, 2010; Beglinger et al., 2005; Falleti, Maruff, Collie, & Darby, 2006; Lowe & Rabbitt, 1998; Rabbitt, Diggle, Holland, & McInnes, 2004; Rabbitt, Lunn, Ibrahim, & McInnes, 2009; Rabbitt, Lunn, Wong, & Cobain, 2008; Rapport, Brines, Theisen, & Axelrod, 1997; Salthouse & Tucker-Drob, 2008; Wilson, Watson, Baddley, Emslie, & Evans, 2000). However, these findings have not always been consistently replicated (cf. te Nijenhuis, Vianen, & van der Flier, 2007). Furthermore, there is evidence that characteristics of the specific test paradigms may be more important for explaining differences in the effect size of the practice effect than the cognitive ability domain measured (cf. Calamia et al., 2012; Hausknecht et al., 2007).

1.2.2. Time interval as a moderator of practice effects

There is evidence that the size of practice effects is negatively related to the time interval between subsequent test administrations (cf. Calamia et al., 2012; Hausknecht et al., 2007). This effect has also been shown to be moderated by characteristics of the specific test used. For instance, practice effects remained detectable for up to seven years for some cognitive ability tests, while shorter time intervals turned out to be sufficient to eliminate practice effects observed for others (e.g. Calamia et al., 2012; Falleti et al., 2006; Rabbitt et al., 2004, 2009; Salthouse, 2010; Salthouse & Tucker-Drob, 2008; Wilson et al., 2000). There were also differences in the time course of practice effects between cognitive ability tests (e.g. Bartels et al., 2010; Beglinger et al., 2005; Benedict & Zgaljardic, 1998; Calamia et al., 2012; Falleti et al., 2006; Wilson et al., 2000). While practice effects continued to increase over subsequent retest sessions for some psychometric tests, they reached a plateau after the first or second retest for others.

1.2.3. General mental ability as moderators of practice effects

Meta-analytic research (Kulik et al., 1984) indicated that the effect size of the practice effect was larger for more intelligent (=high-g) test-takers (meta-analytic Cohen's $d = .80$) than for less intelligent (=low-g) test-takers (meta-analytic Cohen's $d = .40$). This effect has been consistently replicated in several independent studies (cf. Arendasy & Sommer, 2013a; Freund & Holling, 2011; Kulik et al., 1984; Rabbitt et al., 2008; Rapport et al., 1997; Salthouse & Tucker-Drob, 2008). This implies that the mechanisms responsible for practice effects may be sensitive to individual differences in intelligence.

1.3. Explaining practice effects in driving-related cognitive ability tests

Five competing theoretical models have been advanced in the literature to account for practice effects that differ in the processes assumed to be responsible for the practice effect and in their implications for the use of retesting in traffic psychological assessment (cf. Arendasy & Sommer, 2013a; Lievens, Reeve, & Heggestad, 2007; Lievens et al., 2005).

1.3.1. Increase in test scores due to an increase in test-familiarity

The test *familiarity hypothesis* (Anastasi, 1981) attributes practice effects to an increase in test familiarity. According to this model less test familiar test-takers are faced with a dual task at the initial test administration session. They need to familiarize themselves with the item format while simultaneously solving the first few test items. By contrast, more test-familiar test-takers do not need to invest this additional effort. This implies that the first few items of a test should be harder to solve for test-takers unfamiliar with the item format than for equally able test-takers, who are already familiar with the item format. This effect is commonly referred to as warming-up effect (cf. Tanzer, 2005). According to Anastasi (1981), retesting reduces individual differences in test familiarity and therefore leads to a more fair and accurate estimate of