Performance Tests in Driver Assessment

Margit Herle, Matus Sucha, Fatima Pereira da Silva &Thomas Wagner

KEYNOTES AND RECOMMENDATIONS

02/2015

Keynotes and recommendations is a series of statement papers written by the Traffic Psychology International Group, with the aim of quality assurance and harmonization of Traffic Psychology processes within the EU.

1. Introduction

Driving-related psychological assessment affects two legal rights: the protection of the general public from unsuitable drivers, and the right of the individual to mobility. We all live in a society where mobility often means access to a social and professional life, which is why any intrusion upon the independent mobility of an individual has both legal consequences as well as massive personal implications.

Within der EU countries different products for driver assessments are used. Since the lack of EU-wide regulations of minimum standards in driver assessment tools, these different products naturally vary in their quality criteria. This Keynote is a first step towards consistent quality criteria in traffic psychological assessment.

Basically, quality assurance measures must extend across the entire diagnostic process, in particular the following three aspects:

- The qualifications of the persons involved
- The observance of processes and rules (data protection, safeguarding of personal integrity, type of feedback on results), from planning through to the diagnostic decision.
- The quality of methods used, in particular the fulfilment of quality criteria.

The following keynotes and recommendations aim at the third point, the quality of the methods used. The suggestions should come across as minimum standards. Beyond this, higher sophisticated criteria are possible.

To find out minimum standards, the authors screened already published guidelines for psychology, as well as legal acts are taken into account. These keynotes set out the requirements for the psychological assessment of drivers to comply with the conditions of test information systems (e.g. BUROS; www.unl.edu/buros; ETS; www.ets.org/testcoll/index.html) and evaluation systems (e.g. the EFPA Review Model[1], APA[2], ITC[3], COTAN[4] and TBS-TK[5], as well as legal acts from Germany and Austria and the German guidelines of driver assessment.

---

[1] EFPA review model for the description and evaluation of psychological and educational tests
[2] American Psychology Association
[3] International Test Commission
[4] Committee on Test Affairs Netherlands
[5] TBS-TK Testkuratorium. (2010). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologenvereinigungen. Revidierte Fassung vom 09. September 2009. *Psychologische Rundschau, 61, 52-56.*

(1) EFPA Test Review Model

The EFPA Test Review Model provides a description and a detailed assessment of psychological assessment inventories used in the area of Work, Education, Health and other contexts. The latest version of the EFPA Test Review Model has been prepared by a Task Force of the EFPA Board of Assessment.

(2) TBS-TK

These guidelines are published by the Federations of German Psychological Associations (DGPs and BDP) for supporting test authors, publishers, providers and test reviewers. Two independent reviewers evaluate the test. The procedure is composed of 3 steps. At first the test manual is screened, including diagnostic ambition, theoretical frame and test application. Step two targets the classification of the test by using official categories, e.g. taken from EFPA-System. At least the test is evaluated by using a detailed list TBS-TK-system (including objectivity, norms, reliability, validity, and further criteria as interference-sensitivity, faking good or bad). Before the final report is written, the test author may give his comments to the current review results. At least the test review is published by two different prominent journals (Report Psychologie, Psychologische Rundschau).

(3) APA:

The APA Guidelines are published by the American Psychology Association and deal with Evaluation, diagnostic process as well as the interpretation of psychological inventories. They address test publisher as well as test user.

(4) ITC

Guidelines of the International Test Commission: Providing detailed recommendations concerning computerized testing, online testing, quality control, test use and others.

(5) COTAN

The COTAN Guidelines are the Dutch Rating System for Test Quality. This rating system evaluates test inventories based on the following seven criteria: theoretical basis, quality of the testing materials, comprehensiveness of the manual, norms, reliability, construct validity, and criterion validity. The criteria are very similar to those ones of German TBS-TK-system. The COTAN Guidelines help the test users to evaluate the test at hand, but imposes obligations only on test publishers.

(6) Legal acts from Germany and Austria

Within the EU, Germany and Austria have the most elaborated regulations concerning inventories which are used for traffic psychological assessment. Thus it

makes sense, to take these legal requirements into account. Relevant sections refer to Austria: FSG-GV (BGBl. II Nr. 322/1997 and BGBl. II Nr. 64/2006) and Germany: Driving Licence Ordinance (FeV Anlage 5 (e.g. BGBl. I from 23.04.2014, p. 348), whereas the gap between legal standards and their practical implementation is bridged by assessment criteria (Urteilsbildung in der Fahreignungsbegutachtung – Beurteilungskriterien, 3. Auflage 2013, Hrsg.: DGVP/DGVM, Bonn: Kirschbaum Verlag). They represent scientific principles and the current level of knowledge (VkBl. 2014, S. 132), recommended by Annex 4a of Driving Licence Ordinance (FeV).

2. Preface

Responsibilities test publisher

The test publisher has to ensure the accessibility of the procedure instructions to the user. The procedure instructions have to keep up to date. Changes in tests have to be documented clearly and periodically. Empirical references as well as indication of psychometric standards have to be accessible for the test user, particularly detailed information about reliability, validity and fairness.

Test User Responsibilities

The test user has to ensure to have the latest information; he is responsible of being up to date concerning changes in the test or new test-developments, changes in guidelines, legislation and policy. He is responsible for choosing the right test and the right norm population. He has to have knowledge about test administration, test scoring and test interpretation of the used inventories.

3. Psychometric standards in traffic psychological assessment

(7) Objectivity

The administrator, evaluation and interpretation objectivity have to be assumable. All relevant information has to be provided by the test-publisher.

Administrator objectivity

Test administrator objectivity exists when the respondent's test behavior, and thus his test score, is independent of variations (either accidental or systematic) in the behavior of the test administrator.

Evaluation objectivity

Evaluation objectivity exists when the test performance, in terms of responses to each individual item, leads to the same result, regardless of who evaluates the test.

Interpretation objectivity

Interpretation objectivity exists when the same conclusion is drawn from particular test results (test scores) even when they are interpreted by different people.

(8) Reliability

Reliability means the degree to which scores are free from measurement error variance. When evaluating the reliability, not only the numerical value of the reliability coefficient is meaningful. Also the quality of the empirical research (e.g. size of sample, matching to target group) has to be taken into account as well as results from independent repeat and comparable research. Information about the reliability coefficients (kind, value and actuality) has to be provided by the test-publisher.

Aspects of reliability

Reliability has to be tested empirically and several different aspects of reliability (e.g. internal consistency, retest reliability, equivalence reliability…) should be available. Reliability coefficients should be calculated as Cronbach's Alpha, Lambda2, Greatest lower bound. As a rule, internal consistency and retest-reliability should be documented, even for important subgroups (e.g., elderly drivers).

Reliability coefficients

Reliability coefficients vary between 0 and 1 whereas 1 indicates the strongest correlation. Reliability coefficients of performance tests should be at least 0,70.

Periodical check
Statistical values have to be checked periodically, whether they are still valid. This examination should be done at least every 8 years.

Retesting
Information on possible effects on re-testing has to be provided by the test publisher. In case of re-testing (e.g. disruptive factors during the assessment) the psychologist has to decide based on this information, if re-testing is acceptable, and if, after which duration.

(9) Validity

Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests.

Aspects of validity
Due to the extensive consequences of the result of a driver assessment, more than one validity assessment has to be performed. Beyond convergent and divergent

validity the validation procedure has to be assessed among the driver population using correlation and/or experimental designs. Adequacy of the research design as well as the demographic aspects of the research population has to be given. Only in reasonable cases the analysis of the criterion validity can be left out, e.g., when there is no clear legal definition for the term aptitude (as in Germany) which could be operationalized. This has to be documented by the test publisher.

Validity coefficients

Fitness to drive is a multidimensional construct, despite this, in several countries it has to be predicted via one-dimensional psychological inventories. This leads unavoidable to a systematic bias which affects low correlation coefficients between the multidimensional external criterion and the one-dimensional inventories. This is why validation coefficients of r= 0.3 can be considered as sufficiently.

(10)   Reasonability

Due to the high personal importance of the driving licence, most of the clients are undergoing an enormous stress situation during a traffic psychological assessment. It is in the responsibility of the Psychologist to reduce the stress factors as good as possible through an adequate test-setting. Psychologists are responsible to help the client getting over his test nervousness as good as possible. Moreover the Psychologist has to ensure, that he does not use tests beyond the psychological issue.

(11)   Usefulness

It is in the responsibility of the Psychologist, to take tests, which matches the psychological issue.

(12)   Fairness

It is in the responsibility of the Psychologist to choose inventories, which offers the best fairness to traffic psychological related problems (e.g. Age, usability, nonverbal intelligence).

(13)   Ease of handling

This issue addresses to the interaction between test applicant and the test system. Relevant usability features should be: Short working time of test application, interference-sensitivity, minimizing faking strategies, saving individual data by password, shielding the entire clients' data file by specific security mechanisms, in order to avoid manipulations after the test procedure.

(14)   Norms:

Sizes and sources of norms groups have to be described clearly and detailed. Information about size, representativeness and actuality of norms has to be provided by the test-publisher.

- **Sample size**

Samples of less than 200 are regarded as too small, as the resolution provided in the tails of the distribution will be very small. The scope of the norm sample depends on the test material, but should not be below N=300 in any case.

- **Norm population**

As a norm sample, a representative age-independent norm sample has to be used. Note: Representative refers to the population which acts within the traffic-environment. It has to be justified by the test publisher, in which way the provided norm sample can be seen as representative. For Group1 and Group2 drivers, different cut offs should be used and a higher cut off be required for Group 2 drivers (e.g., PR 16; PR 33).

- **Actuality of norms**

Norm check has to be done at least every 8 years. In case of a norm-shift, the old norms have to be replaced by new norms.