On the psychometric quality of new ability tests administered using the WWW

Oliver Wilhelm[1], *Michael Witthöft*[2], *Andreas Größler*[3], *& Patrick McKnight*[4]

Internet or ordinary ability testing methods differ with respect to the nature of the testing situation on several important factors (e.g., person mediated communication and subject behavioral control). Ordinary testing methods ought to be preferred in psychological experiments, however, there are important considerations for choosing alternative (i.e., internet) testing methods. The conditions that may warrant the use of internet testing include necessary resources, availability of experts and bridging geographical distances.

In some applied settings, e.g., personel selection, the goal of measurement is to select the best person from a pool of highly capable subjects. With ordinary testing methods there are frequently not enough experts available or the experts are too costly to assist in constructing, calibrating, evaluating, and validating measurement instruments. New knowledge tests were developed for internet distribution to make use of available experts and gather the largest sample possible for item evaluation and test construction. The new knowledge tests were developed in science and economy - two domains where, to our knowledge, no convenient instruments are available. Each domain had two parallel tests in two languages and used self selected samples participating via internet.

The quality of data was critically evaluated on several levels with different methods. In addition to procedures from classical test theory, probabilistic procedures were applied to the data. Both domain tests were evaluated for a priori internal structure. Specifically, difficulties and validities on an item level could be compared to expert ratings of difficulty and validity for the economics tests. In both domains data gained by internet administration could be compared with data gained

1 Lehrstuhl Psychologie III, Universität Mannheim, D-68131 Mannheim, Phone ++49 (0) 621 181 2143, Fax, ++49 (0) 621 181 3997, email: wilhelm@tnt.psychologie.uni-mannheim.de

2 Lehrstuhl Psychologie II, Universität Mannheim, D-68131 Mannheim witthoeft@tnt.psychologie.uni-mannheim.de

3 Industrieseminar, Universität Mannheim, D-68131 Mannheim agroe@is.bwl.uni-mannheim.de

4 CESATE - VA Puget Sound Health Care System, University of Washington, Seattle, WA 98166, USA, pem@theriver.com

using ordinary methods. In the science test the internal structure was tested for robustness using a quasi-experimental manipulation. Finally the equivalence of parallel tests was determined comparing relevant statistics. To preliminarily validate the new measurement instruments, biographical questions (e.g., education level, relevant prior knowledge, and proxy variables including reading of newspapers and journals) for the interest in the respective domains were used. The results of the analysis essentially support the supposition that internet administration yields quality data, hence, it seems appropriate to use the internet as a testing medium for the means outlined here. Future investigations include calibrating test results to real world behavior. Additionally, future efforts to vary the item format from traditional yes/no or multiple choice response choices will be tested along with an expansion of the knowledge structure.

Keywords: Calibration, Equivalence, Validity, knowledge tests