Schmollers Jahrbuch 127 (2007), 183 – 192
Duncker & Humblot, Berlin

Assessing Cognitive Capacities in Computer-Assisted Survey Research: Two Ultra-Short Tests of Intellectual Ability in the German Socio-Economic Panel (SOEP)

By Frieder R. Lang, David Weiss, Andreas Stocker, and Bernhard von Rosenbladt

1. Introduction

Social and economic survey research can benefit immensely from the inclusion of psychometric variables that reflect individual differences in stable personality traits such as intelligence performance. The construct of intelligence is without doubt the most thoroughly investigated trait in personality and psychometric research over the past century. Since Binet and Simon (1905) and Spearman (1904) published their now-classic works, intelligence research has yielded an enormous, sophisticated, and diverse spectrum of performance test batteries, most covering a broad and heterogeneous set of tasks that allow for a fine-tuned and precise assessment of individual differences in intelligence. In this research tradition, there is a long and still-open debate as to whether a general factor underlies all cognitive abilities, or whether the individual's general intellectual ability can be better represented by the set of several more or less correlated specific sub-factors of intelligence[1].

Survey researchers have generally – and prematurely – assumed that there is no good way of using cognitive performance tasks to provide valid estimation of individual differences in intellectual ability. Such reasoning is based on two considerations. First, the assessment of intellectual ability requires a highly standardized testing procedure to ensure objectivity and comparability of individual performance. Second, large-scale surveys often suffer from enormous economic constraints that require time-efficient and ultra-short assessments of any construct or trait of interest. Consequently, any measurement of intellectual ability that takes more than five minutes will have literally no chance of being included in a costly large-scale survey. To overcome these

[1] In this article, we use the term intellectual ability (and interchangeably, cognitive capabilities or cognitive abilities) to refer to any performance of the individual that indicates the assumed latent trait of intelligence.

problems, we have developed two ultra-short cognitive performance tasks that should allow a rough assessment of general intellectual ability, in addition to assessing specific components of intellectual ability according to the life-span theory of intellectual functioning (Lindenberger, 2002). In the present paper, we report findings on the reliability and validity of these two ultra-short cognitive tests.

We contend that the assessment of intellectual ability among survey participants provides useful and innovative perspectives in socio-economic research, for example, with respect to the potential effects of intellectual ability on wages and income (Anger / Heineck, 2006; Cawley et al., 2000), and on career advancement (McManus et al., 2003). Unfortunately, however, to the best of our knowledge there exists no established standard performance task that provides a reliable and valid estimation of intellectual ability and, at the same time, is feasible for use in a survey interview. The present study aims at closing this gap and, thus, proposes two newly adopted ultra-short tests of cognitive performance for use in computer-assisted personal interviewing (CAPI) and in large-scale surveys such as the German Socio-Economic Panel (see Wagner et al., 2007; Schupp / Wagner, 2002).

In selecting the two ultra-short tests described below, we used the theoretical framework of life-span psychology (Baltes et al., 1999; Lindenberger, 2002), which distinguishes between two components of intellectual functioning: the mechanics and the pragmatics of intellectual ability. Taken together, the mechanics and the pragmatics represent all of an individual's cognitive abilities that are required for acting or performing competently over the life course. The mechanics of cognition, on the one hand, pertain to the hard-wired, biology-related capacities of information processing, a component that is captured well by measures of perceptual speed (e.g., Lindenberger, 2002; Lindenberger / Baltes, 1995). On the other hand, the pragmatics of intellectual performance refer to educational and experience-related competencies, reflected, for example, in knowledge-based indicators such as fluency (Lindenberger, 2002). Hence, we use the newly adopted Symbol-Digit-Test (SDT) as a speed-constrained measure of mechanics (i.e., information-processing capacities), and a CAPI-version of the Animal Naming Task (ANT) as a test of word fluency (i.e., pragmatics). Both tests were modified and adapted in order to fit the requirements of computer-assisted personal interviewing (CAPI).

A major goal in this study was to ensure that the newly developed tests are easy to apply in survey contexts without a need for special training of interviewers while at the same time keeping the sources of error within CAPI comparatively low. In the present study, we report findings of a six-week test-retest multi-method study with 119 participants that examined the reliability and validity of the two ultra-short cognitive tests together with an extensive cognitive test battery.

2. Method

2.1 Sample

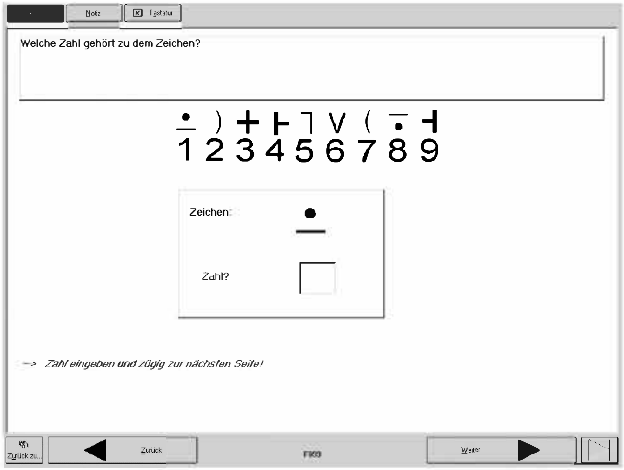
Study participants included 53 men and 66 women aged between 18 to 69 years (M = 31, SD = 11.1). The educational status of participants was comparatively high: Nearly half of the participants had graduated from high school (i.e. Abitur) and about 15% held a university degree. Most participants reported having had prior experience with personal computers, but there were also participants without any prior computer experience.

The *sample recruitment strategy* aimed at generating a heterogeneous sample of study participants in order to obtain considerable variance in intellectual performance ability. Interviewers (students) contacted the participants directly in diverse public places–for example, at the city employment office or in front of a supermarket. While this strategy may not have yielded a representative sample, it served well to achieve heterogeneity in a spectrum of relevant socio-economic variables, including education, income, and employment status.

2.2 Two Ultrashort Tests of Cognitive Capacity

The Symbol-Digit-Test (SDT) was constructed after the Symbol-Digit-Modality-Test (Smith, 1973 / 1995). For the CAPI version, no paper and pencil is needed. For the CAPI-version of the SDT, respondents give their answers directly by typing the correct number on the PC keyboard. The test requires individuals to match numbers with graphical symbols as quickly as possible. In the CAPI version of the SDT, a screen image is presented, showing a band of nine graphical symbols with numbers ranging from 1 to 9 below. After reading brief test instructions, participants respond to each symbol that pops up one after the other in the centre of the computer screen (while at the same time the full band of nine symbols with numbers remain visible on top of the screen). When a symbol appears in the centre, respondents have to enter the correct number assigned to that symbol in the band above as quickly as possible. The test ends automatically after 90-seconds. The CAPI software calculates the number of correct responses as well as the response latencies (Rosenbladt / Stocker, 2005). The maximum amount of correctly assigned digits provides an estimate of the respondent's perceptual information-processing speed.

Knowledge-based word fluency was assessed with the Animal Naming Task (Lindenberger / Baltes, 1995). In the CAPI-adapted Animal Naming Task (ANT) participants name as many different animals as possible during a 90-second time interval. When participants name an animal correctly for the first time, interviewers click on a button on the computer screen. Interviewers make separate entries for each animal named repeatedly and for ambiguous responses.



Note: Participants are asked to type the correct number on the keyboard, in this example a "1". After that, a new screen appears with a different symbol shown in the middle.

Figure 1: Screen shot of the first page of the CAPI Version of the Symbol-Digit Test

The total time needed to complete both tests of intellectual functioning, including the time needed for instructing respondents and setting up computers, is approximately five minutes.

2.3 Cognitive Test Battery

In order to determine the validity of the two short cognitive tests, participants also completed a battery of additional cognitive performance tasks testing perceptual speed, reasoning, knowledge, memory, and attention (for a detailed description, see Lang, 2005). In the present study, we focus on results related to convergent and discriminant validation of the two ultrashort tests with the digit-symbol test (Tewes, 1994), the digit-letter test (Lindenberger/Baltes, 1995), a vocabulary knowledge test (Lehrl, 1995), and a word recognition test (Horn, 1983).

3. Results

Reliability of the two ultrashort cognitive performance tests was tested with indicators of internal consistency and retest-reliability. In addition, we examined the convergent and discriminant validity of the tests with a set of external criterion variables from a cognitive test battery.

3.1 Internal Consistency and Retest-Reliability

In order to obtain a measure of internal consistency, the two 90-second performance tests were subdivided into three test intervals of 30 seconds each. Cronbach's Alpha as well as the mean inter-item correlation (MIC) coefficients obtained for the three 30-second subtests in each of the two tests at first and second measurement are reported in Table 1. Internal consistency of the standardized tests reached an Alpha of .87 at first measurement and .95 in the retest of the Digit Test (DT) with a Test-Retest-Coefficient of .55. In the Animal Naming Task (ANT) Alpha reached .64 and .68, respectively, and a test-retest coefficient of .46. After correcting the test-retest coefficient for unreliability among the three 30-second tests on both occasions, the test-retest coefficient was .70.

Table 1

Internal Consistency of the Symbol-Digit Test (SDT) and the Animal Naming Task (ANT)

	First Occasion (N = 119)		Retest Occasion (N = 116)	
	SDT	ANT	SDT	ANT
Standardized Alpha	.87	.68	.95	.64
MIC	.70	.42	.87	.37
Test-Retest-Coefficient			.55*	.46*

Note: MIC = Mean-Inter-Item Correlation, * Test-retest coefficient is .60 for the SDT and .70 for the ANT when correcting for unreliability at both occasions.

Nevertheless, the findings lead to the conclusion that the internal consistency of the Animal Naming Task is somewhat limited. One explanation is that the ANT is more susceptible to situational effects than the SDT and might also be affected by the interviewer's attention level during the CAPI entry phase. However, it needs to be taken into consideration that there is greater heterogeneity in performance when testing crystallized or pragmatic components of cognitive abilities. This may also account for a generally lower level of internal consistency.

In addition, all responses of participants in the Animal Naming Task (ANT) were tape-recorded and subsequently transcribed in order to obtain an objective test of interviewer performance while entering the responses during the interview session. The number of correctly named animals in CAPI differed from the tape-recorded entries by –0.4 animals on average (SD = 4.3) with a reliability coefficient of .89.

3.2 Construct Validation of the Two Ultra-short Cognitive Tests

Convergent and discriminant validation of the two cognitive performance tests was based on a set of well-established tests from differing cognitive test batteries. Convergent validation of the SDT was tested with two tests of perceptual speed of information processing (Digit-Symbol Test; Digit-Letter Test). Convergent validation of the ANT relied on two knowledge tests of vocabulary knowledge and word recognition. The upper part of Table 2 displays the correlation coefficients for the speed and knowledge tests with SDT and ANT. Both show an acceptable pattern of convergent and discriminant validity. The Symbol-Digit Test proves to differentiate speed-related capacities from knowledge-based performance in the fluency tasks. Correlation coefficients with the Animal Naming Task do not differentiate the constructs as well, but show an acceptable pattern of convergence.

Table 2

Validation of the Symbol-Digit Test (SDT) and the Animal Naming Task (ANT)

Construct / Instrument	First Occasion (N = 119)		Retest Occasion (N = 116)	
	SDT	ANT	SDT	ANT
Speed				
Digit-Symbol Test	**.70**	*.32*	**.54**	*.34*
Digit-Letter Test	**.68**	*.30*	**.56**	*.29*
Fluency				
Vocabulary Knowledge	*.15*	**.33**	*.09*	**.39**
Word Recognition	*.33*	**.36**	*.23*	**.30**
External Criteria of Validation				
Age Cohort	–.56	–.03	–.36	.05
Education (years)	–.05	.42	.08	.45
*School Grade Average**	–.09	–.11	–.11	–.25
Expertise of computer use (years)	.34	.09	.12	.13

Note: Coefficients > |.18| are significant at p < .05. Coefficients show synchronous correlation at the respective measurement occasion. * Average of last school grades in Mathematics, English, and German (higher grades indicate lower performance).

The lower part of Table 2 shows validation with external criterion for both ultra-short cognitive tests. The observed pattern of correlations underscores the external validity of the two cognitive performance tests. SDT scores are strongly associated with chronological age, whereas ANT scores are basically unrelated to the age cohort, but show associations with educational variables in the expected direction and strength. In general, the associations of SDT and ANT with external variables substantiate the aim of this study to represent the width of relatively culture-free mechanical components and the width of the education-based pragmatic component of intellectual functioning. Performance in the two ultra-short tests showed a weak but substantial correlation ($r = .18$ in the first test, $r = .48$ in the retest).

4. Discussion

The two newly adopted ultra-short tests of intellectual ability provide a rough estimation of individual differences in general intellectual ability. The present research shows that the two ultra-short cognitive performance tests achieve an acceptable reliability and sufficient validity in a heterogeneous sample of adults. In particular, the Symbol-Digit Test (SDT) provides a fast and valid assessment of mechanical cognitive capabilities that are relatively unrelated to educational or cultural variables. The SDT proves to be robust against situational effects of the interviewing situation. Furthermore, minimal interviewer training is required when applying this test. In comparison, the Animal Naming Task (ANT) requires greater attention and training of interviewers, and therefore achieves only limited reliability and validity, when seeking to assess a pragmatic component of intellectual ability. It was shown, however, that ANT yields additional information on pragmatic aspects of cognitive performance that is not captured by the SDT. The ANT proves to be more closely associated with the experiential and education-related aspects of intellectual performance. In sum, both tests together achieve acceptable reliability and validity, and provide a good basis for estimation of cognitive capacity and intellectual ability within large-scale surveys using CAPI techniques.

Findings showed that the two ultra-short cognitive tests serve well to reflect the assumed life-course pattern of mechanics and pragmatics of intellectual functioning in a reliable and robust way. For example, the high negative correlation of age cohort with the scores in the Symbol-Digit Test is reflective of the expected and often observed age-associated decline in mechanical components of intellectual functioning. In contrast, the expectation that pragmatic performance is relatively robust against age-related cognitive decline is reflected in the finding that performance in the ANT is uncorrelated with chronological age.

One caveat needs to be considered when applying the CAPI-version of the Animal Naming Test in survey research. The ANT turns to be more vulnerable to situational effects and interviewer training effects. As a consequence, we found relatively low coefficients of retest stability and internal consistency for the ANT as compared with the SDT. Interviewer attention and motivation may be a source of systematic bias in some interviewer situations. One possible solution would be to include attention capacity and variables of interviewer motivation (e.g., satisfaction, a baseline perceptual speed). However, the observed correlations with external criteria such as education suggest that the test captures a substantial amount of the valid information related to the pragmatic components of cognitive abilities.

We contend that overall, the two ultra-short cognitive tests meet the requirements of tests aiming to assess general intellectual ability in the context of a CAPI-based survey. In this context, the Animal Naming Test may serve as a test of education-prone cognitive pragmatics in intellectual functioning in addition to the more speed-related mechanics of relatively culture-free cognitive capabilities.

The Symbol-Digit Test and the Animal Naming Task also proved to sufficiently represent diverse types of intellectual performance as measured with a broad spectrum of well-established tasks from intelligence test batteries. While the two tests serve well to represent general intellectual ability, caution is needed with respect to the specific components thereof. For example, the two tests may not serve well for prediction of specific aptitude effects in particular life circumstances. Also, the two tests are not suited for case-wise diagnostics of individual cognitive potential: rather, they reflect the distribution and rank-order of relatively stable individual differences in cognitive performances. Finally, both tests, like most cognitive performance tasks, are somewhat vulnerable to situational influences and interviewer effects, and thus need to be implemented with great caution in CAPI protocols of surveys. Despite such shortcomings, both tests provide a robust, albeit rough, estimation of two general components of cognitive capacity.

References

Anger, S. / *Heineck,* G. (2006): Cognitive abilities and labour market outcomes – First evidence for Germany. DIW Berlin Diskussionspapier No. 655.

Baltes, P. B. / *Lindenberger,* U. / *Staudinger,* U. M. (1998): Life-span theory in developmental psychology, in: R. M. Lerner (ed.): Handbook of child psychology (5th edition, Vol. 1: Theoretical models of human development, New York, 1029–1143.

Binet, A. / *Simon,* T. (1905): Methodes nouvelles pour le diagnostique du niveau intellectuel des anormaux, Année Psychologique 11, 191–244.

Cawley, J. / *Heckman,* J. / *Vytlacil,* E. (2001): Three observations on wages and measured cognitive ability, Labour Economics (8) 4, 419–442.

Galton, F. (1988): Co-relations and their measurement, chiefly from anthropometric data, Proceedings of the Royal Society of London 45, 135–145.

Horn, W. (1983): Leistungsprüfungssystem: L-P-S, Göttingen.

Lang, F. R. (2005): Erfassung des kognitiven Leistungspotentials und der “Big Five” mit Computer-Assisted-Personal-Interviewing (CAPI): Zur Reliabilität und Validität zweier ultrakurzer Tests und des BFI. Deutsches Institut für Wirtschaftsforschung (DIW Berlin), Research Notes 9.

Lehrl, S. (1995): Mehrfachwahl-Wortschatz-Intelligenz-Test (MWT), Balingen.

Lindenberger, U. (2002): Erwachsenenalter und Alter, in: R. Oerter / L. Montada (Hrsg.), Entwicklungspsychologie (5. Auflage), Weinheim, 350 – 391.

Lindenberger, U. / *Baltes,* P. B. (1995): Kognitive Leistungsfähigkeit im Alter: Erste Ergebnisse aus der Berliner Altersstudie, Zeitschrift für Psychologie 203, 283 – 317.

McManus, I. C. / *Schmithers,* E. / *Partridge,* P. / *Keeling,* A. / *Fleming,* P. R. (2003): A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study, British Medical Journal 327, 139 – 142.

Rosenbladt, B. von / *Stocker,* A. (2005): Testerhebung zum SOEP 2006. “Persönlichkeit und Alltag”, Verhaltensexperimente, Retest-Studie. München: TNS Infratest Sozialforschung (PB20 – 44167).

Schupp, J. / *Wagner,* G. G. (2002): Maintenance of and Innovation in Long-term Panel Studies: The Case of the German Socio-Economic Panel (GSOEP). Allgemeines Statistisches Archiv 86, 163 – 175.

Smith, A. (1973 / 1995): Symbol Digit Modalities Test. Los Angeles, California.

Spearman, C. (1904): “General intelligence”, objectively determined and measured. American Journal of Psychology 15, 201 – 293.

Tewes, U. (1994): HAWIE-R: Hamburg Wechsler Intelligenztest für Erwachsene, Revision 1991, Bern.

Wagner, G. G. / *Frick,* J. R. / *Schupp,* J. (2007): The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. Schmollers Jahrbuch 127 (in this issue).