



DIW Berlin

Deutsches Institut
für Wirtschaftsforschung

Research Notes

2005

9

**Erfassung des kognitiven
Leistungspotenzials und der „Big Five“ mit
Computer-Assisted-Personal-Interviewing
(CAPI):
Zur Reliabilität und Validität zweier
ultrakurzer Tests und des BFI-S**

Frieder R. Lang

IMPRESSUM

© DIW Berlin, 2005

DIW Berlin
Deutsches Institut für Wirtschaftsforschung
Königin-Luise-Str. 5
14195 Berlin
Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
www.diw.de

ISSN 1860-2185

Alle Rechte vorbehalten.
Abdruck oder vergleichbare
Verwendung von Arbeiten
des DIW Berlin ist auch in
Auszügen nur mit vorheriger
schriftlicher Genehmigung
gestattet.



Research Notes 9/2005

Frieder R. Lang *

unter Mitwirkung von Dorothee Hahne, Stefanie Gymbel,
Stefan Schröpfer und Katharina Lutsch

Erfassung des kognitiven Leistungspotenzials und der „Big Five“ mit Computer-Assisted-Personal-Interviewing (CAPI): Zur Reliabilität und Validität zweier ultrakurzer Tests und des BFI-S

Berlin, Dezember 2005

* Martin-Luther-Universität Halle-Wittenberg, Institut für Psychologie, flang@psych.uni-halle.de

Inhaltsverzeichnis

1	Einleitung	3
2	Die Zweikomponententheorie der intellektuellen Fähigkeiten	5
3	Entwicklung der kognitiven Funktionstests für die vorliegende Studie.....	7
4	Darstellung des Vorgehens und Datenbasis	8
4.1	Stichprobenmerkmale	8
4.2	Erhebungsinstrumente der Validierungsstudie	9
4.2.1	Kurzversion des CAPI-Pretest Instruments	11
4.2.2	Wahrnehmungsgeschwindigkeit (<i>speed</i>)	11
4.2.3	Schlussfolgerndes Denken (<i>reasoning</i>)	11
4.2.4	Wortschatz und Wissen (<i>knowledge</i>)	12
4.2.5	Gedächtnis und Erinnern (<i>memory</i>).....	12
4.2.6	Aufmerksamkeit	13
4.2.7	Persönlichkeitsinventar (<i>NEO-FFI</i>)	13
4.3	Ablauf der Erhebung in Halle (Saale).....	13
5	Testgüte der Tiere-Nennen-Aufgabe und des Zeichen-Tests	15
5.1	Kennwerte, Objektivität und Reliabilität der Tiere-Nennen-Aufgabe (TA).....	15
5.2	Kennwerte und Reliabilität des Zeichen-Tests	19
5.3	Validierung der Tiere-Nennen-Aufgabe und des Zeichentests.....	20
6	Testgüte des BFI-S.....	25
7	Zusammenfassung und Ausblick	29
	Literatur.....	31

Tabellenverzeichnis

Tab. 1: Beschreibung der Stichprobe (Erste Erhebung; N = 119)	9
Tab. 2: Übersicht der verwendeten Erhebungsinstrumente zur Validierung der CAPI- Tests.....	10
Tab. 3: Verteilungskennwerte, Trennschärfe und Reliabilität der Tiere-Nennen Aufgabe	17
Tab. 4: Interne Konsistenz der zwei kognitiven Funktionstests.....	19
Tab. 5: Verteilungskennwerte, Trennschärfe und Retest-Reliabilität des Zeichentests.....	20
Tab. 6: Konstruktvalidität der kognitiven CAPI-Tests	21
Tab. 7: Multitrait-multimethod-Matrix	23
Tab. 8: Außenkorrelationen der kognitiven Funktionstests des CAPI in der Erst- und Retesterhebung (synchrone Korrelationen).....	24
Tab. 9: Faktorladungen (Varimax PCA) der BFI-S15 Items je Erhebungszeitpunkt	25
Tab. 10: Innere Konsistenz der fünf Persönlichkeitsskalen des BFI-S	26
Tab. 11: Mittelwertsstabilität und Retest-Reliabilität der Persönlichkeitsskalen (BFI-S) nach 6 Wochen	27
Tab. 12: Synchrone Konstruktvalidität des BFI-S zur ersten Erhebung und im Retest.....	27

1 Einleitung

Aufgrund der zunehmend interdisziplinären Ausrichtung und entsprechenden Erweiterungen zahlreicher sozial- und verhaltenswissenschaftlicher Untersuchungen und Panelstudien gibt es ein großes Interesse an einer effizienten, gleichwohl aber robusten und validen Erfassung etablierter Verhaltens- und Persönlichkeitskonstrukte der empirischen Psychologie, wie beispielsweise der kognitiven Leistungsfähigkeit und den fünf „großen“ Persönlichkeitsmerkmalen Extraversion, Verträglichkeit, Gewissenhaftigkeit, Neurotizismus und Offenheit für Erfahrung (Lang, Lüdtke & Asendorpf, 2001). Damit verbunden ist der Anspruch in den Sozialwissenschaften, eine an sozialstrukturellen Überlegungen ausgerichtete Forschungsperspektive auf mögliche Einflüsse der kognitiven Verhaltens- und Handlungskapazitäten des Individuums – das als „statistische Einheit“ anonym bleibt – auszuweiten (vgl. auch Solga et al. 2005). Von beispielhaftem Interesse sind theoretische Überlegungen zum Zusammenspiel sozialstruktureller Bedingungen (z.B. Bildung, Erwerbssituation) mit je individuellen Verhaltensdispositionen und –kapazitäten (z.B. Leistung, Persönlichkeit) sowie deren mögliche Auswirkungen auf differenzielle Entwicklungs- und Lebensverläufe (vgl. Heckhausen, 1999; Lang & Heckhausen, in Druck). Die Abschätzung kognitiver Verhaltens- und Leistungsparameter erlaubt es darüber hinaus solche Teilgruppen in der Bevölkerung zu identifizieren, bei denen die Qualität der Selbstauskünfte aufgrund kognitiver Belastungen strukturell abweichen (z.B. bei älteren Menschen). Die Berücksichtigung psychologischer und kognitiver Indikatoren eröffnet neben inhaltlichen Erkenntnisgewinnen auch zahlreiche Möglichkeiten einer methodologisch orientierten Betrachtung von Survey- und Paneldaten.

Zielsetzung der Auswahl und Modifikation von erprobten kognitiven Tests für das Sozio-oekonomische Panel (SOEP, vgl. Projektgruppe SOEP, 1995; Schupp & Wagner, 2002) ist es, solche Aspekte der kognitiven Funktionstüchtigkeit zu erfassen, die eine gute Abschätzung von individuellen Differenzen im Bereich der kognitiven Fähigkeiten gewährleisten und damit einen über Bildungs- und Erwerbseinflüsse hinaus gehenden gruppenspezifischen Vorhersage- und Erkenntnisgewinn in Panel- und Survey-Erhebungen versprechen. Dies wird erreicht, wenn neben kulturfreien auch stärker bildungsnahe Indikatoren der intellektuellen Kompetenz kombiniert einbezogen werden, um so verschiedene Komponenten der mentalen Fähigkeiten differenzieren zu können. Im Rahmen dieses Anspruchs wurden gemeinsam mit dem SOEP (Gert G. Wagner und Jürgen Schupp) und mit Infratest Sozialforschung (Bernhard von Rosenblatt und Andreas Stocker) zwei ultrakurze Tests zur Erfassung von Indikatoren

der Wahrnehmungsgeschwindigkeit („speed“) und der Wortflüssigkeit („word fluency“) entwickelt, die im Kontext etablierter und standardisierter Testbatterien (HAWIE-R; BASECog) überprüft wurden.

Der vorliegende Bericht informiert über die theoretischen Grundlagen und die Validierung der beiden kognitiven Leistungstests sowie des für das SOEP entwickelte kurze Big-Five-Inventory (BFI-S) zur Erfassung der Persönlichkeitsstruktur (Gerlitz & Schupp, 2005).

2 Die Zweikomponententheorie der intellektuellen Fähigkeiten

In den an einer Lebensspannenperspektive orientierten kognitiven Wissenschaften und *life sciences* hat sich in den vergangenen Jahrzehnten die Zweikomponententheorie der intellektuellen Fähigkeiten empirisch gut bewährt (Baltes, 1997; Lindenberger, 2002). Ausgangspunkt ist die Annahme, dass sich alle kognitiven Leistungen des Menschen im Wesentlichen einem von zwei übergeordneten Fähigkeitsbereichen zuordnen lassen. Die erste Komponente umfasst die durch biologische Prozesse geformte Mechanik der kognitiven Funktionstüchtigkeit des Menschen, die zweite Komponente beschreibt die von kulturellen Bedingungen abhängige Pragmatik des Intellekts. Die Zweikomponententheorie der intellektuellen Fähigkeiten beschreibt die lebenslange kognitive Entwicklung als ein dynamisches Wechselspiel kultureller und biologischer Einflüsse, welches unterschiedliche Verlaufsmuster der intellektuellen Kapazitäten im Lebenslauf begründet.

Die Mechanik des Denkens bezieht sich auf Leistungsunterschiede, beispielsweise in der Geschwindigkeit, Genauigkeit, Aufnahmekapazität, Koordination und Inhibition elementarer Grundprozesse des Denkens. Hervorgehobene Beispiele sind insbesondere die Wahrnehmungsgeschwindigkeit, die Kapazität des Arbeitsgedächtnisses oder die Fähigkeit zu schlussfolgerndem Denken. Der Entwicklungsverlauf mechanischer Komponenten folgt in der Regel einem Anstieg bis zum frühen Erwachsenenalter, auf den ein langsamer, kontinuierlicher Abfall folgt, der sich im höheren Alter in einigen Bereichen weiter beschleunigen kann.

Die Pragmatik des Denkens bezeichnet die prozeduralen und deklarativen Wissensbestände, die über Prozesse lebenslangen Lernens erworben werden. Die Ausbildung pragmatischer Kompetenzen ist dabei auch Ergebnis der früh im Lebenslauf erfolgten Investitionen mechanischer Fähigkeiten in ausgewählte Verhaltensbereiche dar (z.B. in Bildungsverläufe; in Training). Der Entwicklungsverlauf der intellektuellen Pragmatik zeigt meist einen kontinuierlichen Anstieg im Lebenslauf, der spät seinen Höhepunkt erreicht und dann oft erst im Alter geringfügig abnimmt. Nicht überraschend korrelieren auf pragmatischer Intellektualität aufbauende kognitive Leistungen (z.B. Wissen, Wortschatz, Weisheit) in der Regel deutlich stärker mit sozioökonomischen Ressourcen wie Bildung, Einkommen oder Berufsprestige, während die Entwicklung der elementaren kognitiven Mechanik im Lebenslauf stärker durch individuelle sensorische und psychomotorische Ressourcen bedingt wird (Baltes, Lindenberger & Staudinger, 1998; Lindenberger, 2002).

Auf der Grundlage dieser zentralen Unterscheidung wurde jeweils ein zentraler Indikator der Mechanik und der Pragmatik intellektueller Funktionstüchtigkeiten ausgewählt und für eine effiziente Erfassung im Rahmen des Computer-Assisted-Personal-Interviewing (CAPI) weiter entwickelt. Im Kontext der kognitiven Mechanik haben sich Tests der Wahrnehmungsgeschwindigkeit (speed) vielfältig bewährt, insbesondere im Rahmen von Studien mit altersheterogenen Stichproben (Lindenberger & Baltes, 1995), aber auch im Kontext der klinischen Diagnostik. Ein verbreiteter „speed“-Indikator ist der Zahlen-Zeichen-Test aus dem HAWIE-R (Tewes, 1994). Für den Bereich der intellektuellen Pragmatik stellen lexikalische Tests, wie die Tiere-Nennen-Aufgabe, einen effizienten und meist robusten Indikator des deklarativen Wissens dar.

3 Entwicklung der kognitiven Funktionstests für die vorliegende Studie

Für den Zweck einer Erfassung der kognitiven Funktionstüchtigkeit in einer nicht-standardisierten mündlichen Interviewerbefragung wurden die ausgewählten Einzeltests der Wahrnehmungsgeschwindigkeit (Zahlen-Zeichen-Test) und der Wortflüssigkeit (Tiere-Nennen-Aufgabe) abgewandelt und für eine computer-assistierte Erhebung tauglich gemacht.¹ Entscheidend war hierbei, dass die Tests ohne spezielle Schulung der Interviewer und zugleich mit möglichst geringen Fehlerquellen im Rahmen des Computer-Assisted-Personal-Interviewing (CAPI) einsetzbar sein müssen. In Anlehnung an den von Smith (1973 / 1995) entwickelten Symbol-Digit-Modalities-Tests (SDMT), wurde ein Test entwickelt, bei dem anhand einer auf dem Bildschirm permanent sichtbaren Zuordnungstafel innerhalb von 90 Sekunden das jeweils korrekte Zeichen einer möglichst großen Zahl von numerischen Zahlen zugeordnet werden mussten. Die maximale Anzahl korrekt zugeordneter Zeichen erlaubt eine Abschätzung der individuellen Wahrnehmungsgeschwindigkeit. Die Protokollierung der von den Probanden eingegebenen Zeichen erfolgte durch eine PC-Software im Hintergrund (vgl. Rosenblatt & Stocker, 2005).

Die Wortflüssigkeit („*fluency*“) wird mit der „Tiere-Nennen-Aufgabe“ (TA; Lindenberger & Baltes, 1995) erfasst. Hierbei sollen die Befragten innerhalb von 90 Sekunden möglichst viele unterschiedliche Tiere laut nennen. Die PC-Eingabe erfolgte durch den Interviewer, der für jede korrekte, wiederholte oder unklare Tier-Nennung jeweils eine entsprechende Eingabe (Tastendruck) am PC durchführte. Die Zählung erfolgte durch die PC-Software und wurde nach 90 Sekunden automatisch abgebrochen (vgl. Rosenblatt & Stocker, 2005). Die Gesamtdauer beider Tests inklusive der Instruktion beträgt im Durchschnitt ca. 5 Minuten.

¹ An dieser Entwicklung direkt beteiligt waren Frieder R. Lang, Bernhard von Rosenblatt, Jürgen Schupp, Andreas Stocker und Gert G. Wagner.

4 Darstellung des Vorgehens und Datenbasis

Die Datenerhebung für diese Studie erfolgte zu zwei Zeitpunkten im Abstand von 6 Wochen am Psychologischen Institut der Martin-Luther-Universität Halle-Wittenberg. Die Teilnehmer der Studie wurden über Aushänge im Bereich des Arbeitsamtes und in Supermärkten der Stadt Halle sowie durch direkte Anwerbung durch die Interviewer gewonnen. Angestrebt wurde eine heterogene Zusammensetzung der Stichprobe nach Erwerbsstatus und Bildung, wobei allerdings aufgrund der Länge und des Umfangs der Leistungstests mit einer positiven, hoch motivierten Teilnehmersauswahl zu rechnen war.

4.1 Stichprobenmerkmale

An der Studie nahmen 53 Männer und 66 Frauen im Alter von 18 bis 69 Jahren teil. Tabelle 1 gibt einen Überblick über die Verteilung sozialstruktureller und demografischer Personenmerkmale in der Stichprobe.

Das Durchschnittsalter der Stichprobe lag bei 31 Jahren, wobei nur ein Drittel der Befragten älter als 35 Jahre waren. Auf eine gesonderte Rekrutierung von älteren Erwachsenen (über 70 Jahre) wurde aus Kosten- und Zeitgründen verzichtet. Die Teilnehmer hatten einen vergleichsweise hohen Bildungsstand: Rund die Hälfte hatte Abitur und rund 15% hatten zusätzlich einen weiterführenden Universitäts- oder Hochschulabschluss absolviert. Mehr als ein Drittel der Teilnehmer (39.5%) waren bislang ohne Berufsabschluss, hierzu zählten auch die Studierenden. Dem relativ hohen Bildungsstand entspricht, dass die überwiegende Mehrheit der Teilnehmer bereits über mehrjährige PC-Erfahrung verfügte.

Tabelle 1:
Beschreibung der Stichprobe (Erste Erhebung; N = 119)

Merkmale	Männer (N = 53)	Frauen (N = 66)	Gesamt (N = 119)
Geschlecht (%)	44.5	55.5	
Alter in Jahren (SD)	31.8 (11.4)	30.4 (11.0)	31.0 (11.1)
davon: 18 – 25 Jahre (%)	35.8	42.4	36.1
26 – 35 Jahre (%)	33.3	27.3	33.6
36 – 45 Jahre (%)	19.2	19.7	19.3
46 – 69 Jahre (%)	10.8	10.6	10.9
Schulabschluss (in %)			
Ohne Abschluss	1.9	0.0	0.8
Hauptschule	7.5	1.5	4.2
Realschule	10.8	30.3	26.1
Fachhochschule	28.0	15.2	16.0
Abitur	52.8	51.5	52.1
Andere	0.0	1.5	0.8
Berufsausbildung (%^a)			
Ohne Abschluss (inkl. Studierende)	41.5	37.9	39.5
Lehre (gew., landw., kaufm., sonst.)	17.0	28.8	23.5
Fachschule (Handel-, Meister-)	15.1	15.2	15.1
Beamtenausbildung	3.8	1.5	2.5
Fachhochschule, Ingenieurschule	1.9	1.5	1.7
Universitätsabschluss	15.1	15.2	15.1
Sonst. Abschluss (z.B. Ausland)	5.7	-	2.5
<i>Bildung und Ausbildung in Jahren (SD)</i>	14.4 (2.3)	14.5 (2.6)	14.4 (2.5)
Erwerbstätigkeit (%)			
Nicht erwerbstätig (inkl. Studierende)	54.7	47.0	50.4
Geringf./Unregelm. beschäftigt	17.0	21.2	19.3
Ausbildung / Lehre / Umschulung	-	6.1	3.4
Teilzeitbeschäftigt	3.8	10.6	7.6
Voll erwerbstätig	24.5	15.2	19.3
<i>Erfahrung mit PCs seit X Jahren (SD)</i>	7.9 (5.7)	7.0 (4.9)	7.4 (5.3)
<i>PC-Nutzung (Tage im Monat, SD)</i>	17.1 (9.2)	13.2 (9.5)	14.9 (9.6)

Anm.: ^a = Höchster Ausbildungsabschluss (für 8.4% Mehrfachnennungen)

4.2 Erhebungsinstrumente der Validierungsstudie

Zur Validierung der CAPI-Tests wurden Tests der fluiden und kristallinen intellektuellen Leistung einbezogen: Tests zur Wahrnehmungsgeschwindigkeit, zum schlussfolgernden Denkens, Wissenstests (Wortschatz), Gedächtnistests sowie Tests der Aufmerksamkeitsleistung. Zur Überprüfung des BFI-S wurde zusätzlich ein Persönlichkeitsinventar einbezogen. Tabelle 2 gibt einen Überblick über die Testbatterie zur Validierung der CAPI-Tests.

Tabelle 2:
Übersicht der verwendeten Erhebungsinstrumente zur Validierung der CAPI-Tests

Konstruktbereich / Testverfahren	Beschreibung	Autor	Limit (sec.)	Re- test
<i>Geschwindigkeit</i>				
• Zahlen-Symbol-Test (ZST)	Zuordnung von Symbolen zu Zahlen (nach Vorlage)	Tewes, (HAWIE-R)	1994 90	X
• Zahlen-Buchstaben-Test (ZBT)	Zuordnung von Buchstaben zu Zahlen (nach Vorlage)	Lindenberg/Baltes, 1995 <i>BASE Cog</i>	90	X
<i>Schlussfolgerndes Denken</i>				
• Raven Matrizen Test (APM)	Auswahl des Teilstückes zu vorgegebenem Muster	Raven/ Kratzmeyer, 1980	420	
• Culture Fair Test 20 (CFT20)	(1) Ergänzung, (2) Streichung inkorrektur Figur, (3) Ergänzung eines Teils, (4) Räuml. Relation	Weiss, 1987	240, 180	
• Rechnerisches Denken (RD)	Textbasierte Rechenaufgaben	Tewes, 1994 (HAWIE-R)	15,30, 60,120	
<i>Wortschatz/Wissen</i>				
• Wortschatztest (WT)	Erkennen sinnhafter Worte	Lehrl, 1995 (MWT)	-	X
• Leistungsprüfsystem	Korrektur von Fehlern in optisch verfremdeten Worten	Horn, 1983 (LPS)	120, 180	X
• Allg. Verständnis (AV)	Fragen zu alltägl. Handlungen	Tewes, 1994 (HAWIE-R)	-	
<i>Gedächtnis</i>				
• LGT Deutsch-Türkisch	Paar-assoziations-Lernen mit „free recall“ am Ende	Bäumler, 1974	240	X
• Activity Recall (AE)	„Free Recall“ aller zuvor durchgeführten Aufgaben	Lindenberg/Baltes, 1995 <i>BASE Cog</i>	120	
<i>Aufmerksamkeit</i>				
• Test d2	Durchstreichen von Zeichen	Brickenkamp, 2002	je 20	
<i>Persönlichkeit</i>				
• NEO – FFI	Selbstbericht (5stufiges Rating)	Borkenau/Ostendorf 1993		X

4.2.1 Kurzversion des CAPI-Pretest Instruments

Zum Einsatz kam eine von TNS Infratest für die Studie in Halle auf 17 Fragen reduzierte Version des im SOEP-Pretest 2005 verwendeten CAPI Instruments. Erhoben wurden neben Geschlecht und Geburtsjahr, aktuelle Stimmung, Religionszugehörigkeit, Bildung, Noten des letzten Schulzeugnisses, Berufsausbildung, Erwerbsstatus, PC-Erfahrung und -Nutzung, persönliche Einstellungen (Vertrauen, Einsamkeit), Persönlichkeit (BFI-S), Kontrollmeinungen (LOC-I, LOC-E), die Lebenszufriedenheit sowie der Zeichentest (ZT, vgl. Punkt 3) und die Tiere-Nennen-Aufgabe (TA, vgl. Punkt 3).

4.2.2 Wahrnehmungsgeschwindigkeit (*speed*)

Der *Zahlen-Zeichen-Test (ZZT)* entstammt dem Handlungsteil des HAWIE-R (Tewes, 1994) und dient als Maß der Wahrnehmungsgeschwindigkeit. Hier wurde ein Zuordnungsschlüssel vorgegeben, nach dem der Proband innerhalb von 90 Sekunden Symbole zu Zahlen zuordnen soll. Als Leistungsmaß resultiert die Summe der richtig zugeordneten Zeichen.

Auch der *Zahlen-Buchstabentest (ZBT; Lindenberger & Baltes, 1995)* dient als Maß der Wahrnehmungsgeschwindigkeit. Hier wird ein Zuordnungsschlüssel vorgegeben, nach dem der Proband innerhalb von 90 Sekunden Buchstaben zu Zahlen zuordnen soll. Als Leistungsmaß resultierten die Summen der richtig zugeordneten Zeichen für drei 30-sekündige Intervalle.

4.2.3 Schlussfolgerndes Denken (*reasoning*)

Die *Advanced Progressive Matrices (APM; Raven & Kratzmeyer, 1980)* wurden zur Messung der schlussfolgernden Denkfähigkeit eingesetzt. Jedes Item besteht aus einem 9teiligen logischen Muster, in dem ein Teilstück fehlt, das aus 8 Figuren korrekt auszuwählen ist. Die Bearbeitungszeit für diesen Test wurde auf 7 Minuten begrenzt. Der Testwert ergab sich aus der Summe der korrekt gelösten Aufgaben.

Die Aufgabe *Rechnerisches Denken* gehört ebenfalls zum Handlungsteil des HAWIE-R (Tewes, 1994) und erfasst einen Aspekt schlussfolgernden Denkens. Hier wurden dem Probanden mündlich rechnerische Textaufgaben gestellt, die in einem vorgegebenen Zeitraum zu lösen waren. Die Zeitbegrenzung variiert in Abhängigkeit des Schwierigkeitsgrades jeder Aufgabe und bewegen sich zwischen 15 und 120 Sekunden. Für jede richtig gelöste Aufgabe wurde ein

Punkt vergeben, für besonders schnelles Lösen der schwierigeren Aufgaben je 1 Zusatzpunkt. Ermittelt wurde die Summe der Punkte.

Culture Fair Test. Die vier eingesetzten Subtests des CFT-20 (Weiß, 1987) dienen als Indikator des schlussfolgernden Denkens. Die Items der Aufgabe *Reihenfortsetzen* bestehen aus fünfteiligen Figurenreihen, deren letzte Position durch eine von je fünf vorgegebenen Figuren ergänzt werden soll. Im Subtest *Klassifikationen* wählt der Proband aus fünf Figuren diejenige aus, die nach einem zu ermittelnden Kriterium nicht in die Kategorie der vier anderen gehört. In der Aufgabe *Matrizen* werden Lücken in vier- bis neunteiligen logischen Mustern durch eine von je fünf vorgeschlagenen Figuren ergänzt. Im Subtest *topologische Schlussfolgerungen* wählten die Probanden aus fünf Abbildungen von Elementen die Abbildung aus, in der die räumliche Relation der jeweiligen Vorgabe entsprach. Für jeden der vier Subtests wurde die Summe korrekt ausgewählter Figuren bestimmt.

4.2.4 Wortschatz und Wissen (*knowledge*)

Der *Mehrfachwahl-Wortschatz-Test* (MWT; Lehl, 1995) erfasst den Wortschatz. Jedes Item besteht aus vier Nonsens-Worten und einem sinnvollen Wort, welches der Proband herauszufinden hat. Berechnet wurde die Summe der korrekt ausgewählten Wörter.

Aus dem *Leistungsprüfsystem* (LPS; Horn, 1983) wurden drei Subtests zur Erfassung des Wortschatzes eingesetzt. Aufgabe ist es, innerhalb von zwei bis zweieinhalb Minuten, den Druckfehler in je vierzig Substantiven von absteigender Gebräuchlichkeit herauszufinden. Berechnet wurde die Summe der korrekt identifizierten Druckfehler je Aufgabe.

Allgemeines Verständnis ist ein Subtest des Handlungsteil des HAWIE-R (Tewes, 1994) und erfasst Alltagswissen. Hier wurden den Probanden mündlich Fragen zu alltäglichen Problemstellungen und zum Sinn alltäglicher Handlungen gestellt. Die Antworten wurden nach vorgegebenen Richtlinien mit 0, 1 oder 2 Punkten bewertet. Der Testwert ergab sich aus der Summe der Punkte.

4.2.5 Gedächtnis und Erinnern (*memory*)

Die Aufgabe *Activity Recall* wurde in Anlehnung an eine Aufgabe aus der BASE-Testbatterie (Lindenberger & Baltes, 1995) übernommen und dient der Messung eines Aspekts der Gedächtnisleistung. Hier wurden die Probanden zu einem bestimmten Zeitpunkt im Untersu-

chungsablauf gefragt, an welche Aufgaben aus der bisherigen Untersuchung sie sich erinnern könnten. Als Leistungsmaß resultiert die Summe der erinnerten Aufgaben.

Die Aufgabe *Deutsch-Türkisch* aus dem Lern- und Gedächtnistest (LGT-3, Bäumlner, 1974) wurde zur Messung eines Aspekts der Gedächtnisleistung eingesetzt. Der Proband erhielt eine Minute Zeit, um zwanzig türkische Vokabeln zu lernen. Nach etwa zehn Minuten sollte der Proband innerhalb von vier Minuten die richtige türkische Vokabel aus je fünf türkischen Wörtern wieder erkennen. Als Leistungsmaß resultierte die Summe der korrekt ausgewählten Wörter.

4.2.6 Aufmerksamkeit

Der *Test d2* (Brickenkamp, 2002) dient der Messung der Aufmerksamkeits-Belastungsleistung. Er besteht aus 14 Zeilen mit den Buchstaben p und d, die jeweils mit einem bis vier Strichen versehen sind. Der Proband soll innerhalb von je 20 Sekunden aus jeder Zeile so viele Buchstaben d mit zwei Strichen wie möglich durchstreichen. Als Leistungsmaß resultiert neben verschiedenen Geschwindigkeits- und Fehlermaßen der so genannte Konzentrationsleistungswert. Diese Konzentrationsleistung ergibt sich aus der Summe aller durchgestrichenen d2-Zeichen abzüglich der versehentlich durchgestrichenen anderen Zeichen.

4.2.7 Persönlichkeitsinventar (NEO-FFI)

Das *Neo-Fünf-Faktoren Inventar* (NEO-FFI; Borkenau & Ostendorf, 1993) dient als Maß der Big Five Persönlichkeitsmerkmale Neurotizismus, Extraversion, Offenheit für Erfahrung, Verträglichkeit und Gewissenhaftigkeit. Im NEO-FFI soll der Proband sich im Hinblick auf Aussagen über seine Persönlichkeit auf einer Fünf-Punkte-Skala selbst einschätzen. Es resultiert ein Summenwert über alle Antworten für jede der fünf Skalen.

4.3 Ablauf der Erhebung in Halle (Saale)

Die Teilnehmer nahmen im Abstand von 6 Wochen an zwei Einzelsitzungen von jeweils 100 und 60 Minuten teil. Die erste Sitzung umfasste die dargestellte CAPI-Kurzversion (ca. 15 Minuten) sowie die kognitive Testbatterie und den NEO-FFI. Um mögliche Aktivierungs- und Trainingsgewinne der kognitiven Testbatterie im Hinblick auf die Anordnung der Tests während der Erhebung identifizieren zu können, wurde die Sitzung bei der Hälfte der Teilnehmer mit dem CAPI-(54%) und bei der anderen Hälfte (46%) mit der kognitiven Testbatte-

rie begonnen. Die Retest - Erhebung nach 6 Wochen wurde stark verkürzt und auf eine Wiederholung des CAPI und der konvergent validierenden kognitiven Testverfahren (Speed, Wortschatz, Persönlichkeit) begrenzt (vgl. Tabelle 2).

5 Testgüte der Tiere-Nennen-Aufgabe und des Zeichen-Tests

Für die Testgüte beider kognitiven Funktionstest ist entscheidend, in welchem Umfang es aufgrund der vergleichsweise wenig kontrollierten Erhebungsbedingungen im CAPI zu möglichen Verzerrungen und Informationsverlusten kommt. Neben Aspekten der Durchführungsobjektivität waren daher auch insbesondere mögliche Interviewereffekte sowie die Effekte der Reihenfolge der Testverfahren von Bedeutung.

Für die Bewertung der Reliabilität eines psychologischen Testverfahrens werde unterschiedliche Kriterien heran gezogen. Von Bedeutung sind vor allem die Homogenität, Trennschärfe und innere Konsistenz der Items sowie die Retest-Stabilität einer Skala (z.B. Bortz & Döring, 1995; Lienert & Raatz, 1998). Eine Abschätzung der *Homogenität* einer Skala ermöglicht die „Mean-Inter-Item-Correlation (MIC)“, die sich aus dem Durchschnitt aller Korrelationskoeffizienten eines jeweiligen Items mit allen anderen Items ergibt. Die *Trennschärfe* stellt die Korrelation eines Items mit dem gültigen Skalengesamtwert dar und zeigt an, wie gut der Gesamtwert durch jedes einzelne Item vorher gesagt wird. Die innere Konsistenz einer Skala wird üblicherweise durch Cronbach's Alpha indiziert und spiegelt den inneren Zusammenhang zwischen den Items einer Skala wider. Der Alpha-Wert kann sowohl auf der Basis von Rohwerten als auch auf der Basis der standardisierten Itemwerte (standardisiertes Alpha) ermittelt werden. Ein zusätzlichen Hinweis auf die Reliabilität eines Verfahrens erlaubt die Ermittlung der Retest-Stabilität: Der Test-Retest-Koeffizient (bzw. Stabilitätskoeffizient) zeigt an, wie gut die Testergebnisse einer Testwiederholung innerhalb eines kurzen Zeitabstands durch die Werte der ersten Testdurchführung vorhergesagt werden können. Ein Wert von +1 indiziert perfekte Stabilität, bei einem Wert von „0“ besteht keine Stabilität.

5.1 Kennwerte, Objektivität und Reliabilität der Tiere-Nennen-Aufgabe (TA)

Ein besonderes Merkmal der Tiere-Nennen-Aufgabe besteht darin, dass die von den Befragten genannten Tiere jeweils nur auf der Basis der direkten Eingabebestätigung des Interviewers (Tastendruck in vorgegebener Maske) im Softwareprogramm gezählt wurden. Dieses Vorgehen bedroht die Objektivität der Testdurchführung und zugleich die Reliabilität der Aufgabe. Aus diesem Grund wurden alle Tiernennungen der Probanden während dieser Auf-

gabe zusätzlich per Tonband aufgenommen, um somit eine Objektivitätskontrolle durchführen zu können.

Tabelle 3 gibt einen Überblick über Kennwerteverteilung, Homogenität, Trennschärfe und Test-Retest-Koeffizienten der Tiere-Nennen-Aufgabe. Aus Tabelle 3 geht hervor, dass in der ersten Erhebung und im Retest eine vernachlässigbare Abweichung zwischen dem Summenwert der über PC eingegebenen und der per Tonband objektivierten Tier-Nennungen bestand. So wurden im ersten Test durchschnittlich nur 0.4 (SD = 4.8) mehr Tiere im CAPI eingegeben als tatsächlich genannt wurden. Im Retest lag die Überschätzung bei durchschnittlich 1.8 Tieren (SD = 3.8). Ein weiterer Hinweis auf Durchführungsobjektivität ist, dass die Anordnung der CAPI-Tests zu Beginn oder am Ende der Testsitzung sich nicht auf die durchschnittliche Zahl genannter Tiere auswirkte, weder in der ersten Erhebung ($F_{(1;117)} = 0.9, p > .25; N = 119$) noch im Retest ($F_{(1;115)} = 0.6, p > .25; N = 116$). Dies belegt, dass die TA robust gegen Situations- und Übungseinflüsse ist.

Wie Tabelle 3 zu entnehmen ist, liegt die Test-Retest-Korrelation für die Tiere-Nennen-Aufgabe mit .46 im unteren Toleranzbereich. Berücksichtigt man allerdings, dass die Stabilität des Tests aufgrund der Tonbandaufnahme einen Wert von $r = .60$ erreicht, so ergibt sich für die um diesen Minderungsfaktor korrigierte Retest-Stabilität des CAPI-Tests ein Wert von .59. Der Minderungsfaktor korrigiert den Retest-Koeffizienten um das Ausmaß nachgewiesener Unreliabilität in einem oder beiden beteiligten Skalenwerten (Lienert & Raatz, 1998). Wird die Unreliabilität beider beteiligten Skalenwerte (auf der Basis der Tonbandaufnahmen) doppelt minderungskorrigiert, ergibt sich sogar ein Retest-Koeffizient von .70 für die Tiere-Nennen-Aufgabe. Die geringste Retest-Reliabilität zeigt sich für den Zeitblock der ersten 30 Testsekunden. Eine mögliche Erklärung ist, dass in aller Regel fast die Hälfte aller Nennungen in diesem ersten Zeitintervall erfolgt, was besondere Anforderungen an den Interviewer stellt. Allerdings deuten die Verteilungsmerkmale der Items wie die Trennschärfen in der ersten Erhebung und im Retest darauf hin, dass sich die Testzuverlässigkeit und -präzision zwischen den ersten und letzten 30 Sekunden nicht substantiell unterscheiden.

Tabelle 3:
Verteilungskennwerte, Trennschärfe und Reliabilität der Tiere-Nennen Aufgabe

MZP	Intervall (sec)	Wert	M (SD)	Schiefe ^a	MIC ^b	Trennschärfe ^c	Test-Retest-Koeff.
Erste Erhebung (N = 119)	0-30	Korrekt	17.7 (5.0)	.9	.41	.82	
		<i>Wiederholt</i>	<i>0.1 (0.3)</i>	<i>3.3</i>		<i>.02</i>	
		<i>Undeutlich</i>	<i>0.0 (0.2)</i>	<i>8.7</i>		<i>.02</i>	
	30-60	Korrekt	9.4 (4.0)	.5	.45	.80	
		<i>Wiederholt</i>	<i>0.1 (0.4)</i>	<i>2.7</i>		<i>.06</i>	
		<i>Undeutlich</i>	<i>0.1 (0.2)</i>	<i>3.8</i>		<i>.02</i>	
	60-90	Korrekt	7.2 (3.5)	.3	.39	.71	
		<i>Wiederholt</i>	<i>0.1 (0.3)</i>	<i>3.6</i>		<i>.19</i>	
		<i>Undeutlich</i>	<i>0.1 (0.3)</i>	<i>4.3</i>		<i>-.03</i>	
	Gesamt (CAPI)	Korrekt	34.3 (9.8)	.5	.42	1.0	
		<i>Wiederholt</i>	<i>0.3 (0.6)</i>	<i>2.0</i>		<i>.15</i>	
		<i>Undeutlich</i>	<i>0.2 (0.5)</i>	<i>3.7</i>		<i>-.00</i>	
Tonband (N=112)	Korrekt	34.7 (8.6)	.2		.89		
	<i>Differenz zu CAPI</i>	<i>-0.4 (4.3)</i>	<i>-.3</i>		<i>.36</i>		
	<i>Wiederholt</i>	<i>0.4 (0.8)</i>	<i>2.5</i>		<i>.17</i>		
	<i>Undeutlich</i>	<i>0.6 (1.1)</i>	<i>2.9</i>		<i>.14</i>		
	Retest (N = 116)	Korrekt	17.0 (5.5)	.1	.36	.82	.27
		<i>Wiederholt</i>	<i>0.1 (0.3)</i>	<i>5.5</i>		<i>-.00</i>	<i>.19</i>
<i>Undeutlich</i>		<i>0.1 (0.4)</i>	<i>4.5</i>		<i>.06</i>	<i>-.03</i>	
30-60	Korrekt	9.6 (4.4)	.2	.42	.80	.42	
	<i>Wiederholt</i>	<i>0.1 (0.3)</i>	<i>4.2</i>		<i>.06</i>	<i>-.02</i>	
	<i>Undeutlich</i>	<i>0.1 (0.4)</i>	<i>7.2</i>		<i>.07</i>	<i>.29</i>	
60-90	Korrekt	6.7 (3.4)	.1	.33	.64	.32	
	<i>Wiederholt</i>	<i>0.1 (0.4)</i>	<i>2.8</i>		<i>.15</i>	<i>-.03</i>	
	<i>Undeutlich</i>	<i>0.0 (0.3)</i>	<i>6.6</i>		<i>.02</i>	<i>.15</i>	
Gesamt (CAPI)	Korrekt	33.3 (10.2)	-.1	.37	1.0	.46	
	<i>Wiederholt</i>	<i>0.3 (0.6)</i>	<i>2.3</i>		<i>.13</i>	<i>.09</i>	
	<i>Undeutlich</i>	<i>0.2 (0.8)</i>	<i>4.9</i>		<i>.08</i>	<i>.35</i>	
Tonband (N=112)	Korrekt	35.3 (10.0)	.1		.92	.60	
	<i>Differenz zu CAPI</i>	<i>-1.8 (3.9)</i>	<i>-3.5</i>		<i>.15</i>	<i>-.28</i>	
	<i>Wiederholt</i>	<i>0.6 (1.0)</i>	<i>2.2</i>		<i>.29</i>	<i>.22</i>	
	<i>Undeutlich</i>	<i>0.3 (0.6)</i>	<i>2.4</i>		<i>.05</i>	<i>.19</i>	

Anm.: * p<.05, *** p<.001, ^a Standardfehler der Schiefe = .2; ^b Mittlere Inter-Item Korrelation; ^c Korrelation mit korrektem Gesamtwert (p < .05 fuer Koeff. >.18 bzw. <-.18).

Der Tabelle 3 kann entnommen werden, dass „wiederholte“ oder „unklare“ Tier-Nennungen mit den korrekten Testwerten praktisch unkorreliert sind, oder die Korrelationen der entsprechenden Werte der Tonbandaufnahme (r = .17 bzw. .29) widerspiegeln. Allerdings zeigen die unerwartet positiven Korrelationen mit dem Endwert, dass es sich nicht um „echte“ Fehlerwerte handelt (diese müssten negativ korrelieren, vgl. Zeichentest). Wer mehr Tiere wiederholt, sollte am Ende eine geringere Zahl korrekter Nennungen zeigen. Tatsächlich scheinen

aber Wiederholungen von Tieren tendenziell mit besseren Gesamtleistungen einherzugehen. Sich daraus ergebende Verzerrungen stellen keine systematische Beeinträchtigung des „korrekten“ Endwerts dar, weisen aber doch auf eine Anfälligkeit der Aufgabe für Interviewerfehler hin. Kennwerte, Trennschärfe und Retest-Reliabilität belegen eine insgesamt ausreichende Durchführungsobjektivität und Robustheit des Tests.

Tabelle 4 gibt einen Überblick über die internen Konsistenzen (Alpha) für die drei Zeitblöcke der Tiere-Nennen-Aufgabe und des Zeichentests in der ersten Erhebung und im Retest. Die Alpha-Werte der TA liegen im akzeptablen Bereich, wenn die durch die Retest-Reliabilität der objektivierte Tonbandaufnahme markierte obere Grenze der Reliabilität berücksichtigt wird. Somit sind die dargestellten Alpha-Werte der standardisierten Testdrittel von .68 bzw. .64 durchaus im Toleranzbereich. Gleichwohl belegen auch diese Werte die etwas erhöhte Störanfälligkeit der Tiere-Nennen-Aufgabe für Situationseffekte.

Tabelle 4:
Interne Konsistenz der zwei kognitiven Funktionstests

Aufgabe (CAPI)	Erhebung	N	Alpha	Standard. Alpha
Tiere Nennen	1	119	.67	.68
	Retest	116	.62	.64
Zeichentest	1	119	.87	.87
	Retest	116	.95	.95

Anm.: Interne Konsistenz der drei 30-Sekunden-Intervalle je Test.

5.2 Kennwerte und Reliabilität des Zeichen-Tests

Die Tabelle 5 fasst alle Testkennwerte und Retest-Koeffizienten des CAPI-Zeichentests zusammen. Insgesamt belegen die zentralen Kennwerte eine hohe Präzision und Robustheit des Zeichentests. Die Leistungen aller drei Zeitblöcke des Zeichentests tragen homogen und in gleicher Stärke zum korrekten Gesamtwert bei. Erneut zeigen sich keine Effekte der Testanordnung zu Beginn oder Ende der Sitzung (Erste Erhebung: $F_{(1;117)} = 1.7$, $p = .20$; $N = 119$; Retest: $F_{(1;115)} = 0.4$, $p > .25$; $N = 116$).

Fast alle Fehlerwerte der Teilnehmer (bis auf eine Ausnahme) korrelieren im Sinne der Erwartung im marginalen Bereich negativ mit dem korrekten Endwert. Lediglich für den letzten Zeitblock zeigt sich, dass die Verlangsamung der Antworten nach 60 Sekunden dazu führt, dass fehlerhafte Eingaben keinerlei Kosten mehr haben (und sogar insgesamt leicht positiv mit dem Endwert korrelieren). Auch die interne Konsistenz der drei Zeitblöcke des ZT erweist sich mit Alphawerten zwischen .87 und .95 (Retest) als ausgesprochen gut. Der Test-Retest-Koeffizient von $r = .55$ (n. Minderungskorrektur: $r = .60$) liegt im befriedigenden Bereich eines so kurzen Einzeltests. Allerdings bleibt der Retest-Koeffizient deutlich unter der Stabilität des Zahlen-Zeichen-Tests ($r = .92$) und des Zahlen-Buchstaben-Tests ($r = .91$).

Tabelle 5:

Verteilungskennwerte, Trennschärfe und Retest-Reliabilität des Zeichentests

MZP	Zeitintervall	Wert	M (SD)	Schiefef ^a	MIC ^b	Trennschärfef ^c	Test-Retest-Koeff.
Erste Erhebung (N = 119)	0-30 sec	Korrekt	13.0 (3.2)	-0.1	.70	.91	
		<i>Fehler</i>	<i>0.2 (0.5)</i>	<i>3.1</i>		<i>-.20</i>	
	30-60 sec	Korrekt	13.5 (2.5)	-1.1	.69	.87	
		<i>Fehler</i>	<i>0.2 (0.6)</i>	<i>3.5</i>		<i>-.04</i>	
	60-90 sec	Korrekt	12.6 (2.9)	0.4	.70	.90	
		<i>Fehler</i>	<i>0.3 (0.6)</i>	<i>1.8</i>		<i>.03</i>	
Gesamt	Korrekt	39.0 (7.6)	-0.2	.70	1.0		
	<i>Fehler</i>	<i>0.7 (1.2)</i>	<i>3.7</i>		<i>-.09</i>		
Retest (N = 116)	0-30 sec	Korrekt	12.7 (4.5)	-0.2	.88	.97	.56
		<i>Fehler</i>	<i>0.1 (0.3)</i>	<i>3.3</i>		<i>-.10</i>	<i>.36</i>
	30-60 sec	Korrekt	12.6 (3.8)	-0.5	.87	.96	.48
		<i>Fehler</i>	<i>0.1 (0.3)</i>	<i>2.6</i>		<i>-.04</i>	<i>.14</i>
	60-90 sec	Korrekt	12.3 (3.8)	-0.3	.86	.95	.42
		<i>Fehler</i>	<i>0.2 (0.5)</i>	<i>2.7</i>		<i>.23</i>	<i>.26</i>
	Gesamt	Korrekt	37.7 (11.6)	-0.3	.87	1.0	.55
		<i>Fehler</i>	<i>0.4 (0.8)</i>	<i>2.5</i>		<i>.10</i>	<i>.32</i>

Anm.: ^a Standardfehler der Schiefe = .2; ^b Mittlere Inter-Item Korrelation; ^c Korrelation mit korrektem Gesamtwert (p < .05 fuer Koeff. >.18 bzw. <-.18).

In der Zusammenschau der Tabellen 4 und 5 stellt sich der ZT insgesamt als reliabel, robust und durchführungsobjektiv dar. Auch in wenig kontrollierten Erhebungssituationen erlaubt der CAPI-ZT eine verlässliche Schätzung des Leistungsparameters.

5.3 Validierung der Tiere-Nennen-Aufgabe und des Zeichentests

Für die Validierung der beiden CAPI-Tests wurde eine kognitive Testbatterie zusammengestellt, in der die entsprechenden Konstruktbereiche mit etablierten Testindikatoren erfasst wurden. Zur konvergenten Validierung des CAPI-Zeichentests im Bereich der fluiden (mechanischen) Fähigkeiten wurden Indikatoren der Wahrnehmungsgeschwindigkeit und des schlussfolgernden Denken einbezogen. Zur konvergenten Validierung der Tiere-Nennen-Aufgabe wurden im Bereich der kristallinen (pragmatischen) Leistungen drei Tests zu Wissen, Wortschatz und allgemeinem Verständnis erfasst. Zusätzlich wurden zwei Gedächtnisparameter und ein Test zur Erfassung der generellen Aufmerksamkeitsleistung erfasst, für die keine differenzierten Zusammenhänge mit den CAPI-Tests erwartet wurden.

Tabelle 6:
Konstruktvalidität der kognitiven CAPI-Tests

Konstruktbereich/ Testname	Erste Erhebung (N = 119)		Retest (N = 116)		CAPI
	Zeichen Test	Tiere Nennen	Zeichen Test	Tiere Nennen	
Geschwindigkeit					
ZZT	.70	.32	.54	.34	
ZBT	.68	.30	.56	.29	
Schlussfolgerndes Denken					
APM	.54	.25			
CFT20	.62	.32			
RD	.29	.28			
Wortschatz/Wissen					
WT	.15	.33	.09	.39	
LPS	.33	.36	.23	.30	
AV	.15	.26			
Gedächtnis					
AE	.39	.31			
LGT*	.30	.10	.15	.16	
Aufmerksamkeitsleistung					
D2	.57	.39			

Anm.: Koeffizienten > .18 sind signifikant ($p < .05$); ZZT = Zahlen-Zeichen-Test, ZBT – Zahlen-Buchstaben-Test, APM – Advanced Progressive Matrices, CFT20 – Culture Fair Intelligence Test 20, WT – Wortschatztest, AV – Allgemeines Verständnis, AE – Aufgaben Erinnern Aufgabe, *LGT – Deutsch-Türkische-Lernaufgabe (*Recall im Retest nach 6 Wochen), RD – Rechnerisches Denken, D2 – Korrekte Lösungen im D2-Test, LPS – Lernprüfungssystem.

Tabelle 6 gibt einen Überblick der Konstruktvalidierung nach der ersten Erhebung und im Retest. Wie die Tabelle 6 zeigt, sind alle Zusammenhänge der CAPI-Tests mit den jeweils konvergenten Testindikatoren im Sinne der Erwartung. Der Zeichentest zeigt dabei eine insgesamt deutlich höhere Konstruktvalidität, während die Korrelationen mit der Tiere-Nennen-Aufgabe im moderaten Bereich sind. Jedoch ist erkennbar, dass die konvergenten Zusammenhänge der TA mit dem LPS und dem WT am höchsten ausfielen und sich im Retest nochmals deutlich verbesserten. Bei beiden Tests bestand ein moderater Zusammenhang mit der generellen Aufmerksamkeitsleistung (d2).

Die Korrelationen der beiden CAPI-Leistungstests lag in der ersten Erhebung mit $r = .18$ ($n = 119$, $p < .05$) unter der Erwartung, im Retest lag der Zusammenhang in der erwarteten moderaten Höhe von $r = .48$ ($n = 116$, $p < .01$). Die geringe Korrelation in der ersten Erhebung kann nicht auf mögliche Reihenfolgeeffekte der Testanordnung (CAPI zuerst vs. zuletzt) zurückgeführt werden. Eine mögliche Erklärung ist, dass zahlreiche Teilnehmer mit der Test-situation generell unvertraut waren. Wie die Tabellen 3, 4 und 5 belegen, wurden im Retest

insgesamt weniger Fehler in den kognitiven Tests gemacht, wobei sich zugleich eine verbesserte Kohäsion der Testergebnisse (z.B. höhere Trennschärfen) zeigt. Allerdings wurden die bisherigen Vorerfahrungen der Teilnehmer mit Leistungstests und Testsituationen in der vorliegenden Studie nicht erfasst. Die geringe Korrelation der Langzeit-Gedächtnisaufgabe (LGT*) im Retest mit den CAPI-Testwerten weist darauf hin, dass mögliche Erinnerungseffekte bei der Durchführung des Retest eine insgesamt untergeordnete Rolle gespielt haben dürften.

Multitrait-multimethod-Matrix (MTMM)

Die Multitrait-multimethod-Methode überprüft, wie gut ein identisches Konstrukt durch verschiedene Testverfahren erfasst wird (konvergente Validität) und wie stark sich verschiedene Konstrukte dabei mit einer Methode differenzieren lassen (divergente Validität). Im minimalen Fall beruht das MTMM-Verfahren darauf, dass zwei unterschiedliche Merkmale (z.B., speed; word fluency) mit jeweils zwei unterschiedlichen Methoden (z.B. CAPI, „paper-pencil“) erfasst werden (Campbell & Fiske, 1959, Bortz & Döring, 1995).

Die sich in dieser Studie ergebende Matrix von kognitiven Testparametern ermöglicht einen Vergleich zwischen den beiden CAPI-Leistungstests mit herkömmlichen Testverfahren, die hier als „Anker“ der Validierung dienen. Für den Zeichentest (ZT) wurden zwei etablierte Tests der Wahrnehmungsgeschwindigkeit (ZZT, ZBT) und für die Tiere-Nennen-Aufgabe (TA) zwei etablierte Tests des Wortschatzes (WT, LPS) heran gezogen. Tabelle 7 gibt einen Überblick über resultierende MTMM-Matrix, in der aus Gründen der besseren Übersichtlichkeit auch einige der in Tabelle 6 bereits dargestellten Koeffizienten nochmals wieder gegeben werden. Wie erwartet, liegen die in der Diagonale dargestellten Retest-Reliabilitäten der standardisierten Testverfahren deutlich höher als die der CAPI-Tests. Alle konvergenten Validitätskoeffizienten (grau unterlegt) sind im moderaten Bereich und fallen für den Zeichentest höher aus als die (*kursiv* dargestellt) divergenten *heterotrait-heteromethod*-Koeffizienten. Für die Tiere-Nennen-Aufgabe besteht nur im Retest eine akzeptable diskriminante Validität. Die mittleren *heterotrait-monomethod* Korrelationen ($r = .36$) der beiden CAPI-Tests ist niedriger als der mittlere Koeffizient der (grau unterlegten) *monotrait-heteromethod* Matrix ($r = .52$). Beide Tests erfassen somit unterschiedliche Konstruktbereiche der intellektuellen Fähigkeit.

Tabelle 7:
Multitrait-multimethod-Matrix

	CAPI				Validierung (Erste Erhebung)			
	Zeichentest (ZT)		Tiere-Nennen (TA)		ZZT	ZBT	WT	LPS
	1	Retest	1	Retest				
CAPI								
ZT Retest	.551							
TA	.18	.07						
TA Retest	.23	.48	.456					
Validierung								
ZZT	.71	.54	.33	.34	.915			
ZBT	.65	.49	.26	.30	.83	.847		
WT	.19	.09	.32	.39	.23	.16	.691	
LPS	.32	.16	.38	.24	.31	.29	.66	.918

Anm.: Die grau unterlegte Fläche markiert die monotrait-heteromethod-Matrix (konvergente Validitätskoeffizienten). Kursiv dargestellte Koeffizienten markieren die heterotrait-heterotrait Matrix (diskriminante Validität). Koeffizienten > .18 bzw. < -.18 sind signifikant ($p < .05$).

Von zentraler Bedeutung für die Validität der beiden kognitiven CAPI-Tests sind schließlich auch die Außenkorrelationen zu anderen Variablen, mit denen sich zugleich einige substanzwissenschaftliche, inhaltliche Fragestellungen verbinden. Tabelle 8 gibt einen Überblick über ausgewählte externe Validierungsfaktoren. Wie erwartet besteht eine deutliche hohe negative Korrelation des Zeichentests mit dem chronologischen Alter ($r = -.56$ bzw. $-.36$ im Retest). Dieser Zusammenhang steht im Einklang mit dem erwarteten Altersverlauf von Indikatoren der intellektuellen Mechanik. Die Korrelationen mit der Dauer von Bildung und Ausbildung in Jahren ist hingegen völlig unkorreliert mit dem Zeichentest ($r = -.05$ bzw. $.08$). Die Korrelation der Tiere-Nennen-Aufgabe zeigt umgekehrt den erwarteten positiven Zusammenhang mit der Bildung ($.42$ bzw. $.45$), ist jedoch nicht korreliert mit Alter. Diese Befunde validieren den Anspruch beider Tests an eine differenzierte Erfassung eines kulturfreien und eines stärker bildungsabhängigen kognitiven Leistungsindikators.

Aus Tabelle 8 geht auch hervor, dass die kognitiven Leistungsparameter weit gehend unabhängig sind von den Persönlichkeitsmerkmalen, was dafür spricht, dass Einstellungen und konsistente Verhaltenstendenzen (z.B. Gewissenhaftigkeit), die Testleistungen nur geringfügig beeinflussten. Für die erste Erhebung zeigt sich eine erkennbare Korrelation der PC-Vorerfahrung mit der Leistung im Zeichentest. Diese wurde aber bereits bis zum Retest nivelliert. Offensichtlich genügt bereits eine geringe Vertrautheit mit dem CAPI-Test, um fehlende PC-Vorerfahrung im Zeichentest zu kompensieren.

Tabelle 8:

Außenkorrelationen der kognitiven Funktionstests des CAPI in der Erst- und Retesterhebung (synchrone Korrelationen)

Merkmal	Ersterhebung		Retest	
	Zeichen-test	Tiere Nennen	Zeichen-test	Tiere Nennen
Alter	-.56	-.03	-.36	.05
Schul-/Berufsausbildung in Jahren	-.05	.42	.08	.45
Schulnote (Durchschnitt: D, M, Fsp)	-.09	-.11	-.11	-.25
PC-Vorerfahrung (Dauer in J.)	.34	.09	.12	.13
E-Extraversion	.23	.12	.13	.13
A-Verträglichkeit	.07	-.02	-.03	.05
C-Gewissenhaftigkeit	-.17	-.16	-.01	.08
N-Emotionale Labilität	-.10	-.18	-.12	-.20
O-Offenheit f. Erfahrung	.06	.10	.05	.21
Kontrollerleben (LOC-I, 3 Items)	-.06	-.22	-.14	-.14
Kontrollaufgabe (LOC-E, 5 Items)	-.29	-.30	-.08	-.16
Zukunftszuversicht (1 Item)	-.20	-.16	-.14	-.14
Momentane Stimmung („ <i>first</i> item“)	.03	.07	.05	-.06
Lebenszufriedenheit („ <i>last</i> item“)	.34	.18	.25	.15

Anm.: Synchrone Korrelationen des jew. Testzeitpunkts ($p < .05$ für Koeff. $>.18$ bzw. $<-.18$).

6 Testgüte des BFI-S

Im Rahmen des SOEP Pretests 2005 wurde eine Kurzversion des Big Five Inventory nach Oliver John entwickelt und validiert (Gerlitz & Schupp, 2005; John & Srivastava, 1999; Lang, Lüdtke & Asendorpf, 2001). Im Rahmen der vorliegenden Studie wurde die CAPI-Version des BFI-S erneut überprüft. Zielsetzung war es, auf der Grundlage einer weiteren unabhängigen Datenbasis, eine weitere Konstruktvalidierung mit dem NEO-FFI (Borkenau & Ostendorf, 1993) durchzuführen, sowie die Retest-Reliabilität des BFI-S zu überprüfen.

Tabelle 9:
Faktorladungen (Varimax PCA) der BFI-S15 Items je Erhebungszeitpunkt

Erste Erhebung (N = 119)	E	C	N	A	O	h^2
Ich bin jemand, der ...						
- aus sich herausgehen kann, gesellig ist.	.848	-.086	-.011	-.045	.086	.74
- kommunikativ, gesprächig ist.	.788	.107	-.053	.154	.196	.70
- zurückhaltend ist.	-.694	-.058	.199	.049	-.049	.53
- gründlich arbeitet.	-.020	.846	-.074	.016	-.026	.72
- Aufgaben wirksam und effizient erledigt.	.218	.809	-.046	-.002	.108	.72
- eher faul ist.	.099	-.757	-.128	-.143	-.011	.62
- leicht nervös wird.	-.186	-.069	.804	.013	-.138	.71
- sich oft Sorgen macht.	.128	.206	.779	.067	-.037	.67
- entspannt ist, mit Stress gut umgehen kann.	.290	.141	-.694	.187	-.035	.62
- rücksichtsvoll und freundlich mit anderen umgeht.	.180	.241	.051	.767	.068	.69
- verzeihen kann.	.206	-.024	.213	.438	-.182	.31
- manchmal etwas grob zu anderen ist.	.095	-.091	.153	-.780	.027	.65
- eine lebhaft Phantasie, Vorstellung hat.	.157	-.111	-.027	.220	.849	.81
- originell ist, neue Ideen einbringt.	.176	.234	-.108	-.182	.802	.77
- künstlerische Erfahrungen schätzt.	-.260	-.173	-.128	.579	.261	.52
Varianzaufklärung (%)	14.8	14.6	12.7	12.6	10.4	65.1
Retest (N = 116)						
Ich bin jemand, der ...						
- aus sich herausgehen kann, gesellig ist.	.860	.060	-.162	.161	.161	.82
- kommunikativ, gesprächig ist.	.781	.086	-.035	.273	.248	.75
- zurückhaltend ist.	-.779	-.033	.217	.247	.056	.72
- gründlich arbeitet.	.047	.795	.014	.143	.156	.68
- Aufgaben wirksam und effizient erledigt.	.170	.724	-.344	.193	.099	.72
- eher faul ist.	.008	-.738	-.090	-.107	.120	.58
- leicht nervös wird.	-.191	-.081	.842	.029	-.072	.76
- sich oft Sorgen macht.	-.034	.291	.726	-.057	-.105	.63
- entspannt ist, mit Stress gut umgehen kann.	.140	.291	-.793	.059	.039	.74
- rücksichtsvoll und freundlich mit anderen umgeht.	.081	.278	-.131	.810	.146	.78
- verzeihen kann.	.251	.107	.131	.715	-.060	.61
- manchmal etwas grob zu anderen ist.	.126	-.086	.094	-.786	-.124	.67
- eine lebhaft Phantasie, Vorstellung hat.	.219	-.015	-.080	.051	.840	.76
- originell ist, neue Ideen einbringt.	.266	.108	-.100	-.084	.813	.76
- künstlerische Erfahrungen schätzt.	-.166	.004	-.035	.262	.563	.42
Varianzaufklärung (%)	15.2	14.2	14.0	13.3	12.5	69.2

Anm.: Koeffizienten > .12 sind signifikant ($p < .05$; > .25: $p < .01$), grau unterlegt = konvergente Ladungen,

Zunächst wurde die Faktorenstruktur des BFI-S mittels einer Hauptkomponentenanalyse mit Varimax-Rotation getrennt für jeden Erhebungszeitpunkt berechnet. Tabelle 9 gibt die Faktorladungsmatrix für die erste Erhebung und den Retest wieder. Es zeigt sich, dass mit einer Ausnahme alle Faktorladungen im Sinne der Erwartung ausfallen. Für das Item „(Ich bin jemand der) künstlerische Erfahrung schätzt“ zeigte sich in der ersten Erhebung eine höchste Ladung auf dem Verträglichkeitsfaktor. Die Faktorladungsmatrix des Retests zeigt, dass das Item in der wiederholten Abfrage im Sinne der Erwartungen lädt, und dort nicht mehr auffällig abwich. Im SOEP-Pretest 2005 wurde keine auffällige Abweichung für dieses Item beobachtet (Gerlitz & Schupp, 2005; N = 1012). Angesichts der generellen Inkonsistenz der Offenheitsitems sollte gleichwohl in zukünftigen Anwendungen, die ursprüngliche Formulierung des Items „(Ich bin jemand der) künstlerische, *ästhetische* Erfahrungen schätzt“ verwendet werden.

Tabelle 10:
Innere Konsistenz der fünf Persönlichkeitsskalen des BFI-S

BFI-S / Konstrukt	MIC	Alpha	Alpha (standard.)
<i>Erste Erhebung (N = 119)</i>			
E - Extraversion	.50	.74	.75
A - Verträglichkeit	.28	.52	.54
C - Gewissenhaftigkeit	.51	.72	.76
N - Neurotizismus	.42	.68	.68
O - Offenheit für Erfahrung	.24	.44	.48
<i>Retest (N = 116)</i>			
E - Extraversion	.57	.79	.80
A - Verträglichkeit	.47	.71	.73
C - Gewissenhaftigkeit	.43	.66	.70
N - Neurotizismus	.48	.73	.74
O - Offenheit für Erfahrung	.38	.62	.65

Insgesamt stellt sich die Faktorenstruktur zu beiden Erhebungszeitpunkten invariant dar, wobei zwei Items (Ich bin jemand der...„*verzeihen kann*“, „*künstlerische Erfahrungen schätzt*“) im Kontext der Verträglichkeit und Offenheit nicht ganz erwartungsgemäß laden. Die innere Konsistenz der Konstrukte stellt sich mit der Ausnahme der Offenheit robust und homogen dar (vgl. Tabelle 10). Wie meist im Kontext des lexikalischen Big Five Ansatzes zu beobachten ist, besteht eine relativ geringe innere Konsistenz für das Offenheits-Konstrukt.

Allerdings zeigt sich bei allen Skalen des BFI-S eine hohe Retest-Stabilität der Werteverteilung. Dies ist angesichts der geringen inneren Konsistenz der Offenheit-Items überraschend.

Wie Tabelle 11 zeigt, sind auch die Mittelwerte und Streuung im Retest praktisch unverändert. Alle Retest-Koeffizienten liegen über .75. Für alle Skalen besteht somit eine hohe Retest-Reliabilität.

Tabelle 11:
Mittelwertsstabilität und Retest-Reliabilität der Persönlichkeitsskalen (BFI-S) nach 6 Wochen

Konstrukte	Mittelwerte (SD)		Test-Retest Koeffizient (N = 115)
	1. Erhebung (N = 119)	Retest (N = 116)	
E – Extraversion	5.2 (1.2)	5.2 (1.2)	.84
A – Verträglichkeit	5.4 (1.0)	5.4 (1.1)	.76
C - Gewissenhaftigkeit	5.3 (1.2)	5.2 (1.1)	.76
N – Neurotizismus	3.8 (1.3)	3.7 (1.3)	.80
O - Offenheit für Erfahrung	4.9 (1.1)	5.0 (1.1)	.82

Tabelle 12 zeigt die konvergenten und divergenten Korrelationen des BFI-S mit dem NEO-FFI. Wie erwartet zeigen bis auf O alle Konstrukte des BFI-S hohe Korrelationen mit den korrespondierenden NEO Skalen. Die divergenten Korrelationen liegen unter .36.

Tabelle 12:
Synchrone Konstruktvalidität des BFI-S zur ersten Erhebung und im Retest

EZP	BFI-S	NEO-FFI (60 Items)				
		E	A	C	N	O
1	E Extraversion	.71	.06	.12	-.25	.02
	A Verträglichkeit	.14	.67	.22	-.12	.09
	C Gewissenhaftigkeit	.04	.11	.81	.05	-.17
	N Neurotizismus	-.36	.13	-.02	.71	-.19
	O Offenheit für Erfahrung	-.28	.09	.07	-.13	.43
Re-test	E Extraversion	.69	.08	.13	-.34	.08
	A Verträglichkeit	.18	.67	.27	-.01	-.04
	C Gewissenhaftigkeit	.15	.14	.72	-.07	-.16
	N Neurotizismus	-.35	-.02	-.07	.74	-.19
	O Offenheit für Erfahrung	.36	.15	.08	-.15	.48

Anm.: Kursive Korrelationskoeffizienten sind signifikant ($p < .05$), Retest (MZP 2) nach 6 Wochen.

N und O des BFI korrelieren jedoch unerwartet hoch mit der Extraversionsskala des NEO-FFI. Für die Offenheitsindikatoren weisen deren hohe Retest-Stabilität bei geringer innerer Konsistenz und Divergenz darauf hin, dass nicht die gesamte Breite des Offenheitskonstrukts repräsentativ und adäquat abgedeckt wird. Vor diesem Hintergrund wird empfohlen, in zukünftigen Applikationen des BFI-S ein zusätzliches Offenheits-Item aufzunehmen (BFI-S16). Das Item „*(Ich bin jemand, der) wissbegierig ist*“ deckt eine zusätzliche Facette der Offenheit ab und verbreitert die Basis des Konstrukts, ohne die innere Konsistenz und Retest-Reliabilität zu gefährden. Das empfohlene Item lehnt sich an ein robustes, ähnliches Item des NEO-FFI an. Die Hinzunahme eines weiteren Item in die Offenheitsskala (4 Items) ist insofern gerechtfertigt, da sich O in den meisten Instrumenten erfahrungsgemäß heterogen und wenig robust darstellt, gleichwohl aber dispositionelle Grundlagen der Werte-Bindung, Wertorientierung und Informationssuche erhebt.

7 Zusammenfassung und Ausblick

In einer Validierungsstudie wurden Reliabilität und Validität zweier ultrakurzer CAPI-Tests zur Erfassung der kognitiven Funktionstüchtigkeit sowie eine CAPI-Version des BFI-S überprüft. Die kognitiven Leistungstests (Zeichentest, Tiere-Nennen-Aufgabe) wurden gemeinsam mit dem DIW und TNS Infratest entwickelt und für eine Erfassung mit Computer-Assisted-Personal-Interviewing (CAPI) für den Einsatz im Rahmen von breit angelegten Survey-Erhebungen adaptiert. Die Auswahl der beiden kognitiven Funktionsparameter erfolgte dabei auf der Basis der Zweikomponententheorie der intellektuellen Fähigkeiten (Baltes et al., 1998), die zwischen der kognitiven Mechanik und der Pragmatik mentaler Leistungen differenziert und für diese unterschiedliche Lebenslauf-Trajektorien vorhersagt. Die Leistungstests bilden die zugrunde liegende Differenzierung mechanischer und pragmatischer kognitiver Leistungen zuverlässig und robust ab. Für beide ultrakurze Testverfahren kann eine zufriedenstellende Konstruktvalidität konstatiert werden, wobei sich die Güte des Zeichentests merklich positiver darstellt, als die der Tiere-Nennen-Aufgabe.

Die CAPI-Version der Tiere-Nennen-Aufgabe erweist sich in der Durchführung als leicht stör anfällig, wobei insbesondere Leistungsbereitschaft und Aufmerksamkeit der Interviewer entscheidend sein dürften. Dies führt zu einer insgesamt eher geringen Stabilität des Tests und einer im Vergleich zum Zeichentest suboptimalen Homogenität und inneren Konsistenz des Testverfahrens. Zu berücksichtigen ist, dass Testverfahren der kristallinen Fähigkeiten in der Regel eine größere Heterogenität aufweisen als fluide Testverfahren. So zeigt die Tiere-Nennen-Aufgabe eine im Sinne der Erwartung eindrucksvoll hohe Korrelation mit der Bildungsdauer der Studienteilnehmer, die sogar höher ausfällt als die konvergente Validität der TA mit dem Wortschatztest. Dieser Befund belegt, dass die Tiere-Nennen-Aufgabe sich im Prinzip gut bewährt, wenn es darum geht, in groß angelegten Umfragen neben einem bewährten, fluiden (und kulturfreien) Fähigkeitsparameter auch einen bildungsabhängigen, differenzierenden kognitiven Leistungstest zu berücksichtigen.

Die Durchführung der beiden CAPI-Tests benötigte einen Zeitumfang von jeweils 90 Sekunden und bedarf eines nur minimalen Interviewertrainings. Trotz einer begrenzten Anfälligkeit für Interviewer-Effekte, erlauben die zwei Testparameter eine reliable, robuste und valide, wenn auch nur grobe Abschätzung der intellektuellen Leistungsfähigkeit der Teilnehmer, die sich über ein heterogenes Spektrum fluider (mechanischer) und kristalliner (pragmatischer)

Komponenten der kognitiven Kapazitäten erstreckt. Das Testverfahren ist damit gut geeignet zum Einsatz in Interviewer basierten, altersheterogenen Umfragen sowie in Fällen, bei denen keine Einzelfalldiagnostik erforderlich ist.

Literatur

- Baltes, P. B. (1997). On the incomplete architecture of human ontogeny: Selection, optimization, and compensation as foundation of developmental theory. *American Psychologist*, 52, 366 - 380.
- 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, pp. 1029 – 1143). New York: Wiley.
- Bäumler, G. (1974). *Lern- und Gedächtnistest: LGT-3*. Göttingen: Hogrefe.
- Borkenau, P. & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa und McCrae*. Göttingen: Hogrefe.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. Berlin: Springer.
- Brickenkamp, R. (2002). *Test d2: Aufmerksamkeits-Belastungs-Test*. Göttingen: Hogrefe.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin*, 103, 276-279.
- Gerlitz, J.-Y. & Schupp, J. (2005, Juli). Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP. Berlin: DIW Berlin (DIW Research Note 4/2005).
- Heckhausen, J. (1999). *Developmental regulation in adulthood*. New York: Cambridge.
- Horn, W. (1983). *Leistungsprüfungssystem: L-P-S*. Göttingen: Hogrefe.
- John, O. & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102-138). New York: The Guilford Press.
- Lang, F. R. & Heckhausen, J. (in Druck). Allgemeine Entwicklungspsychologie der Lebensspanne: Begriffe, Theorien und Befunde. In K. Pawlik (Hrsg.), *Psychologie*. Berlin: Springer.
- Lang, F. R., Lüdtke, O., & Asendorpf, J. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen. *Diagnostica*, 47, 111 – 121.
- Lehrl, S. (1995). *Mehrfachwahl-Wortschatz-Test (MWT)*. Balingen: PERIMED-spitta.
- Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz PVU.
- Lindenberger, U. (2002). Erwachsenenalter und Alter. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (5. Auflage, S. 350-391). Weinheim: Beltz PVU.
- Lindenberger, U. & Baltes, P. B. (1995). Kognitive Leistungsfähigkeit im Alter: Erste Ergebnisse aus der Berliner Altersstudie. *Zeitschrift für Psychologie*, 203, 283-317.
- Projektgruppe Sozio-oekonomisches Panel* (1995). *Zehn Jahre Sozio-oekonomisches Panel (SOEP). Vierteljahreshefte zur Wirtschaftsforschung*, 62, 27-42.
- Raven, J. C. & Kratzmeyer, H. (1980). *APM: Raven Matrizen Test – Advanced Progressive Matrices*. Weinheim: Beltz-Test.
- Rosenblatt, B.v. & Stocker, A. (2005, August). *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: Western Psychological Services.

Solga, H. et al. (2005). The measurement and importance of general reasoning potentials in schools and labor markets. Berlin: DIW Berlin (DIW Research Note 10/2005).

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

Weiss, R. H. (1987). Grundtest Skala 2: CFT-20 mit Wortschatztest und Zahlenfolgetest (ZF). Göttingen: Hogrefe.