

## IV. METHODE

### 4.1 Stichprobe

Kognitive Plastizität wurde in vorliegender Studie in zwei Altersgruppen untersucht: Zum einen in einer Hauptstichprobe von 96 Personen im hohen Alter, die sich aus Teilnehmern der Berliner Altersstudie zusammensetzte ( $M_{\text{Alter}} = 83.7$  Jahre;  $SD = 4.8$ ,  $Range = 75\text{--}101$  Jahre), zum anderen in einer Vergleichsstichprobe von 20 jüngeren Erwachsenen zwischen 20 und 30 Jahren, die anhand persönlicher Kontakte gewonnen wurde ( $M_{\text{Alter}} = 24.0$  Jahre;  $SD = 2.4$ ,  $Range = 20\text{--}29$  Jahre).<sup>5</sup> Die junge Vergleichsstichprobe hatte hierbei primär die Funktion, eine Einschätzung über das Ausmaß altersbedingter Reduktion kognitiver Plastizität im letzten Lebensabschnitt zu erhalten.

#### 4.1.1 Stichprobe der 75- bis 101jährigen

Das Hauptziel bei der Rekrutierung der Stichprobe der Personen im hohen Alter lag in der Gewinnung aller in den Jahren 1998 und 1999 Überlebenden der Berliner Altersstudie, die das Intensivprotokoll der vierten Erhebungswelle von BASE vollständig abgeschlossen und ihre Einwilligung für eine weitere Teilnahme an der Berliner Altersstudie gegeben hatten ( $N = 126$ ). Hierbei befand sich die Berliner Altersstudie zu Beginn der Datenerhebung im Januar 1998 gerade in ihrer vierten Erhebungswelle, die von Februar 1997 bis April 1998 andauerte (für eine Beschreibung der Stichprobe sowie des Studiendesigns der längsschnittlichen Berliner Altersstudie siehe Anhang A und Abb. 3).

Von den 126 Überlebenden der Berliner Altersstudie nahmen 96 (76 %) Personen an der vorliegenden Studie teil. Die Ausfallgründe für die verbleibenden 30 Personen sind in Tabelle 4 im Rahmen der Darstellung des Rekrutierungsprozesses aufgeführt.

---

<sup>5</sup> Alle im folgenden Text dargestellten Altersangaben beziehen sich auf das Alter, das die Probanden am Tage der ersten Sitzung (Prätest I) hatten. Aus Gründen der Übersichtlichkeit wird für die Darstellung der Altersverteilungen abgerundet, für alle statistischen Analysen hingegen die kontinuierliche Altersvariable herangezogen.

#### 4.1.1.1 Beschreibung

Unter den 96 Studienteilnehmern der entgeltigen Studienstichprobe ( $M_{\text{Alter}} = 83.7$  Jahre;  $SD = 4.8$ ,  $\text{Range} = 75\text{--}101$  Jahre) waren 60 % ( $n = 58$ ) Frauen und 40 % ( $n = 38$ ) Männer. Zum Zwecke der Beschreibung der Alters- und Geschlechterverteilungen der vorliegenden Stichprobe der Personen im hohen Alter wurden in Tabelle 2 die 96 Studienteilnehmer in vier Altersgruppen aufgeteilt.

Tabelle 2  
Alters- und Geschlechterverteilung der Stichprobe der 75- bis 101jährigen ( $N = 96$ )

<i>n</i>	Altersgruppe (Jahre)			
	75–79	80–84	85–89	90+
Gesamt	24	37	24	11
Frauen	13	22	16	7
Männer	11	15	8	4

Die Mehrheit der Teilnehmer (64 %) verteilte sich auf die zwei jüngeren Altersgruppen (unter 85 Jahren). Auch nahm der prozentuale Anteil an Frauen in den vier Altersgruppen mit zunehmendem Alter zu (in der Gruppe der 75- bis 79jährigen lag der Anteil an Frauen bei 54 %, in der Gruppe der 80- bis 84jährigen stieg er auf 59 % und in der Gruppe der über 90jährigen war er bei 63 %). Die beobachteten ungleichen Alters- und Geschlechterverteilungen in der vorliegenden Stichprobe entsprachen demographischen Beobachtungen einer ungleichen Alters- und Geschlechterverteilung in der alten Bevölkerung zugunsten der 'jungen' Alten und der Frauen. So gibt es beispielsweise in der Westberliner Gesamtpopulation etwa fünfmal mehr 70- als 90jährige und etwa fünfmal mehr 90jährige Frauen als Männer (vgl. Mayer & Baltes, 1996).<sup>6</sup>

<sup>6</sup> Es sei daran erinnert, daß die Ausgangsstichprobe der Berliner Altersstudie (1990–1993) ursprünglich nach Alter und Geschlecht geschichtet und daher nicht repräsentativ für die Demographie der Population der 70- bis über 100jährigen war. Die beobachteten ungleichen Alters- und Geschlechterverteilungen in der vorliegenden Stichprobe sind daher das Ergebnis selektiver Mortalitätseffekte in dem Zeitraum zwischen dem ersten Meßzeitpunkt von BASE und dem Beginn der Trainingsstudie. Die geringere Anzahl von Personen in der Gruppe der 75- bis 79jährigen im Vergleich zu der in der Gruppe der 80- bis 85jährigen ist auf die

Tabelle 3  
Demographische Charakteristika der Stichprobe der 75- bis 101jährigen ( $N = 96$ )

Demographische Charakteristika	Gesamt $N = 96$		Frauen $n = 58$		Männer $n = 38$	
	$M$	$SD$	$M$	$SD$	$M$	$SD$
Alter in Jahren	83.7	4.8	84.1	5.1	83.2	4.4
Bildung in Jahren	11.4	2.3	11.0	2.1	12.0	2.4
Kognitiver Gesundheitsstatus (Nachdiagnose mit MMSE)	27.3	2.2	27.4	2.0	27.1	2.3

In der Stichprobe der Personen im hohen Alter waren zwei Teilnehmer in stationären Einrichtungen untergebracht. Die Mehrheit der 96 Probanden ( $n = 94$ ) lebte in Privathaushalten, davon 9 in Seniorenwohnheimen und 85 in ihren Privatwohnungen. Von den 96 Studienteilnehmern hatten 49 % einen Volksschulabschluß (8 Jahre) und 51 % einen höheren Schulabschluß, davon 10 Personen das Abitur (13 Jahre) und 39 einen Realschulabschluß (10 Jahre). 64 % der Probanden mit Volksschulabschluß und 69 % der Teilnehmer mit höherem Schulabschluß hatten zusätzlich eine Lehre beendet. Im Mittel betrug die Bildungsdauer in der vorliegenden Stichprobe 11.40 Jahre ( $SD = 2.29$ ;  $Range = 8$ –17 Jahre). Hinsichtlich der Klassenzugehörigkeit zählten 19.8 % der Teilnehmer zur Unter- oder unteren Mittelschicht, die Mehrheit (67.7 %) zur mittleren oder gehobenen Mittelschicht und 12.5 % zur oberen Mittelschicht. Das mittlere, monatliche bedarfsgewichtete Haushaltseinkommen lag bei 2.574,- DM ( $SD = 1467.69$ ;  $Range = 950$ –12.000 DM). Mehr als die Hälfte der Teilnehmer (55 %) waren gebürtige Berliner. Alle Teilnehmer gaben an, Berlin sehr gut zu kennen.

#### 4.1.1.2 Rekrutierung

Das einzige Selektionskriterium bei der Zusammenstellung der Stichprobe lag in dem Ausschluß von Personen mit einer mittleren oder schweren Demenz. Aus zahlreichen Studien war bekannt, daß ein dementielles Syndrom mit schwerwiegenden Funktionsstörungen des

---

Tatsache zurückzuführen, daß das jüngste Alter in der vor im Mittel 6.5 Jahren begonnenen Berliner Altersstudie bei 70 Jahren lag.

episodischen Gedächtnis und der Lernfähigkeit einhergeht (Bäckman et al., 2000; Baltes, Kühl & Sowarka, 1992; Reischies & Lindenberger, 1996). Zudem ist bei Dementen auch die Fähigkeit reduziert, von unterschiedlichen Formen kognitiver Hilfestellung (wie beispielsweise einer Instruktion in eine leistungsoptimierende Strategie oder der Darbietung von Hinweisreizen) zu profitieren, die bei 'gesunden älteren Erwachsenen' erfolgreich eingesetzt werden (für Übersichten siehe Bäckman et al., 1990, 2000).

Die Selektion dementer Personen im Vorfeld der Stichprobenrekrutierung erfolgte anhand der psychiatrischen Demenzdiagnosen, die nach DSM-III-R-Kriterien gestellt worden waren (Diagnostic and Statistical Manual of Mental Disorders, 3rd rev., American Psychiatric Association, 1987). Hierin wird zwischen leichter (1), mittel schwerer (2) und schwerer Demenz (3) unterschieden. Aus der Stichprobe ausgeschlossen wurden die Personen mit mittel schwerer (2) und schwerer (3) Demenz, die als 'leicht' dement (1) diagnostizierten Probanden durften teilnehmen, da der a priori Ausschluß von Personen minimiert werden sollte und bekannt war, daß sowohl die Validität als auch die Reliabilität der 'leichten' Demenzdiagnose nicht perfekt sind. Da die psychiatrischen Demenzdiagnosen nach DSM-III-R-Kriterien aus der vierten Erhebungswelle der Berliner Altersstudie (1997–1998) nicht zur Verfügung standen, dienten zur Vorauswahl die Demenzdiagnosen der dritten Erhebungswelle (1995–1996).

Der Erstkontakt mit den Teilnehmer der Berliner Altersstudie erfolgte über ein Anschreiben, in dem Informationen über das Hauptziel der Untersuchung, den einzuplanenden Zeitaufwand und den Unkostenbeitrag enthalten waren (siehe Anhang B). Nach einer Woche kontaktierten zwei Versuchsleiter, deren Namen im Anschreiben erwähnt wurden, die Teilnehmer nochmals telefonisch, um zusätzliche Informationen zu geben und im Falle einer Zusage bereits die ersten Termine zu vereinbaren.

Die Datenerhebung der vorliegenden Studie begann Anfang Januar 1998 und wurde Ende Dezember 1998 abgeschlossen. Im Mittel lagen zwischen dem Anschreiben und der ersten Testsitzung (Prätest I) 43.17 Tage ( $SD = 55.65$ ).<sup>7</sup> Die Auswahl der Personen, die jeweils angeschrieben wurden, erfolgte nach dem Zufall. Es mußte lediglich zwischen der

---

<sup>7</sup> Da vermieden werden sollte, daß ein zu großer Zeitraum zwischen erstem Briefkontakt und dem tatsächlichen Beginn an der acht Sitzungen umfassenden Trainingsstudie lag, wurden Anfang Januar 1998 nicht alle potentiellen Teilnehmer gleichzeitig angeschrieben; es wurde immer nur die Anzahl an Personen schriftlich kontaktiert, die innerhalb eines Zeitraums von zwei Wochen nach dem ersten Anschreiben auch tatsächlich getestet werden konnten. Das mittlere Zeitintervall von etwa 1.5 Monaten war vor allem auf jene Personen zurückzuführen, die zwar ihre Teilnahme an der Trainingsstudie aufgrund des Anschreibens zusicherten, den Beginn der Studie aber aufgrund von Reiseplänen oder Krankheiten auf spätere Erhebungsblöcke verschoben.

Abschlußsitzung der vierten Erhebungswelle der Berliner Altersstudie und dem Anschreiben mindestens ein Monat vergangen sein. Im Mittel lagen 5.01 Monate ( $SD = 2.48$ ) zwischen der letzten Sitzung der vierten Erhebungswelle der Berliner Altersstudie und dem Anschreiben und 6.43 Monate ( $SD = 3.03$ ) zwischen der letzten BASE-Sitzung und der ersten Sitzung der vorliegenden Trainingsstudie (Prätest I).

Die Probanden erhielten für ihre Teilnahme an der acht Sitzungen umfassenden Trainingsstudie 350,- DM, was ungefähr 30,- DM pro Zeitstunde ergab. Dieser Betrag entspricht in etwa dem Unkostenbeitrag, den die Teilnehmer aus früheren Erhebungen im Rahmen der Berliner Altersstudie gewohnt waren.

Von den 126 Überlebenden der Berliner Altersstudie, die das Intensivprotokoll des vierten Meßzeitpunktes (1997–1998) vollständig abgeschlossen und ihre Zusage für eine weitere Kontaktaufnahme gegeben hatten, nahmen 76 % an der Studie teil. Tabelle 4 gibt einen Überblick über die Gründe für den Ausfall der verbleibenden 30 Personen.

Tabelle 4  
Gründe für den Stichprobenausfall

<i>n</i>	%	Gründe für den Ausfall
126	100	Anzahl der Überlebenden mit vollständig abgeschlossenem Intensivprotokoll des vierten Meßzeitpunktes (1997–1998) und Einwilligung für eine weitere Kontaktaufnahme
5	4	Mittel schwere oder schwere Demenz nach DSM-III-R (APA, 1987) zum dritten Meßzeitpunkt von BASE
5	4	Schriftlich und telefonisch nicht erreichbar
13	10	Teilnahme verweigert
7	6	Studienabbruch
96	76	Vollständige Teilnahme an der Studie

Fünf Personen (4 %) mußten aus der Stichprobe ausgeschlossen werden, da sie bereits zum dritten Meßzeitpunkt der Berliner Altersstudie (1995–1996) eine mittel schwere oder schwere Demenzdiagnose (nach DSM-III-R, APA, 1987) erhalten hatten ( $M_{\text{Alter}} = 93.6$  Jahre;  $\text{Range} =$

87–102 Jahre).<sup>8</sup> Fünf weitere Personen (4 %) waren weder schriftlich noch telefonisch erreichbar ( $M_{\text{Alter}} = 85.8$  Jahre;  $\text{Range} = 76\text{--}100$  Jahre). Lediglich 13 Personen (10 %) verweigerten ihre Teilnahme an der Studie ( $M_{\text{Alter}} = 84.15$  Jahre;  $\text{Range} = 76\text{--}95$  Jahre). Angegebene Gründe für die Absage waren vor allem längere Abwesenheiten aufgrund von Reisen oder gesundheitlichen Problemen. Von den 103 (82 %) Studienteilnehmern brachen sieben Personen (7 %) aufgrund von gesundheitlichen Problemen oder zu hoher Belastung die Studie nach einigen Sitzungen ab ( $M_{\text{Alter}} = 92$  Jahre;  $\text{Range} = 84\text{--}105$  Jahre).

Die Altersangaben der Nicht-Teilnehmer zeigen deutlich, daß die Altersverteilung der vorliegenden Stichprobe vor allem durch den Ausschluß mittel schwer und schwer dementer Probanden und der Personen, die aus gesundheitlichen Gründen die Studie abbrechen mußten, im oberen Bereich reduziert war. So fielen aus diesen Kategorien beispielsweise drei Überlebende im Alter von 101, 102 und 105 Jahren und vier über 90jährige aus ( $M_{\text{Alter}} = 92.66$  Jahre;  $\text{Range} = 84\text{--}105$  Jahre).

Zusammengenommen waren die Teilnehmer der Berliner Altersstudie hoch motiviert, was sich sowohl in der unerwartet hohen Teilnehmerate von 76 % als auch in der geringen Anzahl von Terminverschiebungen nach Studienbeginn äußerte. Einige Probanden bestanden sogar dann noch auf die Einhaltung ihres wöchentlichen Termins, wenn sie sich aufgrund leichter gesundheitlicher Komplikationen im Krankenhaus aufhalten mußten.

#### 4.1.1.3 Erfassung von Demenz

Die Stichprobe der 75- bis 101jährigen enthielt nur drei Personen (3 %), die zum dritten Meßzeitpunkt der Berliner Altersstudie (1995–1996) als ‘leicht’ dement (nach DSM-III-R, APA, 1987) klassifiziert worden waren. Da die aktuellen Demenzdiagnosen nach DSM-III-R des vierten Meßzeitpunktes der Berliner Altersstudie nicht zur Verfügung standen, wurde zur Erfassung des aktuellen mentalen Status der Probanden zusätzlich ein klinisches Maß, der neuropsychologische *Mini Mental State Examination Test* (MMSE; Folstein, Folstein & McHugh, 1975) erhoben, der häufig in epidemiologischen Demenzstudien als Screeningverfahren eingesetzt wird. Die Untersuchung erfolgte nach der deutschen Übersetzung des SIDAM (Zaudig et al., 1991). Der MMSE-Test ist eine Checkliste aus Fragen zur Orientierung

---

<sup>8</sup> Da die Prävalenzrate für Demenz in der Ausgangsstichprobe der Berliner Altersstudie bei 14 % lag, war ein a priori-Ausschluß von nur 4 % der potentiellen Teilnehmer unerwartet gering.

(zeitlich, örtlich), einfachen Gedächtnisaufgaben und Items zu anderen wichtigen psychopathologischen Symptombereichen. Die Items sind so konstruiert, daß sie von gesunden Personen ohne Probleme gelöst und beantwortet werden können. Es kann eine maximale Punktzahl von 30 Punkten erreicht werden. Der MMSE-Test ist heute wahrscheinlich die international am häufigsten eingesetzte Demenzskala.

Die Analysen des zu Beginn der Trainingsstudie, das heißt etwa drei Jahre nach den berichteten Demenzdiagnosen, erhobenen MMSE-Tests ergaben, daß diese drei als 'leicht dement' diagnostizierten Personen Gesamtpunktwerte von 21, 26 und 26 Punkten erzielten. Diese relativ hohen MMSE-Werte sprechen dafür, daß es sich zumindest bei den letzten beiden Fällen nicht um demente Personen handelte, da aufgrund des progredienten Verlaufs der dementiellen Erkrankung zu Beginn der Studie niedrigere MMSE-Werte hätten erwartet werden müssen (< 20 Punkte). Die mittlere Leistung der gesamten Stichprobe lag im MMSE-Test bei 27.26 Punkten ( $SD = 2.15$ ;  $Range = 21-30$  Punkte). Dieser Wert entspricht dem Mittelwert von 27.6 Punkten, den Folstein, Folstein und McHugh (1975) in einer Stichprobe von 63 'normalen', nicht-klinisch erkrankten älteren Erwachsenen vorfanden.

#### **4.1.2 Stichprobe der 20- bis 30jährigen: Rekrutierung und Beschreibung**

Die Vergleichsstichprobe der jungen Erwachsenen setzte sich aus 20 Personen (9 Frauen, 11 Männern) zwischen 21 und 29 Jahren mit einem Altersdurchschnitt von 24 Jahren zusammen.

Die Stichprobe der jungen Erwachsenen wurde anhand persönlicher Kontakte gewonnen. Ziel bei der Rekrutierung war es, eine möglichst heterogene Stichprobe zu erhalten. Studierende der Psychologie durften an der Studie nicht teilnehmen, da mitunter befürchtet werden mußte, daß mnemonische Techniken bereits aus dem Studium bekannt waren.

Die jungen Erwachsenen bekamen für die acht Sitzungen umfassende Studie 160,- DM, was in etwa einer Vergütung von 20,- DM pro Zeitstunde entsprach.

Von den ursprünglich 21 Teilnehmern brach eine Person nach drei Sitzungen die Untersuchung aus mangelndem Interesse ab. Zwei (10 %) der jungen Erwachsenen hatten einen universitären Abschluß, 15 Personen (75 %) das Abitur und drei (15 %) die mittlere Reife erlangt. Alle jungen Erwachsenen gaben an, in guter gesundheitlicher Verfassung zu sein und Berlin sehr gut zu kennen.

## 4.2 Erhebungsinstrumente

Die in vorliegender Studie verwendeten Meßinstrumente hatten hauptsächlich zwei Funktionen: Sie dienten (a) der Erfassung individueller Unterschiede in unterschiedlichen Funktionsbereichen zur Vorhersage individueller Unterschiede im Ausmaß an beobachtbarer kognitiver Plastizität im hohen Alter (unabhängige Variablen) und (b) der Erfassung der Gedächtnisleistungen vor und nach Instruktion in und Training mit der Methode der Orte (abhängige Variable). Ausnahmen bildeten (a) die begleitenden Fragen zur Erfassung der Anwendungshäufigkeit und -qualität der Methode der Orte, die der Kontrolle trainingsspezifischer Effekte mit der Methode der Orte dienten, und (b) die von der Psychiatrischen Einheit erfaßten Demenzdiagnosen (nach DSM-III-R, APA, 1987), die für die Vorauswahl der Untersuchungsstichprobe der Personen im hohen Alter herangezogen wurden.

### 4.2.1 Prädiktoren

Zur Vorhersage individueller Unterschiede in der kognitiven Plastizität wurden individuelle Unterschiedsvariablen genutzt, die im Rahmen der vierten und letzten Erhebungswelle der Berliner Altersstudie (1997–1998) erhoben wurden.<sup>9</sup> Die Verwendung der im Rahmen der Berliner Altersstudie erhobenen Maße zur Prädiktion von Plastizität hatte mehrere Vorteile: (a) Es konnten längsschnittliche Veränderungsmessungen zur Vorhersage von Plastizität herangezogen werden; (b) es standen interdisziplinäre Datensätze zur Verfügung, die von Fachkräften der jeweiligen Disziplinen erhoben wurden. Um Mißverständnisse zu vermeiden, sollte nochmals betont werden, daß das zur Verfügung stehende längsschnittliche Prädiktionsystem keine Maße der Entwicklungskapazität bzw. kognitiven Plastizität enthielt.

Als Indikatoren des biologischen Einflußsystems dienten mechanisch-fluide Maße kognitiver Leistungsfähigkeit sowie des Hörvermögens und der Sehschärfe. Als Indikatoren für das kulturelle Einflußsystem wurden von den Soziologen erhobene soziostrukturell-biographische Variablen wie das Einkommen, die Bildung und die soziale Klasse einer Person sowie wissensbasierte Maße pragmatisch-kristalliner Intelligenz herangezogen (vgl. Abschnitt 3.1.2).

---

<sup>9</sup> Im Mittel lagen zwischen der Abschlußsitzung der vierten Erhebungswelle der Berliner Altersstudie und der ersten Sitzung vorliegender Studie 6.43 Monate ( $SD = 3.03$ ).

Zusätzlich zu diesen Statusmaßen, die den aktuellen Leistungsstatus der Probanden vor Beginn der Trainingsstudie repräsentierten, wurden Veränderungsmaße zur Vorhersage von Plastizität gebildet, die die individuelle Leistungsveränderung der BASE-Teilnehmer über einen mittleren Zeitraum von 6.5 Jahren ( $SD = 0.85$ ) abbildeten. Die Tabelle 5 gibt einen Überblick über die individuellen Unterschiedsvariablen, die zur Vorhersage individueller Unterschiede in kognitiver Plastizität verwendet wurden. Eine detaillierte Beschreibung der einzelnen Maße erfolgt anschließend.

#### **4.2.1.1 Kognitive Leistungsfähigkeit**

Die kognitive Leistungsfähigkeit wurde mit einer in der Tradition der psychometrischen Intelligenzforschung (Carroll, 1993) stehenden Testbatterie erfaßt (vgl. Lindenberger et al., 1993). Im Rahmen der Berliner Altersstudie wurden längsschnittlich acht kognitive Tests benutzt, wobei jeweils zwei dieser Tests einer von vier unterschiedlichen kognitiven Fähigkeiten zugeordnet waren. Im einzelnen waren dies

- (a) Wahrnehmungsgeschwindigkeit (Gleiche Bilder, Zahlen-Buchstaben),
- (b) Gedächtnis (Wortpaare, Geschichte erinnern),
- (c) Wortflüssigkeit (Tiere nennen, Wortanfang mit 'S') und
- (d) Wissen (Wortschatz, Wörter finden).

Zur Bildung dieser vier Fähigkeitskonstrukte wurden die Werte der jeweiligen kognitiven Tests  $z$ -standardisiert zusammengefaßt. In Anlehnung an die Unterscheidung zwischen fluider und kristalliner Intelligenz (Cattell, 1971; Horn, 1982) bilden die Fähigkeiten Wissen und Wortflüssigkeit die pragmatisch-kristalline Komponente der Intelligenz und die Fähigkeiten Wahrnehmungsgeschwindigkeit und Gedächtnis die mechanisch-fluide Komponente ab (Lindenberger & Baltes, 1995a, 1997). Eine detaillierte Übersicht der Reliabilitäten und Validitäten der einzelnen kognitiven Tests findet sich in Lindenberger et al. (1993). Eine Darstellung der Interkorrelationen der vier kognitiven Fähigkeiten in der vorliegenden Stichprobe erfolgt an späterer Stelle (siehe Abschnitt 5.2.1, Tab. 10).

Tabelle 5  
Liste der Prädiktoren zur Vorhersage individueller Unterschiede in kognitiver Plastizität

Konstrukte	Indikatoren	Quelle
<i>Maße kognitiver Leistungsfähigkeit</i>		
Wahrnehmungsgeschwindigkeit	Gleiche Bilder	Ekstrom et al., 1976 Lindenberger et al., 1993
	Zahlen-Buchstaben	Lindenberger et al., 1993 Wechsler, 1955
Gedächtnis	Wortpaare	Lindenberger et al., 1993
	Geschichte erinnern	Lindenberger et al., 1993 Engel & Satzger, 1990
Wortflüssigkeit	Tiere nennen	Lindenberger et al., 1993
	Wortanfang mit 'S'	Lindenberger et al., 1993
Wissen	Wortschatz	Lindenberger et al., 1993 Wechsler, 1982
	Wörter finden	Lindenberger et al., 1993 Lehrl, 1977
<i>Maße sensorischer Leistungsfähigkeit</i>		
Sehschärfe	Nah- und Fernvisus	Geigy, 1977 Marsiske et al., 1999
Gehör	Audiometrie	Lindenberger & Baltes, 1994
<i>Kulturbezogene Indikatoren: Soziostrukturell-biographische Variablen</i>		
Bildung	Bildungsdauer in Jahren	Mayer & Wagner, 1996
Einkommen	Bedarfsgewichtetes Haushaltseinkommen	Motel & Wagner, 1993
Soziale Schicht	Klassenindex (5 Klassen)	Handl, 1977; Mayer, 1977 Mayer & Wagner, 1996

### Gleiche Bilder

Es wurde eine computerisierte Version des Identical Pictures Tests aus dem ETS Kit of Factor-Referenced Tests verwendet (Ekstrom et al., 1976). Bei jedem der 32 Items wurde eine geometrische Figur in der oberen und fünf Antwortmöglichkeiten in der unteren Bildschirmhälfte dargeboten. Die Probanden sollten die mit dem oberen Bild identische Figur möglichst schnell berühren. Nach 80 Sekunden wurde der Test automatisch beendet. Die Leistungswerte beziehen sich auf die Anzahl korrekter Antworten in diesem Zeitraum.

### Zahlen-Buchstaben-Test

Bei diesem Test handelt es sich um eine adaptierte Version des Zahlensymboltest aus dem WAIS von Wechsler (1955; vgl. Lindenberger et al., 1993). In diesem Test waren neun verschiedene Buchstaben jeweils einer Zahl von 1 bis 9 zugeordnet. Der wesentliche Unterschied zum Zahlensymboltest bestand darin, daß der Proband Buchstaben nennen und nicht Symbole aufschreiben sollte, was die Auswirkungen motorischer Beeinträchtigungen minimieren sollte. Der Test bestand aus 39 Tafeln, auf denen sich sechs Zahlen befanden, unter denen jeweils ein Fragezeichen stand. Der Proband sollte von links nach rechts die den Zahlen entsprechenden Buchstaben nennen. Hierbei war eine Zuordnungstabelle, in der neun verschiedene Großbuchstaben jeweils einer Zahl von 1 bis 9 zugeordnet waren, während der ganzen Testphase sichtbar. Auch waren Zahlen und Buchstaben in großer Schrift gedruckt, um den Einfluß möglicher visueller Beeinträchtigungen zu reduzieren. Die Leistungswerte entsprechen der in einem Zeitraum von drei Minuten genannten Anzahl richtiger Buchstaben.

### Wortpaare

Es handelte sich um eine Paarassoziationsaufgabe mit Hinweisreizen. Acht Wortpaare wurden zweimal hintereinander in unterschiedlicher Reihenfolge mit einer Geschwindigkeit von fünf Sekunden pro Wortpaar auf dem Bildschirm dargeboten. Hierbei bezogen sich die Wörter auf konkrete Objekte und hatten in einer Normierungsstudie auf einer Bildhaftigkeitsskala von 1 (wenig bildhaft) bis 7 (sehr bildhaft) Werte über 6 erhalten (Baschek, Bredenkamp, Öhrle & Wippich, 1977). Beim Abruf wurde jeweils das erste Wort eines Wortpaares gezeigt und der Proband sollte sich an das dazugehörige Wort erinnern. Die Leistungswerte entsprechen der Gesamtzahl der richtigen Antworten und variieren zwischen 0 und 16 (vgl. Lindenberger et al., 1993).

### Geschichte erinnern

Bei diesem Test handelte es sich um eine leicht veränderte Version des Tests von Engel und Satzger (1990; vgl. Lindenberger et al., 1993). Es wurde die Geschichte eines Jungen erzählt, der

Angeln ging, ausrutschte und von seinem Hund gerettet wurde. Die Geschichte erschien in großen Buchstaben und wurde gleichzeitig von dem/der Forschungstechnischen AssistentIn laut und deutlich vorgelesen. Nach der etwa 38 Sekunden dauernden Vorlesephase erfolgte die Erinnerungsphase. In dieser wurden sukzessiv sechs, auf unterschiedlichem Detailniveau angesiedelte Fragen gestellt und von zwei unabhängigen Ratern mit falsch (0) oder richtig (1) bewertet. Die Leistungswerte entsprechen der Anzahl richtig beantworteter Fragen und reichen daher von 0 bis 6.

### Tiere nennen

Hier mußten in 90 Sekunden möglichst viele verschiedenen Tiere genannt werden. Die Antworten wurden von zwei unabhängigen Ratern ausgewertet, wobei zwischen richtigen Antworten, morphologischen Varianten, bemerkten und nicht bemerkten Wiederholungen und Kategoriefehlern unterschieden wurde. Die hier berichteten Werte beziehen sich auf die Anzahl korrekter Antworten (vgl. Lindenberger et al., 1993).

### Wortanfang mit 'S'

Ähnlich wie im Test 'Tiere nennen' sollten die Teilnehmer in diesem Test in 90 Sekunden möglichst viele Wörter nennen. Die in diesem Test zu nennenden Wörter sollten mit dem Buchstaben 'S' beginnen. Die Auswertung der Ergebnisse erfolgte ansonsten analog zu dem Test 'Tiere nennen'.

### Wortschatz

16 Fragen aus dem HAWIE von Wechsler (1982) wurden in diesem Test verwendet. Auch hier wurden jede der Fragen in sehr großen Buchstaben auf dem Bildschirm dargeboten und gleichzeitig von der/dem FTA laut vorgelesen. Der Test wurde beendet, wenn der Proband (a) das letzte Item bearbeitet hatte oder (b) hintereinander fünf eindeutig falsche oder gar keine Antworten gab. Die Richtigkeit wurde von zwei unabhängigen Ratern mit falsch oder fehlend (0), teilweise richtig oder unvollständig (1) oder richtig (2) bewertet.

### Wörter finden

20 Items, die eine Auswahl aus dem Mehrfachwahl-Wortschatz-Test B von Lehrl (1977) darstellten, wurden nacheinander auf dem Bildschirm dargeboten. Dabei war ein Wort von sechs ein seltenes Wort und die restlichen fünf gut aussprechbare Nichtwörter (Distraktoren). Die Probanden sollten das seltene Wort mit dem Finger berühren. Der Test war entweder zu Ende, wenn drei falsche Antworten hintereinander gegeben oder das letzte Item des Test bearbeitet wurde (vgl. Lindenberger et al., 1993).

#### **4.2.1.2 Sensorische Leistungsfähigkeit**

##### Sehschärfe

Die Sehschärfe wurde in Snellen-Dezimale in zwei unterschiedlichen Entfernungen mit zwei unterschiedlichen Lesetafeln gemessen (Geigy, 1977). Der Fernvisus wurde für beide Augen zusammen (binokular) mit einer Snellen-Tafel in mindestens 2,5 m Entfernung bestimmt (vgl. Marsiske et al., 1999). Der Nahvisus wurde mit einer Snellen-Lesetafel bei individuell festgelegter Leseentfernung einzeln für das linke und rechte Auge gemessen. Alle Messungen wurden mit und ohne eigene korrigierende Fern- oder Lesebrillen durchgeführt. Die im folgenden berichteten Analysen beruhen auf den besseren Werten, was in den meisten Fällen die korrigierte Sehschärfe war. Als Maße für die Sehschärfe dienten sowohl der Nah- als auch der Fernvisus. Hierbei wurde der Nahvisus durch den Mittelwert der  $z$ -standardisierten Werte für den Nahvisus des rechten und des linken Auges bestimmt.

##### Gehör

Das Gehör wurde mit einem Bosch ST-20-1 Sinuston-Audiometer über Kopfhörer geprüft. Die Verwendung von Kopfhörern machte die Bestimmung von korrigiertem versus unkorrigiertem Gehör unmöglich, da Hörgeräte mit den Kopfhörern nicht bequem getragen werden konnten. Für vier Frequenzen wurden Hörschwellen in Dezibel in der folgenden Reihenfolge bestimmt: 1, 2, 4 und 6 kHz. Die Messung wurde an dem Ohr begonnen, das von dem Probanden als das "bessere" angegeben wurde. Die Lautstärke des Tons wurde solange erhöht, bis der Teilnehmer angab, den Ton zu hören (vgl. Lindenberger & Baltes, 1994). Als Maß für das Gehör wurde der Mittelwert über die vier Frequenzen des rechten und linken Ohres benutzt.

#### **4.2.1.3 Soziostrukturell-biographische Variablen**

##### Einkommen

Als Maß für das Einkommen wurde das nach der Anzahl der Haushaltsmitglieder gewichtete Haushalts-Netto-Einkommen pro Kopf herangezogen (Äquivalenzeinkommen; vgl. Motel & Wagner, 1993; Mayer & Wagner, 1996; Wagner, Motel, Spieß & Wagner, 1996). Da die Verteilung der Einkommenswerte den Anforderungen einer Normalverteilung nicht entsprach, wurden für die Analysen logarithmierte Werte verwendet.

##### Soziale Schicht

Die Teilnehmer wurden aufgrund ihrer beruflichen Stellung zu Schichten gruppiert. In Anlehnung an Mayer (1977) und Handl (1977) wurden sie dann fünf Klassen zugeordnet: der

Unterschicht, der unteren Mittelschicht, der mittleren Mittelschicht, der gehobenen Mittelschicht und der oberen Mittelschicht. Hierbei wurde der Ausdruck Oberschicht für die höchste Schicht vermieden, da die gesellschaftliche Elite (beispielsweise Großunternehmer, Top-Manager, Spitzenpolitiker) in der Berliner Altersstudie nicht hinreichend vertreten waren (vgl. Mayer & Wagner, 1996).

### Bildung

Bildung wurde hier gemessen als die Gesamtdauer allgemeiner und beruflicher Bildung in Jahren. Das heißt zusätzlich zu den Ausbildungsjahren in unterschiedlichen Schulsystemen (Hauptschule, Realschule, Gymnasium), wurden auch Lehr- oder Studienjahre hinzugerechnet (vgl. Mayer & Wagner, 1996).

#### **4.2.2 Fragen zur Erfassung der Anwendung der Methode der Orte**

Am Ende jeder Sitzung beantworteten die Versuchsleiter eine Reihe von Fragen, die der Erfassung individueller Unterschiede in der Anwendungshäufigkeit und -qualität der Methode der Orte dienten. Damit wurden zwei Ziele verfolgt. Erstens sollte die theoretische Annahme überprüft werden, ob individuelle Unterschiede in instruktionsbedingten Leistungszugewinnen mit individuellen Unterschieden in der Anwendungsintensität der Methode der Orte nach Instruktion verknüpft waren. Zweitens sollten die aus den begleitenden Fragen gewonnenen Informationen dazu dienen, unerwarteten oder widersprüchlichen Ergebnissen weiter nachgehen zu können.

Der Fragebogen zur Anwendungshäufigkeit und -qualität der Methode der Orte wurde jeweils am Ende einer Sitzung von den Versuchsleitern ausgefüllt. Da die erste Sitzung der Erfassung des Leistungsniveaus vor Einführung in die Gedächtnistechnik und die zweite Sitzung der ausführlichen Instruktion in der Methode der Orte dienten, stellten die Versuchsleiter die Fragen zum ersten Mal am Ende der dritten Sitzung (Prätest II). In der dritten und achten Sitzung (Prätest II und Posttest) wurden sowohl Selbsteinschätzungen der Probanden als auch Fremdeinschätzungen der Versuchsleiter erfaßt, für die vier Trainingssitzungen (vierte bis siebte Sitzung) hingegen lagen ausschließlich externe Beurteilungen von den Versuchsleitern vor.<sup>10</sup> Der Bogen bestand aus fünf Fragen (siehe Anhang C).

---

<sup>10</sup> Da die Sitzungsdauer 1.5 Stunden nicht überschreiten sollte und die Beantwortung der Fragen bei manchen hochbetagten Personen etwa 20 Minuten dauerte, mußte auf die Erfassung der Selbsteinschätzungen durch die Probanden in den Trainingssitzungen verzichtet werden.

Die erste Frage bezog sich auf die quantitative Erfassung der korrekten Nutzungshäufigkeit der Gedächtnistechnik in einer jeweiligen Sitzung, die Fragen 2 bis 5 auf die Erfassung qualitativer Aspekte bei der Anwendung der Methode der Orte.<sup>11</sup>

Da die erste Frage zur Erfassung der Anwendungshäufigkeit der Methode der Orte als Kontrollvariable für vorliegende Arbeit von Bedeutung war, sind die Meßeigenschaften dieses Items in Tabelle 6 abgebildet. Hierbei sind sowohl die Korrelationen zwischen Fremd- und Selbsteinschätzungen in Prätest II und Posttest (in Klammern) als auch die Stabilitäten der Fremdeinschätzungen über sechs Sitzungen hinweg dargestellt.

Tabelle 6  
Stabilitäten der Fremdbeurteilungen zur Nutzungshäufigkeit der Methode der Orte

	Sitzungen					
	3	4	5	6	7	8
3. Sitzung (Prätest II)	(.80)					
4. Sitzung (Training I)	.80					
5. Sitzung (Training II)	.84	.90				
6. Sitzung (Training III)	.75	.81	.81			
7. Sitzung (Training IV)	.76	.76	.82	.90		
8. Sitzung (Posttest)	.81	.80	.85	.85	.90	(.70)

*Anmerkung.* Alle Korrelationen sind von Null verschieden ( $p < .01$ ). Die Werte in Klammern stellen Korrelationen zwischen Fremd- und Selbstbeurteilungen innerhalb der Prätest II und Posttest Sitzungen dar.

Das Korrelationsmuster zeigt, daß die Rangordnung der Personen nach Instruktion in der Methode der Orte über den weiteren Trainingsverlauf stabil blieb: Die Korrelationen zwischen jeweils aufeinanderfolgenden Sitzungen bewegten sich zwischen  $r = .80$  und  $r = .90$ , die Korrelation zwischen der Sitzung unmittelbar nach Instruktion in der Methode der Orte (Prätest II) und der abschließenden Sitzung nach dem Training mit der Gedächtnistechnik

<sup>11</sup> In bezug auf die Beantwortung der ersten Frage wurden die Versuchsleiter während der Erhebungsphase in wöchentlichen Sitzungen dazu angehalten, sich bei ihrer Einschätzung weder auf die Güte des allgemeinen Gedächtnisses des jeweiligen Probanden noch auf die effektive Anwendung alternativer mnemonischer Strategien zu beziehen, sondern ausschließlich die Häufigkeit der *korrekten* Anwendung der *Methode der Orte* zu beurteilen. Daher beinhalten die Fremdbeurteilungen der ersten Frage nicht nur quantitative Informationen über die Nutzungshäufigkeit, sondern auch qualitative Einschätzungen über den Grad der korrekten Anwendung der Methode der Orte.

(Posttest) lag bei  $r = .81$ . Auch die Korrelationen zwischen Selbst- und Fremdeinschätzungen waren mit  $r = .80$  und  $r = .70$  zufriedenstellend.

### **4.2.3 Erfassung kognitiver Plastizität mit der Methode der Orte**

Die Erhebung der abhängigen Variablen, das heißt der episodischen Gedächtnisleistungen vor, während und nach einem Training mit der Methode der Orte, erstreckte sich über acht Sitzungen (für eine Übersicht siehe Abschnitt 3.2, Abb. 3). Da ein Hauptanliegen der Untersuchung die Ausdehnung der Fragen nach Ausmaß und Grenzen kognitiver Plastizität auf das hohe Alter war, lag der Schwerpunkt bei der Entwicklung des Studiendesigns in der Realisierung von Bedingungen, die eine optimale Erfassung der Gedächtnisleistungen von Personen zwischen 75 und 100+ Jahren erlauben. So wurde versucht, möglichen auditiven und visuellen Beeinträchtigungen im Alter durch eine bimodale Darbietung des Stimulusmaterials entgegenzuwirken. Die Auswirkung altersbedingter motorischer Beeinträchtigungen wurde dadurch reduziert, daß die Teilnehmer während der Abrufphase ihre Antworten verbal gaben.

Bei der Durchführung der Methode der Orte wurde weiterhin durch die Darbietung der Hinweisreize in der Einprägungs- und in der Abrufphase (cued recall) versucht, externale Hilfestellungen zu maximieren und gleichzeitig die Anforderungen an selbstinitiierte Prozesse zu minimieren. Dieses Vorgehen sollte das Erlernen und Durchführen der Methode der Orte erleichtern, um die Zahl der Sitzungen im Interesse der Studienteilnehmer gering zu halten. Es handelte sich demnach in der vorliegenden Studie um eine weniger komplexe Version der Methode der Orte, da die Versuchspersonen für das Generieren der interaktiven Vorstellungsbilder die Reihenfolge der Hinweisorte in der Einprägungs- und Abrufphase nicht erst selbstgenerieren mußten, sondern präsentiert bekamen.

#### **4.2.3.1 Geräte**

Zur Stimulusdarbietung und Erfassung der Antworten dienten fünf Apple Macintosh SE/30 Computer mit berührungsempfindlichen Bildschirmen (Bildschirmdiagonale = 9 Zoll). Die Benutzung dieser Computer hatte den Vorteil, daß sie klein und leicht genug waren, um von den Versuchsleitern in ihren Autos transportiert und in die Haushalte der Probanden getragen werden zu können.

### 4.2.3.2 Stimulusmaterial

#### Berliner Orte

Alle Teilnehmer erhielten die gleiche Sequenz von 32 bekannten Berliner Orten als Hinweisreize (siehe Anhang D3). Die Sequenz der Orte entsprach einer typischen Stadtrundfahrt durch Berlin. Sie begann bei einem bekannten Ort im ehemaligen Westteil Berlins (Olympiastadium), erstreckte sich bis in den früheren Ostteil der Stadt (Rotes Rathaus) und kehrte dann wieder in die Nähe des Anfangspunktes zurück (Funkturn). Um sicherzustellen, daß alle Teilnehmer klare Vorstellungsbilder zu jedem einzelnen der 32 Orte aufbauen konnten, wurde ein Bilderbuch angefertigt, in dem von jedem der 32 Orte sowohl eine Außenansicht als auch eine Innenansicht (wenn möglich) in Form einer Farbfotokopie abgebildet waren. Die Versuchsleiter prüften in der Instruktionsphase mit Hilfe des Bilderalbums die Bekanntheit der Orte durch Fragen nach persönlichen Erlebnissen der Teilnehmer an den jeweiligen Orten. Entstand der Eindruck, daß ein Ort nicht gut bekannt war, erarbeiteten die Versuchsleiter anhand der Abbildungen zusammen mit dem Teilnehmer eine bildliche Vorstellung des Ortes.

#### Wortlisten

Die Anzahl der präsentierten Wörter in einer Liste variierte in Abhängigkeit des Trainingszeitpunktes. Im Prätest I, Prätest II und im Posttest war die Listenlänge konstant; in jeder dieser drei Sitzungen wurden zwei 8-Wort- und zwei 16-Wort-Listen verabreicht. Hingegen veränderte sich die Länge der Wortlisten in der adaptiv angelegten Trainingsphase als Funktion der jeweiligen Erinnerungsleistung der Teilnehmer.

Als Stimulusmaterial wurden Wörter gewählt, die einen hohen Grad an Konkretheit und Bildhaftigkeit aufwiesen, in bezug auf ihre Länge vergleichbar und einfach verständlich waren. Die Wörter wurden aus einem Pool von 1387 Wörtern entnommen, die aus einem deutschen Wörterbuch von zwei unabhängigen Ratern danach ausgesucht worden waren, ob sie konkrete und gut vorstellbare Objekte bzw. Lebewesen darstellten.<sup>12</sup> In einem weiteren Schritt wählten wiederum zwei unabhängige Rater 763 Wörter aus, die die Grundlage für die Erstellung der Wortlisten vorliegender Studie bildeten. Bei der Auswahl dieser Wörter wurde zusätzlich zu den Kriterien 'Konkretheit' und 'hohe Bildhaftigkeit' darauf geachtet, lediglich Wörter einzu-

---

<sup>12</sup> Leider existierten nur für eine Teilmenge der Worte Bildhaftigkeits- und Konkretheitsurteile analog zu den Einschätzungen nach Paivio, Yuille und Madigan (1968; Baschek et al. 1977). Alle Worte dieser Teilmenge erhielten auf einer Skala von 1 bis 7 einen Bildhaftigkeitswert von über 5 (der höchste Wert lag bei 7).

schließen, die einfach verständlich, in ihrer Länge vergleichbar und der Kohorte der heute 75- bis über 100jährigen bekannt und in ihrer emotionalen Konnotation nicht vorbelastet waren (so wurden beispielsweise Wörter wie Bombe, Granate, Hure oder Neger ausgeschlossen). Die Zuordnung einzelner Wörter zu den Wortlisten erfolgte nach dem Zufall. In einem abschließenden Schritt wurden die Listen daraufhin untersucht, daß Wörter, die einer gemeinsamen Kategorie angehörten (Tier, Pflanze, Lebensmittel, Person, Beruf usw.) nicht dicht aufeinanderfolgten.

#### Darbietungsform

Die Darbietung des Stimulusmaterials (die Orte und die Wörter) erfolgte sowohl auditiv über Kopfhörer als auch visuell über den Computerbildschirm. Dies sollte die Auswirkungen visueller bzw. auditiver Beeinträchtigungen im Alter minimieren. Das auditive Stimulusmaterial wurde in einem Tonstudio auf digitale Tonträger aufgenommen und anschließend als Sound-Ressource (SND) in den Computer eingespielt. Die Orte und die Wörter wurden von zwei unterschiedlichen, professionellen Radiosprechern gesprochen, um die Unterscheidbarkeit zwischen Hinweisreiz und Stimuluswort akustisch zu unterstützen. Dafür wurden zwei männliche Stimmen gewählt, da tiefe Frequenzen für ältere Personen mit altersbedingten Höreinbußen besser zu verstehen sind. Das visuelle Stimulusmaterial war in fetter, schwarzer Schrift (Schriftgröße: 38 Punkt; Schrifttyp = Geneva) auf hellgrauem Bildschirmhintergrund auch für Personen mit schlechter Sehfähigkeit gut sichtbar.

Abbildung 4  
Darbietungsform der Wortpaare



Zunächst erschien in der Mitte der unteren Bildschirmhälfte der Hinweisreiz (der Ort). Nach 600 msec. folgte die akustische Darbietung des Ortes über Kopfhörer. Nach einer Sekunde Pause sahen die Teilnehmer das Stimuluswort auf der oberen Hälfte des Bildschirms über dem Ort erscheinen und 600 msec. später hörten sie das einzuprägende Wort über Kopfhörer. Ort und Wort blieben anschließend 10 Sekunden auf dem Bildschirm sichtbar. Nach Ausblenden des Wortpaares verstrichen noch 2 Sekunden bis erneut der nächste Ort auf dem Bildschirm erschien. Visuelle und auditive Darbietung erfolgten gestaffelt, um eine sequentielle Verarbeitung der präsentierten Informationen zu gewährleisten.

Das Darbietungsintervall von 12 Sekunden zwischen jeweils zwei Wortpaaren war über alle Listen der Trainingsstudie hinweg konstant. Die Wahl einer Dauer von 12 Sekunden beruhte auf Erfahrungswerten der Pilotuntersuchung sowie auf Ergebnissen früherer Gedächtnistrainingsstudien.<sup>13</sup>

Nach der Darbietung des letzten Ort-Wort-Paares einer Liste erschien auf dem Bildschirm eine Maske, durch die die Teilnehmer aufgefordert wurden, sich an so viele Wörter wie möglich zu erinnern (vgl. Anhang D2). Die Abrufphase erfolgte hierbei sofort im Anschluß an die Einprägungsphase und wurde von dem Versuchsleiter durch Tastendruck initiiert. In analoger Reihenfolge zu der Darbietung während der Einprägungsphase erschienen die Orte auf der unteren Bildschirmhälfte und die Teilnehmer sollten die dazugehörigen Wörter laut nennen. Die Antworten wurden von dem Versuchsleiter in den Computer eingegeben. Diese Form der Eingabe minimierte die Auswirkungen altersbedingter motorischer Beeinträchtigungen und reduzierte Fehlantworten aufgrund falscher Rechtschreibung.

---

<sup>13</sup> So berichteten Personen im hohen Alter im Rahmen der Pilotuntersuchung, daß ihnen ein Darbietungsintervall von 12 Sekunden als optimal und Zeiten über 12 Sekunden tendentiell als zu lang erschienen. Einer Metaanalyse von Stigsdotter (1994) zufolge lag die mittlere Darbietungsgeschwindigkeit bei 38 Gedächtnistrainingsstudien für ältere Erwachsene bei 10 Sekunden, das heißt unter dem gewählten Wert von 12 Sekunden. Weiterhin zeigten Ergebnisse aus der von Kliegl et al. (1989) durchgeführten Trainingsstudie mit der Methode der Orte, daß alle älteren Probanden in der Lage waren, bei Präsentationsraten von 20, 15 und 10 Sekunden zweimal hintereinander 15 oder mehr Wörter einer 30-Wort-Liste korrekt zu erinnern (eine Anzahl von 15 Wörtern entsprach dem Adaptationskriterium für das Herabsetzen der Präsentationszeiten in der nächstfolgenden Liste im Laufe einer adaptiven Trainingsphase). Hierbei ist zu berücksichtigen, daß es sich in dieser Studie zwar einerseits um jüngere Personen zwischen 65 und 80 Jahren handelte, andererseits jedoch um eine schwierigere Aufgabe, da die Probanden die Methode der Orte ohne externe Hinweisreize durchführen mußten und die präsentierten Wortlisten mit 30 Wörtern sehr viel länger waren.

### 4.3 Durchführung der Untersuchung

Die Untersuchung fand für die gesamte Stichprobe der Personen im hohen Alter bei den Probanden zu Hause in Einzelsitzungen statt. Die Testung der Stichprobe der jungen Erwachsenen hingegen erfolgte in Gruppensitzungen im Max-Planck-Institut für Bildungsforschung. Die Testung beider Altersgruppen erfolgte hiermit zwar nicht unter identischen, jedoch unter für die jeweilige Altersgruppe optimalen Lernbedingungen.<sup>14</sup>

Bei den älteren Probanden vereinbarten die Versuchsleiter die jeweiligen Sitzungstermine auf individueller Basis. Hierbei versuchten sie, sich bei der Terminplanung nach den Wünschen der Teilnehmer zu richten, da dies die Chancen erhöhte, daß diese sich die Tageszeit aussuchten, an der sie sich am leistungsfähigsten und motiviertesten fühlten. Ziel war es, die acht Sitzungen in einem regelmäßigen, wöchentlichen Abstand durchzuführen.<sup>15</sup> Die mittlere Untersuchungsdauer für die Stichprobe der 75- bis 101jährigen betrug 1.7 Monate ( $SD = .35$ ), das heißt im Mittel etwas weniger als zwei Monate. Die mittlere Dauer zwischen zwei jeweils aufeinanderfolgenden Sitzungen lag bei 7.2 Tagen und entsprach damit dem geplanten wöchentlichen Untersuchungsrythmus ( $[7.76 + 7.65 + 7.54 + 6.88 + 6.65 + 7.03 + 7.21] / 7 = 7.24$  Tage).<sup>16</sup>

Im Vergleich dazu betrug der mittlere Untersuchungszeitraum für die Stichprobe der jungen Erwachsenen lediglich 0.73 Monate ( $SD = .28$ ), das heißt etwa drei Wochen. Die mittlere Dauer zwischen zwei aufeinanderfolgenden Sitzungen lag bei 3.7 Tagen. Zum einen verkürzte sich der Untersuchungszeitraum bei den 20- bis 30jährigen dadurch, daß die Instruktionssitzung

<sup>14</sup> Zwar zeigte die Metaanalyse von Verhaeghen (1993), daß sich ein Gedächtnistraining in Gruppensitzungen vorteilhaft auf die Trainingsgewinne auswirkt. Bei der vorliegenden Stichprobe Hochbetagter kam das Gruppenformat jedoch aus mehreren Gründen nicht in Betracht. Erstens waren diese häufig aus gesundheitlichen Gründen nicht mehr in der Lage, ihren Wohnort zu verlassen. Zweitens treten bei Personen im hohen Alter häufig während der Untersuchung Verständigungs-, Konzentrations- oder andere unerwartete Schwierigkeiten auf, die von einem einzelnen Versuchsleiter in individuellen Einzelsitzungen besser bewältigt und abgefangen werden können als in Gruppensitzungen.

<sup>15</sup> Bei genauer Einhaltung des einwöchigen Untersuchungsabstands hätten die Teilnehmer die Studie nach zwei Monaten abschließen können. Wegen unvorhergesehener terminlicher Schwierigkeiten auf seiten der Probanden ließ es sich jedoch nicht vermeiden, daß sich der Erhebungszeitraum entweder verkürzte, wenn beispielsweise aufgrund von verfrühten Reiseplänen oder Familienbesuchen zwei Sitzungstermine in eine Woche gelegt werden mußten, oder aber verlängerte, wenn es zu überraschenden Krankenhausaufenthalten oder Krankheiten kam.

<sup>16</sup> Bei nicht vermeidbaren Terminverschiebungen wurde vor allem darauf geachtet, daß nicht mehr als 14 Tage zwischen der Instruktionssitzung und dem Prätest II bzw. zwischen der letzten Trainingssitzung und dem Posttest lagen. Dieses Ziel konnte hinsichtlich des Zeitintervalls zwischen zweiter und dritter Sitzung mit einer Ausnahme (die Teilnehmerin mußte aufgrund eines Rippenbruchs 48 Tage aussetzen) erreicht werden. Das maximal angesetzte Zeitintervall von zwei Wochen Untersuchungsabstand zwischen letzter Trainingssitzung und Posttest wurden lediglich von drei Probanden mit 15, 19 und 21 Tagen zwischen siebter und achter Sitzung nicht eingehalten.

noch am selben Tag im Anschluß an die Baseline-Erhebung (Prätest I) erfolgte. Da mögliche Intrusionsfehler bei den Übungsbeispielen in der Instruktionssitzung für die Ergebnisse der vorliegenden Studie nicht von Bedeutung waren, konnte hierdurch ein Sitzungstermin eingespart werden. Außerdem kamen aus organisatorischen und zeitlichen Gründen die jungen Probanden zweimal wöchentlich zu einer Sitzung. Auch wenn identische Testbedingungen für beide Altersgruppen damit nicht gegeben waren, begünstigt ein solcher Unterschied in den Sitzungsintervallen die Personen im hohen Alter, da bei kürzeren Intervallen zwischen zwei Sitzungen die Wahrscheinlichkeit für Intrusionsfehler steigt.

Im Mittel dauerte die reine Erfassung der Gedächtnisleistungen in der Stichprobe der Personen im hohen Alter pro Sitzung 36.5 Minuten ( $SD = 8.5$ ), wobei die mittleren Erhebungszeiten in den drei Testsitzungen mit 28.6 Minuten im Prätest I, 36.2 Minuten im Prätest II und 35.7 Minuten im Posttest kürzer waren als in den vier Trainingssitzungen (mit 45.2 Minuten in der ersten, 41.7 Minuten in der zweiten, 38.7 Minuten in der dritten und 39.0 Minuten in der letzten Trainingssitzung). Da zu Anfang und Ende der Sitzungen viel Zeit auf Einstimmung, verbale Rückmeldungen und Ausfüllen der begleitenden Fragen zur Erfassung individueller Unterschiede in der Anwendung der Methode der Orte verwendet wurde, lag die gesamte Sitzungsdauer zwischen einer und zwei Stunden, im Mittel bei 1.5 Stunden.<sup>17</sup>

In der Vergleichsstichprobe der jungen Erwachsenen lag die über alle acht Sitzungen hinweg gemittelte Erhebungsdauer der Gedächtnisleistungen mit 41.5 Minuten ( $SD = 4.1$ ) über der der 75- bis 101jährigen. Hierbei waren die mittleren Testzeiten der jungen Erwachsenen in den drei Testsitzungen mit 26.6 Minuten im Prätest I, 22.9 Minuten im Prätest II und 24.8 Minuten im Posttest kürzer als die der Personen im hohen Alter. In den vier Trainingssitzungen hingegen waren die Testzeiten erwartungsgemäß aufgrund besserer Leistungen und damit verbundenen längeren Wortlisten mit 51.6 Minuten in der ersten, 58.2 Minuten in der zweiten, 62.6 Minuten in der dritten und 71.3 Minuten in der letzten Trainingssitzung bedeutend länger.

Insgesamt erstreckte sich der Erhebungszeitraum der vorliegenden Studie über ein ganzes Jahr: Die Studie begann Anfang Januar 1998 und wurde Anfang Januar 1999 abgeschlossen. Sechs Versuchsleiter (Trainer) führten die Studie durch (fünf Frauen und ein Mann). Die Autorin der vorliegenden Arbeit zählte hierbei für die ganze Phase der Datenerhebung zu einer der fünf weiblichen Versuchsleiterinnen. Jeder Teilnehmer wurde über

---

<sup>17</sup> Dieser Wert entsprach sowohl dem im Rahmen der Metaanalyse von Stigsdotter (1994) gefundenen Wert einer mittleren Sitzungsdauer von 1.51 Stunden als auch dem von Verhaeghen (1993) berechneten Mittelwert von 1.53 Stunden.

die ganze Studie hinweg von dem gleichen Versuchsleiter begleitet. Vor Beginn der Hauptuntersuchung fand eine Pilotstudie statt, in der alle Versuchsleiter jede Sitzungsart mit jeweils einer älteren Person unter Beobachtung der anderen Versuchsleiter durchführte. Anschließende Analysen bestätigten die Effektivität dieses Vortrainings, da keine signifikanten Versuchsleitereffekte aufgezeigt werden konnten.

#### **4.3.1 Prätest I, Prätest II und Posttest**

Die erste Sitzung (Prätest I) diente der Baseline-Erhebung der Gedächtnisleistung vor Einführung in die Methode der Orte. Um die standardisierte Erfassung von Leistungsgewinnen zu ermöglichen, wurden die Gedächtnisleistungen unter identischen Bedingungen in der dritten Sitzung (Prätest II) nach Instruktion in der Methode der Orte und nochmals in der letzten Sitzung (Posttest) nach ausgiebigem Training mit der Gedächtnistechnik erfaßt.

Da aufgrund von Erfahrungswerten aus der Berliner Altersstudie bekannt war, daß Personen zwischen 75 und 100+ Jahren wegen Konzentrationsproblemen nicht mehr als 4 bis 6 Listen in einer Sitzung bearbeiten können, wurde in der vorliegenden Arbeit auf eine weitere Variation der Darbietungszeiten zur Untersuchung der Gedächtnisplastizität unter variierenden Schwierigkeitsbedingungen verzichtet. Die Manipulation der Schwierigkeit erfolgte in vorliegender Arbeit durch die Variation der Listenlänge (zwei 8-Wort- und zwei 16-Wort-Listen; vgl. Anhang D4). Die Erfassung der Gedächtnisleistung anhand unterschiedlich langer Wortlisten sollte zudem gewährleisten, daß bei schlechten Probanden im hohen Alter Bodeneffekte in der 16-Wort-Bedingung und bei guten Probanden Deckeneffekte in der 8-Wort-Bedingung ausgeglichen werden konnten. Das Auftreten von Deckeneffekten in der 16-Wort-Bedingung erschien aufgrund von Ergebnissen der Vorstudie mit Personen im hohen Alter auch nach einem Training mit der Methode der Orte unwahrscheinlich.

Ein im Rahmen altersvergleichender Studien generell auftretendes, für die vorliegende Studie jedoch besonders relevante Herausforderung, lag in der Konzipierung eines Studiendesigns, das gleichzeitig die optimale Erfassung der Gedächtnisleistungen von Personen im hohen Alter und junger Erwachsener zwischen 20 und 30 Jahren erlaubte. Hinsichtlich der drei Testsitzungen (Prätest I, II und Posttest) mußte erwartet werden, daß bei jungen Erwachsenen bereits im Prätest II und mit Sicherheit im Posttest, sowohl bei den 8-Wort- als auch bei den 16-Wort-Listen, Deckeneffekte auftreten würden, die sich nur durch das Hinzuziehen von zwei

weiteren, längeren, wie zum Beispiel zwei weiteren 30-Wort-Listen hätten ausgleichen lassen. Das Hinzunehmen von zwei zusätzlichen 30-Wort-Listen in den Testsitzungen wäre jedoch aus ethischen und motivationalen Gründen bei der Hauptstichprobe der 75- bis über 100jährigen nicht mehr vertretbar gewesen. Da das Hauptziel der Untersuchung in der Erfassung individueller Unterschiede in kognitiver Plastizität bei Personen im hohen Alter lag, wurde für folgende Lösung optiert: Während Prätest I, Prätest II und Posttest der optimalen Erfassung individueller Unterschiede im Ausmaß an kognitiver Plastizität in der Stichprobe der Personen im hohen Alter dienten, wurde für den Vergleich der Gedächtnisleistungen der 20- bis 30jährigen und der über 75- bis 100jährigen im Rahmen der *Trainingsphase* ein Design realisiert, das eine kontinuierliche Leistungsverbesserung für alle Teilnehmer und damit die Erfassung altersbedingter Unterschiede im Ausmaß an Gedächtnisplastizität ermöglichte.

Um Interferenzeffekte aufgrund häufiger Wiederholung derselben Hinweisreize innerhalb einer Sitzung zu minimieren, dienten als Hinweisreize 32 unterschiedliche Orte. So erhielten die Probanden als Hinweisreize in der ersten 8-Wort-Liste die Orte 1 bis 8, in der zweiten 8-Wort-Liste die Orte 9 bis 16, in der ersten 16-Wort-Liste die Orte 17 bis 32 und zuletzt in der zweiten 16-Wort-Liste erneut die Orte 1 bis 16.<sup>18</sup>

Der Mini Mental State Examination Test (MMSE) wurde am Ende der ersten Sitzung erhoben. Den Bogen mit den Fragen zur Erfassung der Anwendungshäufigkeit und -qualität der Methode der Orte füllten die Versuchsleiter jeweils am Ende der dritten und der letzten Sitzung (Prätest II und Posttest) aus. Hierbei gaben sie zunächst ihre eigenen Einschätzungen (Fremdeinschätzungen) ab und erst dann stellten sie den Probanden mündlich nochmals die Fragen (Selbsteinschätzungen) und notierten die Antworten schriftlich.

### 4.3.2 Instruktion in der Methode der Orte

Am Ende der ersten Sitzung nach der Baseline-Erhebung erfolgte der Beginn der Instruktion in der Methode der Orte nach Bower (1970). Zunächst informierten die Versuchsleiter die Teilnehmer über die historischen Ursprünge der Methode der Orte. Anhand anschaulicher Beispiele wurde anschließend die Funktionsweise der Methode der Orte erklärt (siehe Anhang D1). Zuletzt durften die Probanden selbst die Methode der Orte an einer Liste von drei Ort-

---

<sup>18</sup> Die Wiederholung der Orte 1 bis 16 in der letzten Liste war unvermeidbar, da es keine weiteren 16 Orte in Berlin gab, die einen genügend hohen Bekanntheitsgrad besaßen.

Wort-Paaren üben. Die drei Wortpaare ermöglichten dem Versuchsleiter, die Prinzipien der Methode der Orte nochmals konkret an speziell zu diesem Zwecke ausgewählten Beispielen zu verdeutlichen. Bei der Ausführung dieser Übungsliste gab es keine zeitlichen Restriktionen.<sup>19</sup>

Die zweite Sitzung, die eigentliche Instruktionssitzung, begann mit einer Wiederholung der Hauptprinzipien der Methode der Orte. Anschließend erfolgte die Einführung der 32 Berliner Orte anhand des Bilderalbums. Bei jeder Abbildung wurden die Probanden gebeten, die Orte namentlich zu identifizieren und ihre persönlichen Erfahrungen mit diesen Orten detailliert zu schildern. Mit wenigen Ausnahmen waren den Studienteilnehmern alle 32 Orte gut bekannt: Die meisten Studienteilnehmer im hohen Alter hatten sogar noch zu vielen Orten lebendige Erinnerungen aus den 1920er und 1930er Jahren. Am Ende der Instruktionssitzung übten die Teilnehmer die Methode der Orte an vier Listen mit jeweils sechs Ort-Wort-Paaren. Das Präsentationsintervall konnte in den ersten zwei Übungslisten von den Teilnehmern selbst durch Tastendruck bestimmt werden (self-paced). Im Gegensatz dazu war in den zwei letzten Listen ein 12-Sekunden-Intervall vorgegeben, um die Teilnehmer an den Zeittakt der Untersuchung zu gewöhnen. Während der vier Übungslisten brachten die Versuchsleiter den Teilnehmern außerdem bei, ihre mentalen Bilder und Gedanken während der Einprägungsphase laut zu verbalisieren. Das laute Verbalisieren beim Einprägen der Wörter sollte die Konzentration der Probanden beim Einprägen fördern. Darüber hinaus ermöglichte das laute Verbalisieren dem Versuchsleiter, die verwendeten Strategien der Probanden während der Einprägungsphase einzuschätzen und dadurch (a) im Laufe der Trainingsphase gezielte Rückmeldungen und Verbesserungsvorschläge zu geben und (b) die Fragen zur Erfassung individueller Unterschiede in der Nutzungshäufigkeit und -qualität zu beantworten.

---

<sup>19</sup> Als positive Begleiterscheinung hatte die abschließende Durchführung der Beispielsliste einen stark motivierenden Effekt, da vor allem leistungsschwache Probanden, die aufgrund schlechter Erinnerungsleistungen in der Baseline-Erhebung frustriert worden waren, die Sitzung mit einem Erfolgserlebnis abschließen konnten.

### 4.3.3 Adaptive Trainingsphase

Die adaptive Trainingsphase umfaßte vier Sitzungen (Sitzungen 4–7). Das Trainingsdesign wurde mit dem Ziel entworfen, vier unterschiedlichen Anforderungen gerecht zu werden: Es sollte

- (a) den Teilnehmern ermöglichen, die Anwendung der Methode der Orte unter guten motivationalen Lernbedingungen zu üben und zu verbessern;
- (b) die standardisierte Erfassung des Potentials zur Leistungsoptimierung während eines Trainings erlauben;
- (c) die standardisierte Erfassung von Leistungsgrenzen bei einem Training mit der Methode der Orte gewährleisten (Erfassung des asymptotischen Leistungsniveaus); sowie
- (d) die Erfassung altersbedingter Unterschiede im Ausmaß an kognitiver Plastizität bei jungen Erwachsenen und Personen im hohen Alter ermöglichen.

In jeder der vier Trainingssitzungen übten die Probanden die Methode der Orte an sechs Wortlisten, so daß die gesamte Trainingsphase 24 Wortlisten umfaßte. Während dieser vier Trainingssitzungen wurde die Listenlänge adaptiv an das jeweilige Leistungsniveau der Probanden angepaßt, das heißt die Anzahl an präsentierten Wörtern in einer Liste richtete sich jeweils nach der vorausgegangenen Leistung des jeweiligen Studienteilnehmers. Die Form eines adaptiven Trainings wurde gewählt, um für Personen im hohen Alter eine motivierende und leistungsfördernde Lernumgebung zu schaffen. Zur weiteren Unterstützung gab der Versuchsleiter jeweils nach Beendigung einer Wortliste Verbesserungsvorschläge und Anregungen in bezug auf eine effektivere Anwendung der Methode der Orte. Hierbei wurden die Versuchsleiter instruiert, nicht während der Einprägungs- und Abrufphase einer jeweiligen Wortliste zu intervenieren, um auch während der Trainingsphase eine standardisierte Erfassung der Gedächtnisleistungen zu gewährleisten.

Die Leistungsadaptation erfolgte in vorliegender Studie auch während der Trainingsphase nicht durch die Variation der Darbietungsgeschwindigkeit sondern durch die Variation der Listenlänge. Die Anzahl der Wörter in der ersten Liste der ersten Trainingssitzung orientierte sich hierbei an der maximalen Anzahl erinnerter Wörter im Prätest II. Bei fehlerloser Leistung, das heißt nur wenn alle Wörter der jeweiligen Liste korrekt wiedergegeben werden konnten, wurde die darauffolgende Liste um ein Wort verlängert. Ein solches adaptives Trainingsdesign

bot im Idealfall die Möglichkeit, die individuelle, im Prätest II erreichte Bestleistung über 24 Listen hinweg stetig um ein Wort zu verbessern. Einem solchen Design zufolge hätten Probanden in der letzten Liste der vierten Trainingssitzung idealerweise bei optimaler Leistung im Prätest II (16 von 16 Wörter in einer der 16-Wort-Listen) und fehlerloser Leistungen über alle 24 Listen der Trainingsphase hinweg zum Ende der Trainingsphase eine Maximalleistung von 36 korrekt erinnerten Wörtern erreichen können. Das Anfangsniveau zu Beginn einer jeweiligen Trainingssitzung richtete sich nach der Bestleistung in der vorangegangenen Sitzung und nicht nach der in der letzten Liste erreichten Anzahl richtig erinnerter Wörter. Hiermit sollte vermieden werden, daß eventuelle, zum Beispiel auf Interferenzeffekten oder Ermüdungserscheinungen beruhende Leistungseinbußen zum Ende einer Trainingssitzung einen Einfluß auf die Bestimmung des Ausgangsniveaus der folgenden Sitzung haben konnten. Ziel der Trainingsphase war es, so effizient wie möglich die latenten Potentiale der Probanden erfassen zu können. Aus diesen Gründen wurde die Liste auch nur dann um ein Wort verkürzt, wenn in zwei aufeinanderfolgenden Listen weniger als 75 % der Wörter korrekt erinnert werden konnten.

Wie in den Testsitzungen (Prätest I, II und Posttest) sollten auch in der Trainingsphase Interferenzeffekte, die durch die wiederholte Verwendung der gleichen Orte bei wechselnden Stimulusworten entstehen, reduziert werden. Daher wurde nach Beendigung einer Liste nicht wie in früheren Trainingsstudien mit der Methode der Orte zu Anfang der nächsten Liste zu Ort 1 der Touristenroute zurückgekehrt, sondern mit dem Ort weitergemacht, der auf dem Rundweg durch Berlin der jeweils folgende war.<sup>20</sup> Auf diese Weise wiederholten sich die Orte erst nach der Darbietung von 32 Wörtern. Diese Methode verhinderte weiterhin, daß die leistungsstärkeren Probanden Orte durchliefen, die die leistungsschwächeren im Laufe des Trainings nie zu sehen bekamen, jedoch in der letzten Sitzung, dem Posttest als Hinweisreize erhielten.<sup>21</sup>

---

<sup>20</sup> Da es sich um eine kreisförmige Route durch Berlin handelte war der Übergang von Ort 32 zu Orte 1 qualitativ vergleichbar mit jedem anderen Übergang zwischen zwei Orten auf dem Rundweg.

<sup>21</sup> Der Nachteil eines solchen Trainingsdesigns lag darin, daß einerseits leistungsstärkere Probanden aufgrund längerer Listen mehr Übungsmöglichkeiten erhielten, da sie im Laufe der Trainingsphase eine größere Anzahl interaktiver, mentaler Vorstellungsbilder generierten, andererseits die Wahrscheinlichkeit für Interferenzeffekte bei diesen Probanden aus dem gleichen Grunde stieg, zumal sie dieselben Orte häufiger durchliefen. Eine Lösung dieses Problems hätte darin liegen können, in jeder Trainingssitzung bei variierender Anzahl von Listen eine konstante Anzahl von Ort-Wort-Paaren vorzugeben. Um jedoch sowohl bei den 75- bis über 100jährigen als auch bei den 20- bis 30jährigen die Erfassung maximaler Leistungsfähigkeit in jeder Trainingssitzung zu gewährleisten und damit altersbedingte Vergleiche im Ausmaß an kognitiver Plastizität zu ermöglichen, hätte jeder Studienteilnehmer in der ersten Sitzung 111 Worte (16 + 17 + 18 + 19 + 20 + 21), in der zweiten Sitzung bereits 141 (21 + 22 + 23 + 24 + 25 + 26), in der dritten 171 (26 + 27 + 28 + 29 + 30 + 31) und in der letzten Trainingssitzung 201 (31 + 32 + 33 + 34 + 35 + 36)

## 4.4 Datenanalyse

### 4.4.1 Erfassung individueller Unterschiede in kognitiver Plastizität

Die Datenauswertung der vorliegenden Arbeit gliedert sich in zwei Teile, einem deskriptiven und einem korrelativen Teil. Im ersten Teil steht die Erfassung kognitiver Plastizität auf Gruppenebene im Vordergrund, im zweiten die Vorhersage individueller Unterschiede im Ausmaß an beobachtbarer Gedächtnisplastizität. Es stellt sich somit die Frage, wie Plastizität sowohl auf der Ebene von Gruppen als auch auf der Ebene von Individuen optimal gemessen werden kann.

Aus meßtheoretisch-statistischer Perspektive ist die Erfassung von Plastizität auf Gruppenebene unproblematisch: Der mittlere Leistungszuwachs sowie die mittlere Gedächtnisleistung im Posttest sind effiziente und erwartungstreue Maße für den Leistungszuwachs und die maximale Leistungsfähigkeit auf Gruppenebene. Dasselbe gilt auf individueller Ebene für die Gedächtnisleistungen im Posttest. Im Gegensatz dazu hat die Verwendung von einfachen Differenzen zwischen Prä- und Posttestleistungen als Maß zur Erfassung von intraindividuellen Leistungsveränderungen in der Literatur zur Veränderungsmessung starke Kontroversen ausgelöst.

#### Die Kontroverse um einfache Differenzmaße zur Erfassung von Leistungsveränderungen

Kritiker haben vor allem zwei Probleme bei der Verwendung von einfachen Differenzmaßen herausgestellt, die sich (a) auf die Validität und Reliabilität dieser Veränderungsmaße und (b) auf die Korrelation zwischen Ausgangsleistung und Leistungszuwachs beziehen (für Übersichten siehe Cronbach & Furby, 1970; Embretson, 1991; Nunnally, 1982; Rogosa, Brandt & Zimowski, 1982).

Demnach gibt es erstens ein Reliabilitätsparadox in Differenzwerten. Wenn die Korrelation zwischen Prä- und Posttestleistungen hoch ist, ist die interindividuelle Varianz der Differenzwerte gezwungenermaßen klein und reduziert deren Reliabilität (Collins, 1991). Wenn die Korrelation zwischen Prä- und Posttestwerten hingegen gering ist, steigt zwar die Reliabilität des Differenzmaßes, doch beklagen einige Autoren, daß dann ein Validitätsproblem vorliegen

---

Wörter erinnern müssen. Ein solches Design hätte bei leistungsschwachen Probanden, das heißt bei der Mehrheit der Personen zwischen 75- und 100+ Jahren zum Abbruch der Studie geführt.

mag. Denn es ist nicht klar, was durch die Differenzwerte gemessen wird, da Prä- und Posttest möglicherweise unterschiedliche Dinge erfassen (Bereiter, 1963).

Zusätzlich zu dem Problem niedriger Reliabilitäten wird häufig ein weiterer Einwand gegen Differenzwerte aufgeführt, der sich auf die statistisch problematische Korrelation zwischen der Ausgangsleistung im Prätest und der Differenz zwischen den Post- und Prätestleistungen bezieht (Bohrnstedt, 1969; Linn & Slinde, 1977; O'Connor, 1972; vgl. Rogosa et al., 1982). Aufgrund der Tatsache, daß Ausgangs- und Differenzwert dieselben Meßfehler aufweisen, jedoch mit entgegengesetzten Vorzeichen, hat die beobachtete Korrelation zwischen Ausgangsleistung und Differenzwert einen negativen Bias.

Eng mit dieser Problematik verbunden ist das statistische Problem der Regression zur Mitte, die immer dann zu beobachten ist, wenn Messungen nicht völlig reliabel sind. Solchen Regressionseffekten zufolge fallen hohe Leistungen bei der zweiten Messung geringer und umgekehrt niedrige Leistungen höher aus. Das Resultat ist gleichermaßen eine negative Korrelation zwischen Ausgangsleistung und Differenzmaß (Bohrnstedt, 1969; Furby, 1973; Marcus, 1979; vgl. Rogosa et al., 1982).

Gleichermaßen häufig finden sich in der Literatur zur Veränderungsmessung Positionen, die das einfache Differenzmaß als solches verteidigen und die oben aufgeführten Kritikpunkte widerlegen oder relativieren (Collins, 1996; Maris, 1998; Nunnally, 1982; Rogosa et al., 1982; Williams & Zimmerman, 1996). So argumentieren beispielsweise Rogosa et al. (1982) gegen die Auffassung, daß geringe Stabilität zwischen Prä- und Posttestmessung notwendigerweise die Interpretation des Differenzwertes als Veränderungsmaß ausschließt, wenn sie schreiben: "Whether or not the variable being studied retains its meaning over time cannot be resolved by the value of the Time1-Time2 correlation." (S. 732) Weiterhin zeigen die Autoren auf, daß Differenzwerte hoch reliabel sein können, wenn die jeweiligen Statusmaße reliabel und genügend individuelle Unterschiede in den Leistungsveränderungen vorhanden sind. Wie auch andere Autoren machen sie außerdem darauf aufmerksam, daß die negative Verzerrung der beobachteten Korrelation zwischen Ausgangsleistung und Differenzwert keinen generell gültigen Grund gegen die Verwendung des einfachen Differenzmaßes darstellt. Im Gegenteil verteidigen sie das einfache Differenzmaß als direkte Schätzung der wahren Veränderung von Personen über die Zeit. Die gleiche Position vertretend schreibt beispielsweise Nunnally (1982): "gain scores are what they are when we find them, and they are at the very heart of the study of effects in developmental investigations" (Nunnally, 1982, S. 138).

Als Reaktion auf die Kritik an einfachen Differenzmaßen kam es zur Entwicklung alternativer Methoden zur Erfassung von Leistungsveränderungen, wie beispielsweise dem residualisierten Veränderungsmaß (residual change score). Hierbei werden die Posttestleistungen durch die Ausgangsleistungen im Prätest vorhergesagt und anschließend Differenzen zwischen den beobachteten und vorhergesagten Posttestwerten gebildet (Cronbach & Furby, 1970; Embretson, 1987; O'Connor, 1972). Dieses residualisierte Veränderungsmaß ist *notwendigerweise* mit der Ausgangsleistung unkorreliert; man erhält somit zwei statistisch orthogonale Maße für Ausgangsniveau und Veränderung.

Residualisierte Veränderungsmaße dürfen jedoch nicht als ein meßfehlerbereinigter Ersatz für den einfachen Differenzwert interpretiert werden, da sie keine Antwort auf die einfache Frage geben, wie sehr sich eine Person hinsichtlich eines Merkmals über die Zeit verändert hat. Das residualisierte Veränderungsmaß bezieht sich vielmehr auf die Diskrepanz zwischen der individuellen und der für die Stichprobe erwarteten Veränderung. In den Worten von Cronbach und Furby (1970) ist das residualisierte Veränderungsmaß "primarily a way of singling out individuals who changed more (or less) than expected" (S. 74). An anderer Stelle räumen die Autoren selbstkritisch ein, daß dieses alternative Veränderungsmaß nicht als statistisch überlegenes Maß der wahren Veränderung einer Person angesehen werden dürfe, da "in most studies the portion discarded includes some genuine and important change in the person" (Cronbach & Furby, 1970, S. 74).

#### Die Erfassung individueller Unterschiede in kognitiver Plastizität in vorliegender Studie

Aus mehreren Gründen schien in vorliegender Studie die Verwendung einfacher Differenzmaße zur Erfassung individueller Unterschiede im Leistungszuwachs angemessener als die Verwendung alternativer Maße, wie zum Beispiel residualisierter Veränderungsmaße.

- (1) Das Heranziehen einfacher Differenzmaße hatte den Vorteil, daß die Erfassung von Plastizität als Leistungszuwachs über beide Teile der Arbeit hinweg mit dem gleichen Maß erfolgte; das heißt sowohl in dem deskriptiven Teil der Arbeit, in dem die Quantifizierung des mittleren Leistungszuwachses auf Gruppenebene im Vordergrund stand, als auch in dem zweiten, dem korrelativen Teil der Arbeit, der zum Ziel die Vorhersage individueller Unterschiede in kognitiver Plastizität hatte.
- (2) Da im Rahmen der Hypothesenüberprüfung die Interpretation des Zusammenhangs zwischen dem Ausgangsniveau im Prätest und der Differenz zwischen Prätest- und

Posttestleistungen nicht erforderlich war, verlor die mit dieser statistisch problematischen Korrelation verbundene Kritik an dem einfachen Differenzmaß an Bedeutung.

- (3) Die Bedenken, die sich auf die mangelnde Reliabilität und Validität einfacher Differenzwerte beziehen, können ausgeräumt werden, wenn gezeigt werden kann, (a) daß die Statusmaße per se hoch reliabel sind und (b) daß es trotz mittlerer oder hoher Stabilität genügend individuelle Unterschiede in den Leistungsveränderungen gibt (siehe Abschnitt 4.4.1).
- (4) Auch kann argumentiert werden, daß das Heranziehen einfacher Differenzmaße zur Erfassung individueller Leistungsveränderungen im Falle eingeschränkter Reliabilitäten der Differenzwerte die konservativste Form der Hypothesentestung darstellt, da signifikante Zusammenhänge mit einfachen Differenzmaßen dann schwieriger aufzuzeigen wären.
- (5) Das Argument, daß die Validität der Veränderungsmessungen bei niedriger Korrelation zwischen erster und zweiter Messung angezweifelt werden muß, da sich in diesem Falle die Frage stellt, ob Prä- und Posttest nicht unterschiedliche Dinge erfassen, war im Kontext vorliegender Untersuchung zumindest in bezug auf den Leistungszuwachs zwischen Prätest I und II (die Sitzung vor und nach Instruktion in der Methode der Orte) ein Bestandteil des theoretischen Rahmens: Es war gerade das Ziel dargestellter Untersuchung, unter experimenteller Kontrolle Probanden in einer zuvor unbekanntem Gedächtnistechnik zu instruieren und damit im Prätest II eine Veränderung der Herangehensweise an die Gedächtnisaufgabe zu bewirken. Es interessierte demnach vor allem die Erfassung individueller Unterschiede im Leistungszuwachs, unter der *a priori* festgesetzten Annahme, daß im Prätest I und II unterschiedliche Dinge gemessen werden; das heißt eine Gedächtnisleistung ohne und eine mit der Nutzung der Methode der Orte.

Nach dem folgenden Überblick über die in der vorliegenden Arbeit verwendeten Plastizitätsmaße werden die Reliabilitäten und Stabilitäten der Gedächtnisleistungen vor und nach Instruktion in und Training mit der Methode der Orte dargestellt.

#### 4.4.2 Operationalisierung kognitiver Plastizität im Überblick

Zusammenfassend wird Plastizität in vorliegender Studie durch mehrere Maße operationalisiert, die Plastizität entweder als *Leistungszuwachs* oder als *maximale Leistungsfähigkeit* erfassen (siehe Abschnitt 2.3.2.3).

Plastizität als maximale Leistungsfähigkeit wurde sowohl auf der Ebene von Gruppen als auch von Individuen durch die Gedächtnisleistung erfaßt, die Personen nach einem ausgiebigem Training mit der Methode der Orte in der letzten Sitzung, dem *Posttest* erzielen konnten. Das Maß für Plastizität als Leistungszuwachs bestand auf der Ebene von Gruppen im mittleren Leistungszuwachs, auf der Ebene von Individuen in der Differenz zwischen den Gedächtnisleistungen einer Person zu jeweils zwei Meßzeitpunkten. Dabei wurde unterschieden zwischen dem Leistungszuwachs, der auf die Instruktion in der Methode der Orte zurückzuführen war (*Zuwachs I* = Prätest II – Prätest I) und dem anschließenden Leistungszuwachs, der auf das vier Sitzungen umfassende adaptive Training mit der Methode der Orte beruhte (*Zuwachs II* = Posttest – Prätest II).

Da ein Vergleich der Gedächtnisleistungen zwischen den Stichproben der 20- bis 30jährigen und der 75- bis über 100jährigen aufgrund von antizipierten Deckeneffekten in Prä- und Posttest auf seiten der jungen Probanden nicht intendiert war, dienten zur Quantifizierung altersbedingter Unterschiede in kognitiver Plastizität die Leistungsverläufe in den vier adaptiven Trainingssitzungen (vgl. Hypothese 2). Weiterhin hatte die Analyse der Leistungsverbesserungen während der adaptiven Trainingsphase in der Stichprobe der 75- bis über 100jährigen die Funktion, die Optimierungshypothese (Hypothese 1b) näher zu spezifizieren. Die Leistungsveränderungen während der vier Trainingssitzungen stellten daher ein weiteres Maß zur Erfassung kognitiver Plastizität dar.

#### 4.4.3 Reliabilitäten und Stabilitäten der Gedächtnisleistungen

Wie im vorhergehenden Abschnitt deutlich wurde, kann ein Problem bei der Verwendung von einfachen Differenzmaßen zur Erfassung individueller Unterschiede im Ausmaß an Leistungszuwachs in deren mangelnden Reliabilitäten liegen. Aus dargestellten Gründen ist daher die Frage nach Reliabilität und Stabilität der Messungen von zentraler Bedeutung.

Tabelle 7 enthält die Interkorrelationen der Gedächtnisleistungen in den sieben Sitzungen der Trainingsstudie (die Instruktionssitzung ist nicht aufgeführt) sowie die Reliabilitätskoeffizienten der Messungen (in Klammern) getrennt für die Stichproben der 75- bis 101jährigen und der 20- bis 30jährigen.

Tabelle 7

Stabilitäten und Reliabilitäten der Gedächtnisleistungen in Prätest I, II und Posttest sowie in den vier Trainingssitzungen getrennt für die Personen im hohen Alter und die jungen Erwachsenen

	Sitzungen						
	1	3	4	5	6	7	8
75- bis 101jährige (N = 96)							
1. Sitzung (Prätest I)	(.89)						
3. Sitzung (Prätest II)	.65**	(.96)					
4. Sitzung (Training I)	.68**	.93**	(.98)				
5. Sitzung (Training II)	.66**	.89**	.96**	(.98)			
6. Sitzung (Training III)	.63**	.90**	.96**	.98**	(.99)		
7. Sitzung (Training IV)	.61**	.88**	.94**	.97**	.98**	(.99)	
8. Sitzung (Posttest)	.60**	.90**	.88**	.86**	.85**	.85**	(.97)
20- bis 30jährige (N = 20)							
1. Sitzung (Prätest I)	(.90)						
3. Sitzung (Prätest II)	.39	(.78)					
4. Sitzung (Training I)	.48*	.76**	(.97)				
5. Sitzung (Training II)	.52*	.73**	.92**	(.97)			
6. Sitzung (Training III)	.43	.66**	.89**	.91**	(.99)		
7. Sitzung (Training IV)	.28	.61**	.79**	.85**	.92**	(.96)	
8. Sitzung (Posttest)	.25	.26	.53*	.58**	.69**	.67**	(.63)

*Anmerkung.* \*  $p < .05$ ; \*\*  $p < .01$ . Die Werte in Klammern stellen Spearman-Brown Split-half (odd-even) Reliabilitätskoeffizienten dar. Die kursiv gekennzeichneten Werte zeigen die Korrelationen an, die aufgrund vorhandener Deckeneffekte bei den jungen Probanden in Prä- und Posttest nicht interpretiert werden sollten.

Die Werte in Tabelle 7 machen deutlich, daß in der Stichprobe der Personen im hohen Alter alle Spearman-Brown Reliabilitätskoeffizienten hoch waren ( $> r = .89$ ). In der Stichprobe

der jungen Erwachsenen lagen die Reliabilitätskoeffizienten nur für die Messungen der Gedächtnisleistungen im Prätest II ( $r = .78$ ) und Posttest ( $r = .63$ ) etwas niedriger. Dies dürfte auf Deckeneffekte zurückzuführen sein.<sup>22</sup> Hierbei können die niedrigen Koeffizienten in diesen beiden Sitzungen jedoch vernachlässigt werden, da die Gedächtnisleistungen der jungen Erwachsenen in Prä- und Posttest in keiner der hier berichteten statistischen Analysen Verwendung finden.

Die Korrelationsmuster zeigen weiterhin auf, daß die Stabilitätsmaße insgesamt zwar sehr hoch waren, aber auch temporalen und altersbezogenen Veränderungen unterlagen, wie dies zu erwarten war. In der Stichprobe der Personen im hohen Alter standen die Ausgangsleistungen im Prätest I signifikant mit den Gedächtnisleistungen nach Instruktion in der Methode der Orte im Prätest II ( $r = .65$ ) in Zusammenhang. Es fällt zudem auf, daß dieser Zusammenhang nicht so hoch war, wie die nach Instruktion in der Methode der Orte zu beobachtenden Korrelationen zwischen aufeinanderfolgenden Sitzungen, die sich zwischen  $r = .85$  und  $r = .98$  bewegten. Es scheint daher, daß die nach Instruktion in der Methode der Orte erreichte Rangordnung der Probanden über die Trainingsstudie hinweg relativ stabil blieb. Der positive, jedoch niedrigere Zusammenhang zwischen Prätest I und II hingegen zeigt an, daß durch die Instruktion in der Methode der Orte noch eine Neuordnung der Personen hinsichtlich ihrer Gedächtnisleistungen im Prätest II stattfand.

Die Befunde für die jüngere Altersgruppe sind in der Größenordnung unterschiedlich. Wie bereits dargestellt muß jedoch hierbei beachtet werden, daß die Zusammenhänge zu den Gedächtnisleistungen in Prätest II und Posttest aufgrund von deutlichen Deckeneffekten *nicht* interpretiert werden sollten (siehe kursiv markierte Korrelationen in Tab. 7). Wie bei den Hochbetagten zeigen die hohen Korrelationen zwischen den Gedächtnisleistungen in den vier Trainingssitzungen (zwischen  $r = .79$  und  $r = .92$ ) bei den jungen Versuchsteilnehmern, daß die nach Instruktion in der Methode der Orte erreichte Rangordnung der Probanden über die Trainingsphase hinweg relativ stabil blieb. Werden hingegen die Höhen der Stabilitätskoeffizienten der Korrelationen zwischen Ausgangsleistung (Prätest I) und der vier Trainings-

<sup>22</sup> Von den jungen Erwachsenen erzielten im Prätest I drei junge Personen (15 %) die maximale Leistung (12 Wörter) und sechs Personen (30 %) erreichten eine Gedächtnisleistung, die entweder genau oder sogar über 11.50 Wörtern lag. Im Prätest II erreichten dann bereits 8 Personen (40 %) die maximale Leistung und sechs (30 %) erzielten einen Mittelwert von oder über 11.50 Wörtern. Im Posttest erinnerten sich 65 % im Mittel an 12 Wörter und keiner der Personen lag unter einem Mittelwert von 10 Wörtern. Die Mittelwerte und SD der Stichprobe junger Erwachsener für die einzelnen Listen in Prätest I, II und Posttest sind im Anhang E.1, Tabelle E.1.1 dargestellt.

sitzungen zwischen beiden Altersgruppen verglichen, wird deutlich, daß (a) diese bei den jüngeren Erwachsenen deutlich geringer sind ( $r = .48, p < .05$ ;  $r = .52, p < .05$ ;  $r = .43, r = .28$ ) und (b) die Höhe des Zusammenhangs zwischen Ausgangsleistung und der Gedächtnisleistungen *nach* Einführung in der Methode der Orte tendenziell im Laufe des Trainings abnimmt. Dies entspricht den theoretischen Erwartungen, nach denen das Einführen und Üben der Methode der Orte bei jüngeren Erwachsenen zu einer Veränderung der Ensembles kognitiver Komponenten führt, die der Ausführung der Gedächtnisaufgabe vor und nach Erlernen der Methode der Orte zugrunde liegen. Die höheren und relativ uniformen Korrelationen in der Stichprobe der Hochbetagten weisen im Gegensatz dazu darauf hin, daß der Ausführung der Gedächtnisaufgaben vor und nach Einführung der Methode der Orte ein vergleichbares kognitives System zugrunde lag. Solche Befunde stehen im Einklang mit der Dedifferenzierungshypothese der Fähigkeitsstruktur im hohen Alter (vgl. Abschnitt 2.1.3.1)

Die bereits im vorangegangenen Abschnitt diskutierten Bedenken bei der Verwendung einfacher Differenzmaße zur Erfassung individueller Unterschiede im Leistungszuwachs innerhalb der Stichprobe der Hochbetagten können aufgrund der dargestellten Ergebnisse ausgeräumt werden. Für die Hochbetagten waren die Messungen der Gedächtnisleistungen alle ausreichend reliabel. Dies gilt auch für die zwei zentralen Differenzmaße, dem Leistungszuwachs zwischen Prätest I und II und zwischen Prätest II und Posttest. Die Reliabilität für den Zuwachs zwischen Prätest I und II (Zuwachs I) lag bei  $r = .68$ , die für den Zuwachs zwischen Prätest II und Posttest (Zuwachs II) bei  $r = .57$ .<sup>23</sup> Es schien demnach (zumindest zwischen Prätest I und II) bei gleichzeitig hoher Stabilität genügend individuelle Unterschiede in den Leistungsveränderungen zu geben, um hohe Reliabilitäten der Differenzmaße zu gewährleisten.

---

<sup>23</sup> Die Reliabilitätskoeffizienten stellen Korrelationen zwischen jeweils zwei Differenzmaßen dar: Das erste Differenzmaß bezieht sich auf die Veränderung der mittleren Leistungen in den ersten und dritten Listen, das zweite Differenzmaß auf die Veränderung in den zweiten und vierten Listen einer jeweiligen Sitzung.