Aging Neuropsychology and Cognition
2004, Vol. 11, No. 2–3, pp. 104–133

Psychology Press
Taylor & Francis Group

Cognition in the Berlin Aging Study (BASE): The First 10 Years

Martin Lövdén[1], Paolo Ghisletta[2,3], and Ulman Lindenberger[1]
[1]School of Psychology, Saarland University, Saarbrücken, Germany, and Max Planck Institute
for Human Development, Berlin, [2]Center for Interdisciplinary Gerontology, University of Geneva, Switzerland,
and [3]Center for Life Course and Life Style Study, Universities of Lausanne and Geneva, Switzerland

ABSTRACT

This paper summarizes and expands research on cognitive aging from the Berlin Aging Study (BASE), a longitudinal, multidisciplinary, and population-based investigation of old and very old individuals. First, we describe previously reported research on five key themes: (a) experimental and mortality-associated components of longitudinal selectivity; (b) comparisons between cross-sectional and cross-sectional/longitudinal convergence age gradients; (c) old-age dedifferentiation of inter-individual differences; (d) possible reasons for the age-based increase in the link between intellectual and sensory domains; and (e) limits to cognitive plasticity in very old age. Second, we make use of multilevel modeling to determine the magnitude and direction of retest effects. Retest effects are classified as either flat (step function from the first to the second measurement occasion) or growing (linear increase after the first measurement occasion). Five of the eight longitudinally administered cognitive tests are found to display significant retest effects of either or both types. Retest adjustment increased the linear negative and decreased the quadratic negative component of cross-sectional/longitudinal convergence gradients in measures of intellectual abilities.

The Berlin Aging Study (BASE; Baltes, Mayer, Helmchen, & Steinhagen-Thiessen, 1999) is a multidisciplinary, population-based, longitudinal investigation of old and very old residents of former West Berlin. It includes information collected by four different research units: internal medicine/geriatrics, psychiatry, psychology, and sociology/economics (for overviews of the initial design, see Baltes & Mayer, 1999; Baltes et al., 1999).

This article summarizes and expands findings from BASE about intellectual functioning in old and very old age. It is organized in four sections. First, we describe general methodological features of BASE, emphasizing aspects of the design that are pertinent to the present focus on intellectual development in old and very old age. Second, we summarize previously reported cross-sectional and longitudinal findings. Third, we report novel evidence about the magnitude and shape of retest effects for each of the eight cognitive tests that have been longitudinally administered in BASE. Fourth, we end this article with suggestions for future research directions.

GENERAL OVERVIEW

As is true for other large-scale multidisciplinary inquiries into human aging, the general design

---

Address correspondence to: Ulman Lindenberger, Center for Lifespan Development, Max Planck Institute for Human Development, Lentzeallée 94, D-14195 Berlin, Germany. E-mail: seklindenberger@mpib-berlin.mpg.de

10.1080/13825580490510982$16.00 © Taylor & Francis Ltd.

features and data collection protocols of BASE are complex. In the following, we restrict our description to those design characteristics that are most relevant for the present focus on intellectual development; for a general overview of BASE, see Baltes and Mayer (1999), Baltes et al. (1999); for an in-depth description of the sampling scheme and initial sample selectivity, see Lindenberger et al., 1999.

Participants and Data Collection Protocols

The first in-depth measurement occasion of BASE, termed T1 Intensive Protocol, consisted of a 14-session multidisciplinary assessment with 516 participants aged 70–103 years (mean age = 84.9 years, *SD* = 8.7). This sample of 516 individuals, which served as the starting point for most cross-sectional and longitudinal analyses, was stratified by age and gender, with 43 women and 43 men in each of six different age groups (70–74, 75–79, 80–84, 85–89, 90–94, and 95+ years). Stratification served two inter-related purposes: (a) to produce equally reliable estimates of population parameters across age groups and gender; and (b) to improve the chances of detecting interactions of chronological age or gender with other variables.

The T1 Intensive Protocol sample of 516 individuals originated from random draws of addresses from the city registry of former West Berlin. Given the relatively high degree of accuracy of registry information (e.g., in Germany, all residents have to register), the random draws of addresses from the city registry assigned equal probabilities to all members in the population to participate in the study. Selectivity analyses revealed that the 516 individuals who were willing and able to complete the T1 Intensive Protocol were positively selected on a broad range of variables (Lindenberger et al., 1999). Owing to the recruitment procedure in BASE, the transitive nature of the samples (i.e., later samples are selections of earlier samples), and the application of specialized statistical tools for the analysis of sample selectivity (cf. Aitkin, 1934; Pearson, 1903), it was possible to estimate mean-level and variance/covariance selectivity for a wide range of variables measured before and after selection. Specifically, prior to the T1 Intensive Protocol, 1264 out of the 1908 initially contacted individuals participated in a short initial assessment, and 928 individuals participated in a comprehensive 1-session intake assessment. Thus, increasing amounts of information were available for samples of decreasing size, and all the available information, including demographic information obtained for the initial city registry sample ($n = 1908$), was used to quantify selection effects.

With respect to means and prevalence rates, the T1 Intensive Protocol sample ($n = 516$) turned out to be positively selected on all analyzed domains and measures. Specifically, compared to the initial city registry sample, the T1 Intensive Protocol sample showed lower mortality risk, lower dementia prevalence, better somatic health, higher scores on Activities of Daily Living (ADL), higher levels of functioning in intellectual, sensory, and motor domains, larger social network size, and presumably more adaptive scores on various self-report measures of personality and self-regulation. The magnitude of these selectivity effects was largest for general intelligence. However, with the exception of dementia prevalence in very old age, all effects sizes were well below 0.5 standard deviations of the parent sample. Furthermore, structural associations among variables (i.e., variances and covariances), which figure prominently in the theoretical rationale and analyses of BASE, were only marginally influenced by selectivity. Thus, the T1 Intensive Protocol sample provided a relatively solid starting point for the investigation of subsequent longitudinal changes, including mortality-associated and experimental selectivity (see below).

Data regarding the T1 Intensive Protocol were collected between mid-1990 and June 1993. Currently, the longitudinal design of BASE consists of five measurement occasions. Data collection and data entry for the fifth occasion have recently been completed but will not be considered in the present article. The first, third (T3), fourth (T4), and fifth (T5) measurement occasions included a reduced version of the T1 Intake Assessment and the T1 Intensive Protocol. The second session (T2) was limited to the Intake Assessment protocol. Table 1 displays the longitudinal design up to T4, with special emphasis on the scheduling of

Table 1. The BASE Assessment Schedule of Cognitive and Sensory Variables.

	Measurement occasion						
	IA_{T1}	IPr_{T1}	IA_{T2}	IA_{T3}	IPr_{T3}	IA_{T4}	IPr_{T4}
Ability							
Perceptual speed	DL	DL, IP	DL	DL	DL, IP	DL	DL, IP
Episodic memory	PA, MT			PA, MT			PA, MT
Fluency		CA, WB	CA	CA	CA, WB	CA	CA, WB
Verbal knowledge		VO, SW			VO, SW		VO, SW
Vision		CV, DV	CV, DV		CV, DV		CV, DV
Hearing		H	H		H		H
Mean time in study	0.00	0.13	1.95	3.76	3.99	5.55	6.03
n	516	516	361	244	208[a]	164	132

Note. IA = Intake Assessment; IPr = Intensive Protocol; DL = Digit Letter; IP = Identical Pictures; PA = Paired Associates; MT = Memory for Text; CA = Categories; WB = Word Beginnings; VO = Vocabulary; SW = Spot-a-Word; CV = close visual acuity; DV = distant visual acuity; H = hearing. T1, T2, T3, and T4 denote the first, second, third, and fourth measurement occasion, respectively. Mean time in study is expressed in years.

[a] For entire Intensive Protocol, $n = 206$.

the various cognitive and sensory measures. Relevant variables, described in detail below, and effective sample sizes are specified for each measurement occasion. Table 2 summarizes a select set of characteristics for the T1 cross-sectional sample ($n = 516$) as well as for the current T1 to T4 longitudinal sample ($n = 132$). Baltes and Mayer (1999) can be consulted for more detailed participant characteristics that include, for example, in depth physical and psychiatric health characteristics.

Longitudinal Psychometric Battery: Rationale and Measures

Theoretical Rationale

The selection and analysis of cognitive measures in BASE has been informed by two-component views of lifespan cognition. Generally, two-component models assume that intellectual development reflects the operation of two interacting influences, one biological and the other cultural (for a summary, see Lindenberger, 2001). Contemporary examples are the theory of fluid and crystallized intelligence (Gf/Gc theory; Cattell, 1971; Horn, 1982, 1989) and the distinction between the mechanics and the pragmatics of cognition (Baltes, 1987; Baltes, Lindenberger, & Staudinger, 1998). The biological component represents fundamental organizational properties of the central nervous system. In terms of psychological operations, these properties are assumed to be indexed by the speed, accuracy, and coordination of elementary processing operations as assessed in tasks measuring the quality of information input, discrimination, categorization, and selective attention, as well as reasoning ability in highly overlearned or novel domains. As a consequence, the biological component is assumed to dominate inter-individual differences in intellectual abilities such as Gf narrowly defined (i.e., reasoning), perceptual speed, and episodic memory. We refer to this component as the broad fluid domain (cf. Horn, 1989), or the mechanics of cognition (Baltes, 1987).

The cultural component, on the other hand, refers to the acquisition and expression of culturally transmitted bodies of declarative and procedural knowledge that are made available to individuals in the course of socialization. Psychometrically, this component pervades individual differences in intellectual abilities such as verbal knowledge (e.g., semantic memory), specialized expertise (e.g., Ackerman, 1996), or knowledge about the fundamental pragmatics of life (e.g., wisdom; cf. Baltes et al., 1998). Henceforth, we

Table 2. Sample Characteristics for the Total Cross-Sectional Sample ($n = 516$) and for the Longitudinal Sample ($n = 132$) at First (T1), Third (T3), and Fourth (T4) Measurement Occasions.

Variables	Age	Women (%)	Education	Intelligence	Perceptual speed	Episodic memory	Fluency	Verbal knowledge
Total cross-sectional sample at T1 ($n = 516$)								
M (*SD*)	84.9 (8.7)	50	10.8 (2.3)	50.0 (10.0)	50.0 (10.0)	50.0 (10.0)	50.0 (10.0)	50.0 (10.0)
r with age				−.52	−.56	−.43	−.45	−.29
Longitudinal sample at T1 ($n = 132$)								
M (*SD*)	78.3 (5.9)	55	11.3 (2.4)	57.3 (7.8)	57.0 (7.3)	55.6 (9.7)	57.2 (9.0)	54.3 (8.2)
r with age				−.23	−.29	−.22	−.16	.00
Longitudinal sample at T3 ($n = 132$)								
M (*SD*)	82.0 (6.0)	55	11.3 (2.4)	55.8 (8.9)	55.8 (8.4)	53.6 (9.8)	55.6 (9.7)	53.5 (7.5)
r with age				−.28	−.41	−.21	−.24	−.12
Longitudinal sample at T4 ($n = 132$)								
M (*SD*)	83.8 (5.9)	55	11.3 (2.4)	55.1 (9.5)	54.5 (9.4)	54.0 (9.7)	55.5 (10.2)	54.4 (8.5)
r with age				−.36	−.39	−.24	−.26	−.22

Note. Perceptual speed, episodic memory, fluency, and verbal knowledge are unit-weighted composites of two tests, scaled as *T* scores ($M = 50$, $SD = 10$), with the cross-sectional sample at T1 as reference. Intelligence is the unit-weighted composite of the four intellectual ability composites.

refer to the cultural component as the broad crystallized domain (cf. Horn, 1989), or the pragmatics of cognition (Baltes, 1987).

Measures

The original (full) version of the battery, which was used at T1 only, consists of a total of 14 tests assessing perceptual speed (three tests), reasoning (three tests), episodic memory (three tests), verbal knowledge (three tests), and fluency (two tests). Due to constraints on testing time, the tests of reasoning were dropped from the battery at later occasions, and the number of tests for perceptual speed, episodic memory, and verbal knowledge was reduced from three to two. Thus, in the longitudinal version of the battery, perceptual speed and episodic memory represent the mechanics, whereas verbal knowledge and fluency primarily represent the pragmatics. Below we restrict our descriptions to the eight tests used in the longitudinal battery; readers may consult Lindenberger, Mayr, and Kliegl (1993) for a detailed description of the six tests not included in the follow-up assessments. A Macintosh SE30 computer equipped with a Micro Touch Systems touch-sensitive screen assisted cognitive testing, which took place at the residence of the participant. The entire session was tape-recorded. Unless specified differently, general information about psychometric properties of the tests such as reliability are based on T1 data (Lindenberger & Baltes, 1997; Lindenberger & Reischies, 1999).

Perceptual Speed

Two tests, Digit Letter and Identical Pictures, served as indicators of perceptual speed. The Digit Letter test resembles the well-known Digit Symbol Substitution of the WAIS, with the exception that participants had to name letters instead of writing symbols. The main reason for this change was to minimize motor task demands. A template with digit-letter pairings was visible for the entire testing period. The test consists of a total of 21 sheets, each sheet containing six digits with a question mark underneath. Participants named the letters corresponding to the digits, by moving from left to right. The dependent measure is normally the number of correct responses after 3 min. The reliability of this test, based on correlations of the scores obtained for the three 1 min segments of the test, is .96 (Cronbach's α; see Lindenberger & Reischies, 1999, for details).

The Identical Pictures test is a computerized and modified version of the corresponding test from the ETS (Ekstrom et al., 1976). A total of 32 items were presented. For each item, a target figure was presented in the upper half of the screen, and five response alternatives were presented in the lower half. Participants touched the figure in the lower half of the screen that corresponded to the target figure. The main dependent measure is the number of correct responses within 80 s. Again, the reliability of this measure is satisfactory ($\alpha = .90$).

Episodic Memory

This ability factor was measured by Memory for Text and Paired Associates. In Memory for Text, a short story was both presented on the screen and read aloud by the research assistant. A test in a cued-recall format followed immediately thereafter. Participants answered six questions regarding the content of the story, each question referring to propositions either at high, intermediate, or low levels of text hierarchy. The number of correctly answered questions served as a measure of performance. Reliability for this score is within the normal range for episodic memory measures ($\alpha = .57$).

In the Paired Associates test, eight pairs of concrete nouns were presented twice at a rate of 5 s per pair. After each of the two presentations, the first noun of each pair was presented as a recall cue. Scores generally refer to the total number of correct responses across the two lists. Reliability for this measure is satisfactory ($\alpha = .87$).

Verbal Knowledge

For both tests of verbal knowledge, Spot-a-Word and Vocabulary, presented items were ordered by difficulty. Testing was terminated when the participants made three consecutive false responses in the Spot-a-Word test, and when they made five such responses in the Vocabulary test. For the Spot-a-Word test, 20 items containing one word and four pronounceable non-words were presented successively on the screen. Participants were asked to select and touch the word. Three practice items were provided. Testing time was unlimited. The standard dependent measure is the number of correct responses. Reliability of this measure is satisfactory ($\alpha = .92$).

For the Vocabulary test, 20 words were selected from the Vocabulary subtest of the German version of the WAIS (HAWIE; Wechsler, 1982). Words were presented one by one on the screen. Participants' answers were coded by two independent raters using a refined version of the instructions provided by Wechsler (1982). Each response received a score of 0 (wrong), 1 (partially correct), or 2 (correct). Testing time was unlimited. The measure of performance corresponds to the sum of the scores over all the 20 items. Cronbach's α for this score was .82, and inter-coder reliability was high ($r = .96$).

Fluency

In the Categories test, participants were asked to name as many different animals as possible within 90 seconds. Responses were classified by two independent as (a) correct responses, (b) morphological variations, (c) repetitions, or (d) wrong categories. Generally, performance refers to the number of correct responses. Inter-coder reliability for this measure was high ($r = .99$).

In the Word Beginnings test (Letter "S"), participants were asked to name as many different words as possible that begin with the letter "S" within 90 s. Again, responses were classified by two independent as (a) correct responses, (b) morphological variations, (c) repetitions, or (d) wrong categories. Inter-coder reliability was high ($r = .99$).

Other Measures

In addition to the psychometric tests of intellectual abilities, a number of other constructs have figured prominently in the previous reports from BASE related to intellectual functioning. Most of these constructs refer to sensory, sensorimotor, and life history domains.

Hearing

Auditory acuity was measured separately for the right and the left ears using a Bosch ST-201 pure-tone audiometer using headphones. Thresholds were measured separately for the right and left ears at four different frequencies (1.00, 2.00, 4.00, & 6.00 kHz). For technical reasons, thresholds were assessed without hearing aids only. An inverted average score of thresholds in dB across both ears over the four frequencies is normally used as an estimate of auditory acuity.

Vision

Visual acuity was measured with standard optometric procedures, with and without the best optical correction provided by the participant (i.e., glasses or contact lenses). Snellen reading charts were presented at about 25 cm distance from the participants' eyes to assess close vision, and at 2.5 m to assess distance vision. Close visual acuity was assessed for both eyes separately, while distant visual acuity was measured binocularly only. In all the reports from BASE summarized in this paper, scores refer to the best values obtained, be it with or without correction.

Socio-Biographical Variables (cf. Mayer, Maas, & Wagner, 1999)

Four variables were used to represent current socio-biographical status as well as socio-biographical life history: (a) Income, defined as the amount of net income per month on a five point scale; (b) occupational prestige, based on a standard sociological measure for Germany, referring to the prestige of the participants' last occupation before retirement (or to the last occupation of the spouse if the participant had never been part of the labor force); (c) social class, arranged on a continuum of social stratification, ranging from lower class (7% of the total T1 sample), lower middle class (20%), middle middle class (31%), upper middle class (31%), and higher middle class (11%), and (d) number of years in formal education.

COGNITION IN BASE: FINDINGS ON KEY RESEARCH THEMES

In the following, we selectively summarize ten years of cognitive research in BASE (i.e., 1993–2003). Specifically, we focus on five inter-related research themes in lifespan psychology (Baltes et al., 1998; Baltes & Labouvie, 1973) that have played a major role in investigations of intellectual functioning in BASE: (a) experimental and mortality-associated components of longitudinal selectivity; (b) comparisons between cross-sectional and cross-sectional/longitudinal convergence age gradients; (c) old-age dedifferentiation of inter-individual differences; (d) possible reasons for the age-based increase in the link between intellectual and sensory domains; and (e) limits to cognitive plasticity in very old age.

Longitudinal Selectivity: Separating Mortality-Associated and Experimental Components

Individuals who participate in longitudinal studies for longer periods of time tend to be younger, healthier (e.g., McArdle, Hamagami, Elias, & Robbins, 1991), of higher cognitive functioning (e.g., Zelinski & Burnight, 1997), and of a higher social class (e.g., Powers & Bultena, 1972) as compared to those who participate for shorter periods of time. At the most general level, this nonrandom, or selective, attrition can be partitioned into two additive components, mortality-associated and experimental (cf. Baltes & Labouvie, 1973). Mortality-associated selectivity denotes differences on relevant characteristics between individuals who do not return because they are deceased and individuals who are still alive. In contrast, experimental selectivity denotes differences between individuals who are willing and able to continue participation and those individuals who are alive but unwilling or unable to do so.

To examine 3.7-year selectivity in BASE, Lindenberger, Singer, and Baltes (2002) compared the T1 sample ($n = 516$) with the T3 sample ($n = 206$) on a wide range of variables covering demographic, sensory/sensorimotor, life history, and intellectual domains. With respect to mean levels, mortality-associated selectivity was computed as follows: $(M_{\text{survivors}} - M_{\text{full sample}})/SD_{\text{full sample}}$, where $M_{\text{survivors}}$ is the mean of a designated variable for individuals who are alive at T3, $M_{\text{full sample}}$ is the mean of the total (original) sample, and $SD_{\text{full sample}}$ is its standard deviation. The magnitude of mean level experimental selectivity was computed in a similar fashion: $(M_{\text{select}} - M_{\text{survivors}})/SD_{\text{full sample}}$, where M_{select} is the mean for individuals measured at a given occasion, $M_{\text{survivors}}$ is the mean for individuals still alive at the same occasion, and $SD_{\text{full sample}}$ is the standard deviation of the original sample. Total selectivity, that is, the extent to which individuals assessed at a given occasion differ from the T1 parent sample ($n = 516$) from which they originated, is equal to the sum of mortality-associated and experimental selectivity, or $(M_{\text{select}} - M_{\text{full sample}})/SD_{\text{full sample}}$. Thus, the formulae express mean differences between the reduced sample of interest and the total sample in an effect-size metric. To examine selectivity effects on variables measured after selection, the Pearson-Lawley selection formulae were used (Lawley, 1943; Pearson, 1903). By means of linear regression, and given the tenability of a number of statistical assumptions such as homoscedasticity and normality, these formulae project selectivity effects on variables measured *after* selection (for details, see Lindenberger et al., 1999, 2002).

Selectivity results are displayed in Figure 1. Across all variables, the mean (zero-order) magnitude of total selectivity was 0.28 *SD*-units. The averaged mortality-associated component (grey bars; 0.18 *SD*) was more pronounced than the experimental component (open bars; 0.10 *SD*). Age-partialed effect sizes (right panel in Fig. 1) were considerably smaller than the corresponding zero-order effects (left panel in Fig. 1). On average, a decrease from 0.18 to 0.08 was observed for the mortality-associated selectivity component and a reduction from 0.10 to 0.05 was seen for the experimental selectivity component. The reduction in the mortality-associated component after controlling for age was more pronounced than the corresponding reduction for the experimental component. Thus, selectivity effects are to a great extent age-linked, especially for mortality-related selectivity, but dropout also tends to be selective among individuals of the same age. Furthermore, as suggested by a closer inspection of Figure 1, selectivity effects were most pronounced for chronological age, intellectual functioning, sensory functioning, and ADL. Finally, mortality-associated and experimental selectivity effects were correlated across measures ($r = .76$;

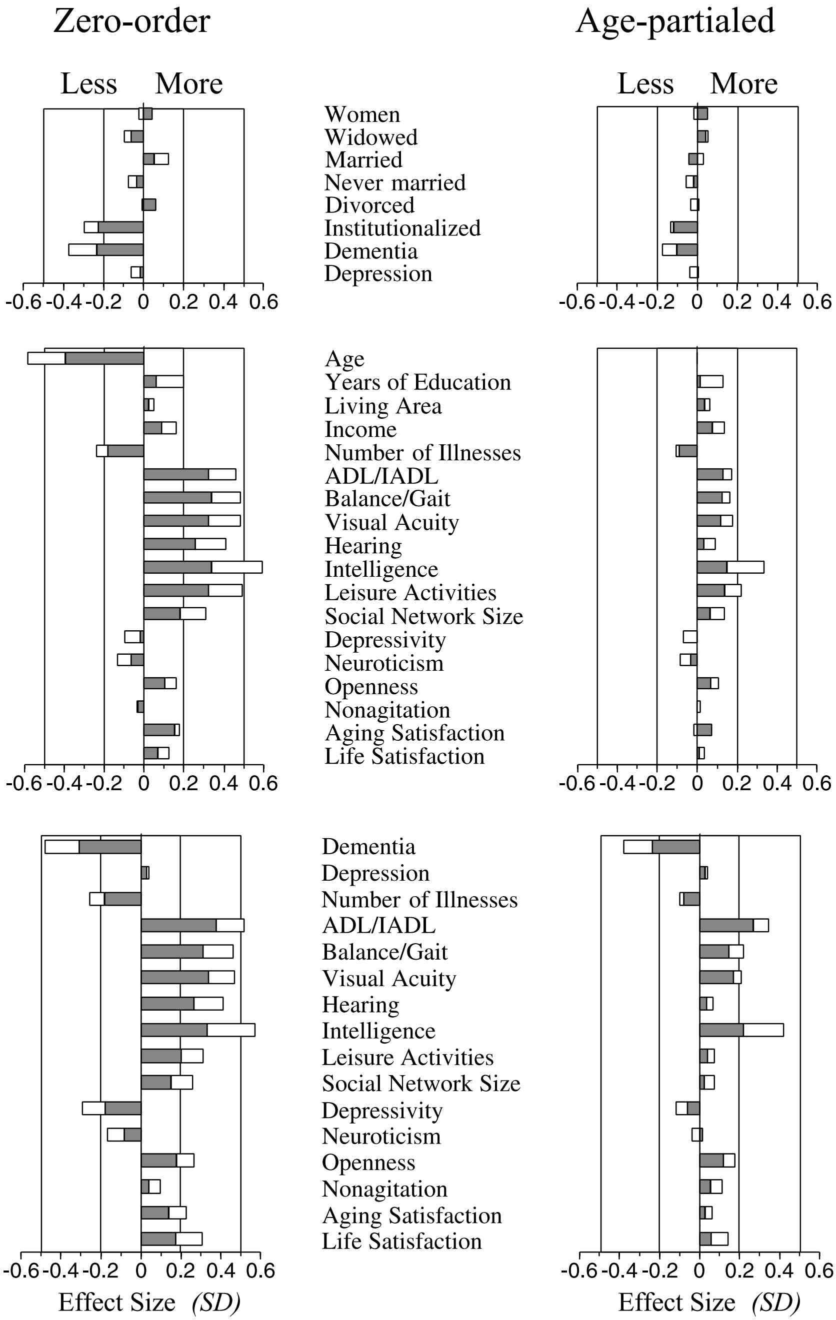

Fig. 1. Mortality-associated and experimental components of selectivity in the BASE T3 sample ($n = 206$) relative to the T1 parent sample ($n = 516$) for zero-order and age-partialed variables assessed at T1 and T3. In all panels, grey bars represent mortality-associated components, and open bars represent experimental components of selectivity. Top panels: Selectivity effects on dichotomous variables assessed at T1. Middle panels: Selectivity effects on continuous variables assessed at T1. Bottom panels: Selectivity effects on variables assessed at T1. Values for T3 variables were estimated using the Pearson-Lawley selection formulae. Throughout, left panels display zero-order effects, and right panels display age-partialed effects. Adapted from Lindenberger, Singer, and Baltes (2002).

after controlling for age, $r = .43$). As argued by Lindenberger et al. (2002), experimental selectivity may be seen as a "precursor" of mortality-associated selectivity in the limited sense that some of the causes underlying experimental and mortality-associated selectivity are identical (for more information on predictors of mortality in BASE, see also Maier & Smith, 1999).

Selection did not only affect means but also variances and covariances. In the results reported by Lindenberger et al. (2002), 26 of the 32 continuous variables depicted in Figure 1 were less variable in the longitudinal T3 sample than in the full T1 sample. Variance decrements tended to be greatest for variables with pronounced mean level selectivity (i.e., age, intelligence, sensory variables, and ADL). Moreover, almost all associations between intelligence, sensory functioning, and chronological age were attenuated, for both mortality-associated and experimental reasons. Hence, these results suggest that restricting longitudinal analyses to individuals observed on all measurement occasions may greatly reduce the magnitude of associations among variables. This tendency appears to be most pronounced for variables that are strongly associated with chronological age.

Singer, Verhaeghen, Ghisletta, Lindenberger, and Baltes (2003b) extended selectivity analyses to T4 data. Experimental and mortality-associated selectivity components were computed by comparing the T1 data of the T4 longitudinal sample ($n = 132$) with those individuals still alive at T4 ($n = 239$), and by comparing the survivors with those of the full cross-sectional T1 sample ($n = 516$), using the methods described above. Not surprisingly, positive total selection effects were found on initial level for the three sets of variables examined; that is, for intelligence (0.74 *SD*-units; 56% due to mortality), for sensory functioning (0.69 *SD*; 70% due to mortality), and for socio-biographical status (0.27 *SD*; 30% due to mortality). Two age groups were distinguished, a younger half called "old" ($n = 66$, *mean age at T1* = 73 years), and an older half called "old–old" ($n = 66$, *mean age at T1* = 83 years). As predicted, the mortality-associated component of selectivity was again larger in the old–old (mean 0.36 *SD*) than in the old sample (mean 0.05 *SD*). The experimental selectivity component was also higher in the old–old sample: Effect sizes ranged from 0.04 to 0.17 in the old and from 0.20 to 0.43 in the old–old. More importantly, among individuals who had survived up to T3, decrements on perceptual speed and verbal knowledge between T1 to T3 were greater for individuals who died before T4 than for individuals who survived and participated in the six-year follow-up.

To conclude, the results from BASE clearly show that mortality-associated and experimental components of selectivity modulate longitudinal observations on cognition in old and very old age. If longitudinal analyses are restricted to individuals observed at all measurement occasions, the results obtained cannot be generalized beyond a positively selected subset of the aging population. Experimental selectivity poses a serious validity threat, and may affect means, variances, and covariances of most or all cognition-related constructs under study. Mortality-related and experimental selectivity fundamentally differ in ontological and methodological status, even though the sources contributing to both types of selectivity may be highly similar. Taken together, both components of selectivity point to marked heterogeneity in development (Lövdén & Lindenberger, in press), especially in very old age (cf. Bäckman, Small, Wahlin, & Larsson, 2000). Their existence also confirms the need to better disentangle aging-related changes in old and very old age from changes associated with impending death (Kleemeier, 1962; Riegel & Riegel, 1972). At a more practical level, the selectivity analyses performed in BASE underscore the need to continuously update mortality-related information in longitudinal panel studies of cognitive aging phenomena.

Cross-Sectional and Longitudinal Age Gradients in Old and Very Old Age

The cross-sectional empirical pattern of monotonic decline across adulthood for the fluid mechanics (e.g., working memory and processing speed) accompanied by stability or increases in the crystallized pragmatics (e.g., verbal knowledge) constitutes the "classic aging pattern" (Botwinick, 1977) of adult intellectual development (see e.g., Hunt, 1949; Jones & Conrad, 1933; Nilsson et al., 1997; Park et al., 2002;

Schaie, 1994, 1996; Wechsler, 1955). Beyond this pattern, the available cross-sectional evidence is mixed with respect to the onset and amount of decline in the crystallized pragmatics within old to very old age (see Bäckman et al., 2000 for review; cf. Salthouse, 2003). Whereas some studies report relatively small age differences during this period (e.g., Christensen, 2001; Nyberg et al., 2003; Park et al., 2002), others report age-related differences starting at about age 50 (Bäckman & Nilsson, 1996; Wechsler, 1997). With respect to BASE, lifespan extensions revealed negative cross-sectional associations between verbal knowledge and age within, but not before, old and very old age (e.g., Baltes & Lindenberger, 1997; Lindenberger & Baltes, 1994, 1997).

For example, Baltes and Lindenberger (1997) administered the cross-sectional (i.e., T1) cognitive battery to a younger sample ($n = 171$; mean age $= 48.2$ years, $SD = 14.7$ years, range = 25–69 years). The two abilities from the pragmatic domain, verbal knowledge and fluency, showed no significant relations with age in the young group ($r = .05$ for verbal knowledge and $r = -.13$ for fluency) but significant negative correlations in the total T1 BASE sample ($r = -.41$ for verbal knowledge and $r = -.46$ for fluency). In light of these cross-sectional findings, and given the selectivity effects discussed above, Singer et al. (2003b) used latent growth curve modeling (cf. McArdle et al., 1991) to compare cross-sectional and longitudinal age gradients under three different data selection conditions: (a) the cross-sectional/longitudinal convergence age gradients for the T4 longitudinal sample ($n = 132$) using all available data points (i.e., T1, T3, and T4 data); these gradients combine cross-sectional and longitudinal information over chronological age (hence convergence); (b) the cross-sectional T1 gradient of the T4 longitudinal sample (i.e., the same sample as before; $n = 132$); here, the T1 cross-sectional age gradient was examined for individuals who survived and participated up to T4; and (c) the cross-sectional T1 gradient of the original T1 sample ($n = 516$).

Figure 2 depicts the three age gradients described above for perceptual speed, episodic memory, fluency, and verbal knowledge, respectively. Each of the ability factors is represented by unit-weighted composites of the two measures of each construct. The resulting scores have been scaled as *T* scores ($M = 50$, $SD = 10$), with the cross-sectional T1 data of the full sample ($n = 516$) serving as the reference sample. With respect to both fluid mechanics and crystallized pragmatics, age-associated decrements in cognition were less pronounced for the longitudinal sample at T1 (thin solid lines) than for the full cross-sectional sample at T1 (dashed lines). Specifically, negative gradients prevailed for all four abilities in the full T1 sample, but verbal knowledge did not decline significantly in the longitudinal sample. This pattern of age gradients suggests that decline in the fluid mechanics is normative and age-based, whereas decline in verbal knowledge appears to be partially or primarily associated with closeness to death (see also Bosworth & Schaie, 1999; Small, Fratiglioni, von Strauss, & Bäckman, 2003).

The third class of age gradients, the longitudinal convergence gradients for the T4 sample (thick solid lines), reinforces this impression. Three of the four intellectual abilities showed signs of accelerated decline in very old age. Also, whereas perceptual speed, episodic memory, and fluency did not differ significantly from each other in rates of change, verbal knowledge showed less negative change, with stability up to about age 90. Beyond this age, mortality-related changes appeared to dominate over aging-related changes, even in highly select samples of survivors and on a pragmatic intellectual ability such as verbal knowledge.

Singer et al. (2003b) also reported evidence for an association of socio-biographical status and sex with initial levels of intellectual functioning. Women scored higher on episodic memory and fluency (cf. Herlitz, Nilsson, & Bäckman, 1997). Similarly, higher levels of socio-biographical status, defined as a composite of income, occupational prestige, social class, and years of education, were associated with higher levels of functioning in all intellectual abilities. Neither socio-biographical status nor sex was related to rates of age-associated change. Thus, in the BASE sample, individuals of different sex or varying socio-biographical status did not differ reliably in

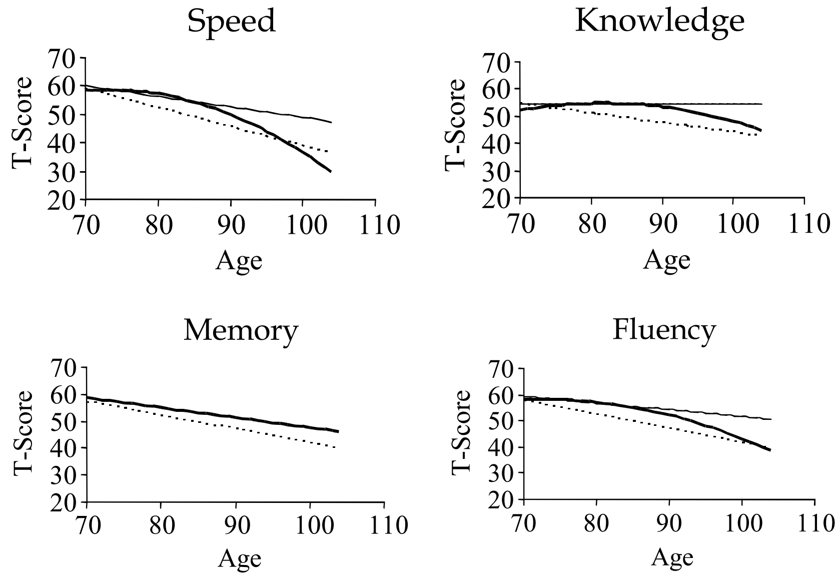

Fig. 2. Intellectual ability age gradients observed in the Berlin Aging Study as a function of sample and measurement occasion. Thick solid lines represent cross-sectional/longitudinal convergence gradients of the longitudinal sample ($n = 132$), and encompass measurements from T1, T3, and T4, which encompass an average longitudinal observation period of 6 years. Thin solid lines represent cross-sectional gradients of the same longitudinal sample ($n = 132$), and are based on measurements taken at T1. Finally, dashed lines represent cross-sectional gradients for the total T1 sample ($n = 516$). Adapted from Singer, Verhaeghen, Ghisletta, Lindenberger, and Baltes (2003).

their amount of age-associated change (cf. Hultsch, Hertzog, Dixon, & Small, 1998; Rabbitt, Diggle, Smith, Holland, & McInnes, 2001).

Dedifferentiation of Psychometric Abilities in Old and Very Old Age

Cross-sectional, and some longitudinal, evidence points to dedifferentiation of intellectual abilities in old and very old age (e.g., Baltes & Lindenberger, 1997; Cunningham, 1980; Hultsch et al., 1998; S.-C. Li et al., in press; Mitrushina & Satz, 1991; Reinert, 1970; Schaie, Maitland, Willis, & Intrieri, 1998; but see Juan-Espinosa et al., 2002; Park et al., 2002). With respect to inter-individual differences in performance on psychometric measures, three forms of dedifferentiation can be distinguished (Baltes & Lindenberger, 1997; Lindenberger & Baltes, 1994, 1997). First, the directions of mechanic (broad fluid) and pragmatic (broad crystallized) age gradients tend to converge in old and very old age (directionality dedifferentiation). Second, correlations among intellectual abilities tend to be higher in old age than during earlier periods of adulthood (covariance dedifferentiation within the intellectual domain). Third, correlations between intellectual abilities and measures of sensory and sensorimotor functioning are higher in old and very old age than during earlier periods of adulthood (across-domain covariance dedifferentiation).

Baltes and Lindenberger (1997) provided cross-sectional evidence for all three empirical patterns by comparing a group of young and middle-aged adults ($n = 171$, 25–69 years) with the T1 cross-sectional BASE sample. As reported above, the three intellectual abilities closely related to the mechanics of cognition, perceptual speed, reasoning, and episodic memory, displayed negative associations with age in both samples. In contrast, the two abilities more closely related to the pragmatics, fluency and verbal knowledge, showed negative age gradients in the BASE sample but not in the young comparison group. Second, correlations among the five intellectual abilities were higher in the BASE sample than in the younger comparison group; for example, the median correlation among the five intellectual abilities was $r = 0.38$ in the younger and $r = 0.71$ in the older group. Third, the correlation between sensory functions (i.e., visual and auditory acuity) and intellectual functioning was considerably lower in the younger comparison group than in the BASE sample. Averaged over the five intellectual abilities and an intellectual ability composite ("intelligence"), individual differences in vision and hearing predicted about 11% of the total variance in cognitive performance in the younger comparison group, but about 31% in the BASE sample. This age-associated increase in correlations was especially pronounced for the two pragmatic intellectual abilities.

To account for directionality and covariance dedifferentiation within and across intellectual and sensory domains, and informed by theoretical propositions from lifespan theories of cognitive development (e.g., Lindenberger, 2001), it was proposed that some of the age-associated changes in sensory, perceptual, and higher-order cognitive processes may be attributable to common causes (Baltes & Lindenberger, 1997; Lindenberger & Baltes, 1994). According to this conjecture, which has been termed "common cause hypothesis," senescent changes in behavior cut across traditional demarcation lines between different domains and levels of processing. Note, however, that cross-sectional analyses of age-heterogeneous data sets do not permit conclusive statements about the degree of functional inter-dependence between different senescent changes in behavior, but need to be complemented by other research designs and statistical methods (e.g., Baltes, Reese, & Nesselroade, 1988; Hertzog, 1996; Hofer & Sliwinski, 2001; Lindenberger & Pötter, 1998; MacDonald, Hultsch, Strauss, & Dixon, 2003; see below).

If dedifferentiation is due, in part, to senescent changes with widespread consequences, then decline in the pragmatics should follow and originate from decline in the mechanics (Ghisletta & Lindenberger, in press-a). Specifically, reductions in the mechanics, as indexed by declining intellectual abilities in the broad fluid domain, are assumed to increasingly limit the expression and accumulation of biographically acquired knowledge, as indexed by declining intellectual abilities from the broad crystallized domain.

To provide direct evidence for lead-lag relations between mechanic and pragmatic declines, Ghisletta and Lindenberger (in press-a) applied a recently developed variant of latent growth curve modeling, the Dual Change Score Model (DCSM; McArdle, 2001; McArdle, Hamagami, Meredith, & Bradway, 2000), to cross-sectional/longitudinal convergence data from BASE ($n = 516$; age range = 70–104; observations up to T4 were included in the analysis). Perceptual speed indexed the mechanics, and verbal knowledge the pragmatics of cognition. The selection of these two intellectual abilities was guided by three considerations: (a) In BASE, perceptual speed and verbal knowledge consistently have emerged as the two most disparate intellectual abilities. For example, in the full cross-sectional sample at T1, perceptual speed showed the most and verbal knowledge the least negative association with age (see Table 2). Furthermore, as predicted by two-component models of cognition, perceptual speed has been found to be more closely related to biological markers such as sensory functioning than verbal knowledge, whereas verbal knowledge is more closely related to various sociobiographical markers than perceptual speed (see Fig. 3; for details, see Lindenberger & Baltes, 1997; cf. Singer et al., 2003b); (b) a large body of evidence suggests that perceptual speed is a sensitive indicator of cognitive decline in old age (e.g., Hertzog, 1989; Lindenberger et al., 1993; Salthouse, 1991; Verhaeghen & Salthouse, 1997); (c) originally, fluency was conceptualized as a second marker of the broad fluid domain, in addition to verbal knowledge (Lindenberger & Baltes, 1994), given that performance on fluency tests taps into semantic memory. However, fluency also requires speeded executive processes and retrieval operations (e.g., Mayr & Kliegl, 2000; Salthouse, 1993). Therefore, fluency appears to be a less valid (pure) marker of the pragmatics than verbal knowledge. In line with this argument, the difference between fluency and perceptual speed in correlational relations to biological and life-history predictors is less pronounced than the corresponding difference between perceptual speed and verbal knowledge

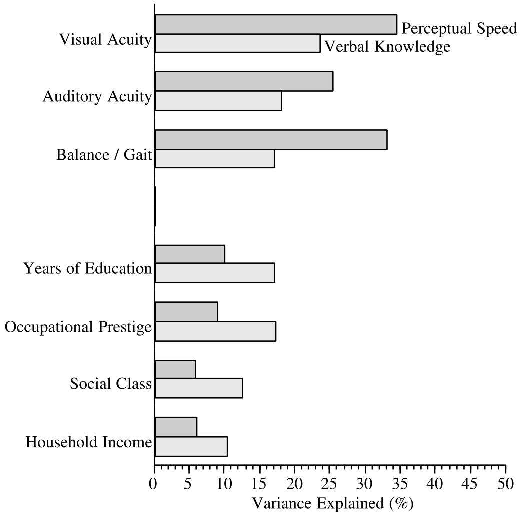

Fig. 3. The divergent validity of the two-component model of lifespan intellectual development subsists into very old age. The figure displays differential correlational links of perceptual speed, a marker of the fluid mechanics, and verbal knowledge, a marker of the crystallized pragmatics, to various indicators of socio-biographical and biological (e.g., sensory) status. Perceptual speed was more highly correlated with biological indicators than verbal knowledge, and verbal knowledge was more highly correlated with socio-biographical indicators than perceptual speed. Thus, despite a general tendency towards dedifferentiation due to age-based losses in the mechanics (Ghisletta & Lindenberger, in press), the two components of lifespan cognition continue to show signs of divergent external validity. Data are taken from the Berlin Aging Study ($n = 516$, age range = 70–103 years). Adapted from Lindenberger and Baltes (1997).

(cf. Lindenberger & Baltes, 1997). In addition, for the longitudinal T3 sample (see Fig. 1), the age gradients of fluency tend to be more similar to the age gradients of perceptual speed and episodic memory than to the age gradient of verbal knowledge (cf. Singer et al., 2003b).

The DCSM is a variant of latent growth curve models (LGM), which in turn share many features with multilevel models (MLM, see also hierarchical linear models, random coefficient models, and mixed effects models). LGM and MLM developed quite independently and in different statistical traditions but are highly similar and at times identical (e.g., Ghisletta & Lindenberger, in press-b; Lindenberger & Ghisletta, in press). In the context of longitudinal data, both LGM and MLM model a population time gradient and represent individual trajectories as deviations from this gradient.

Compared to standard applications of multivariate LGM and MLM, the DCSM allows for empirical testing of lead-lag relations of time-locked associations between different variables. To this end, the DCSM estimates an auto-proportion (dynamic) parameter expressing the effect of a variable at time $t-1$ on the change between $t-1$ and t for this variable. In the multivariate case, the DSCM additionally specifies an analogous dynamic effect of one time-based variable onto change in other time-based variables. At present, these extensions are possible with the LGM but not with the MLM approach.

With a bivariate DCSM, the proposition that the mechanics of cognition drive (i.e., temporally precede and causally influence) decline in the pragmatics can be tested empirically by posing the following statistically specifiable question: Is the effect size of the influence of level of perceptual speed on subsequent change in verbal knowledge greater than the influence of level of verbal knowledge on subsequent change in perceptual speed? When applied to full-information, cross-sectional/longitudinal convergence data from BASE (i.e., data with all 516 individuals included, and with observations up to T4 if available), the corresponding statistical tests indicated that perceptual speed was indeed the leader and verbal knowledge the lagger. Specifically, allowing for a lagged influence from perceptual speed onto changes in verbal knowledge resulted in a χ^2-decrease of 31 points ($df = 1$). In comparison, allowing for a lagged influence from verbal knowledge onto changes in perceptual speed was associated with a χ^2-drop of 10 points ($df = 1$). The equations for the algebraic expectations of both cognitive variables allow a good glimpse at the system considered.

$$\text{Speed}_t = (0.96) \cdot \text{Speed}_{t-1} + (0.07) \cdot \text{Knowledge}_{t-1} + (-2.05) \quad (1)$$

$$\begin{aligned}\text{Knowledge}_t = {} & (0.01) \cdot \text{Knowledge}_{t-1} \\ & + (0.51) \cdot \text{Speed}_{t-1} + (24.01)\end{aligned} \quad (2)$$

Equation (1) specifies the algebraic expectation for the true score of speed at time t as a function of the true scores of speed and knowledge at time $t-1$. Equation (2) specifies the algebraic expectation of the true score of knowledge at time t as a function of the true scores of knowledge and speed at time $t-1$. As can be seen from Equation (1), speed at time t is highly dependent on speed at time $t-1$ (0.96), but little dependent on the previous knowledge score (0.07). On the other hand, knowledge at time t is not highly dependent on the previous knowledge score (0.01), but is strongly influenced by the speed score at time $t-1$ (0.51). Hence, when and if they occur, declines in verbal knowledge were temporally preceded and predicted by lower levels of perceptual speed. The expectations represented by the two equations take into account the differences in reliability between the speed and knowledge composite scores.

These results provide direct support for the hypothesis that old-age decrements in the pragmatics of cognition are driven by decline in the mechanics. In the meantime, they have been replicated in a different sample and with different indicators (Ghisletta & Ribaupierre, 2003). A direct consequence of mechanic-pragmatic lead-lag relations is that intellectual abilities from the broad crystallized domain such as verbal knowledge become increasingly saturated with mechanic variance with advancing age, greater proximity to death, or both. Thus, the bivariate DCSM simultaneously captures the dynamics of both directionality and covariance dedifferentiation.

Initial Attempts to Elucidate the Age-Associated Link Between Intellectual and Sensory Domains

At the descriptive level, strong evidence for an association between cognitive performance and “biomarkers” such as sensory functioning, grip strength, lower limb strength, balance, tactile information processing, and forced expiratory volume has accumulated in recent years, as well as some evidence that this association increases with age (e.g., Anstey, 1999; Anstey, Lord, & Williams, 1997; Anstey & Smith, 1999; Baltes & Lindenberger, 1997; S.-C. Li, Jordanova, & Lindenberger, 1998; Lindenberger & Baltes, 1994, 1997; Salthouse, Hambrick, & McGuthry, 1998; for early evidence, see Heron & Chown, 1967; MacFarland, 1968; Schaie, Baltes, & Strother, 1964). For example, using cross-sectional data from the initial (incomplete) T1 BASE sample ($n = 156$; mean age = 85 years, age range = 70–103 years), Lindenberger and Baltes (1994) examined relations among chronological age, auditory acuity, close and distance visual acuity, and five intellectual abilities. Individual differences in cognition were represented by five first-order intellectual abilities and a second-order factor of intelligence. Approximately 49% of the total and 93% of the age-related variance in intelligence was predicted by individual differences in vision, hearing, or both. Lindenberger and Baltes (1994) did not find any evidence for an age-associated increase in associations between intellectual and sensory functioning within old age. However, Baltes and Lindenberger (1997) found that this association was stronger in the BASE sample than in a reference sample of young and middle-aged adults (25–69 years; see above).

Several hypotheses have been proposed to account for the age-associated link between sensory and intellectual domains (cf. Ghisletta & Lindenberger, 2003). Among the more prominent are the following: (a) the performance factor hypothesis, according to which ability-extraneous sensory performance factors operating during cognitive assessment contribute to the age-associated link between sensory and intellectual domains (for conflicting evidence, see Lindenberger, Scherer, & Baltes, 2001); (b) the cascade hypothesis originally proposed by Birren (1964), which assumes a time-ordered sequence of decline in biomarkers, changes in cognition, and terminal drop (for a similar view, see Anstey, 1999; Anstey & Smith, 1999); (c) the cognitive permeation hypothesis, according to which sensory, sensorimotor, and perceptual aspects of behavior require increasing amounts of cognitive processing with advancing age (e.g., Lindenberger, Marsiske, & Baltes, 2000; cf. Schneider & Pichora-Fuller, 2000; for supportive quasi-experimental evidence, see K.Z.H. Li, Lindenberger, Freund, & Baltes, 2001; Lindenberger et al., 2000); (d) the spuriousness account, which refers to the possibility that independent senescent changes are superimposed onto each other in cross-sectional age-heterogeneous data sets to produce statistical associations between functional domains that bear little or no substantive meaning (e.g., Bäckman et al., 2000; Hofer, Berg, & Era, 2003; Hofer & Sliwinski, 2001; Lindenberger & Pötter, 1998; for an early formalization, see Kalveram, 1965; cf. Reinert, Baltes, & Schmidt, 1966); and (e) the common cause hypothesis (e.g., Baltes & Lindenberger, 1997; Lindenberger & Baltes, 1994), according to which a major portion of the causes of senescent changes are shared among sensory, perceptual, and cognitive domains of functioning.

Lindenberger, Scherer, and Baltes (2001) explored the hypothesis that cognition-extraneous sensory acuity reductions operating during cognitive assessment are a major source of the increase in the link between sensory and intellectual performance from adulthood to old age. According to this account, sensory difficulties in identifying relevant stimuli result in lower psychometric test performance and, considering the pronounced sensory decline with age, induce associations between intellectual and sensory functioning in age-heterogeneous samples of older individuals. To examine this possibility, Lindenberger et al. (2001) administered the BASE battery to middle-aged adults ($n = 218$, age range = 30–50 years) under conditions of reduced auditory acuity, visual acuity, or both. Auditory acuity was reduced through the use of noise protectors, and visual acuity was lowered through the use of partial occlusion filters. The two sensory manipulations successfully reduced visual and auditory acuity to levels comparable to those observed in the younger half of the BASE T1 sample ($n = 258$, age range = 70–84), but completely failed to negatively affect performance on the BASE cognitive battery relative to various placebo and no-treatment control groups. Hence, peripheral input factors related to visual and auditory acuity do not offer a viable explanation for the age-associated increase in sensory-intellectual associations observed with the BASE cognitive battery.

Lindenberger and Pötter (1998) formally demonstrated that hierarchical linear regressions and related variance-partitioning procedures performed on cross-sectional age-heterogeneous data do not allow researchers to conclusively separate independent from shared age-linked influences. Hence, spuriousness and common-cause accounts cannot be unambiguously separated with this combination of statistical tools and research designs. In agreement with others (e.g., Nesselroade & Schmidt McCollam, 2000), Lindenberger and Pötter (1998) stressed the need to gather multivariate longitudinal data on inter-individual differences in change and intra-person associations in change to identify the dimensionality of the causes underlying changes in behavior.

Initial attempts to shed light on the intellectual-sensory puzzle using longitudinal BASE data on inter-individual differences in intellectual and sensory changes have yielded a relatively complex picture. For instance, Ghisletta and Lindenberger (2003) applied a quadrivariate version of the DCSM, described in the preceding section, to longitudinal BASE data. Analyses were limited to two intellectual abilities, perceptual speed and verbal knowledge, and to two sensory abilities, close and distant visual acuity. In contrast to the bivariate application reported by Ghisletta and Lindenberger (in press-a), and primarily for technical reasons, the time dimension was represented in terms of measurement occasions rather than chronological age, which was included in the model as a covariate. Again, all available data, including data from participants not assessed at all occasions, were used (e.g., full-information maximum likelihood estimation; cf. Arbuckle, 1996).

Four general and two more specific results emerged. First, the quadrivariate DCSM seemed to account well for the data structure ($\chi^2_{(n=516, df=61)} = 85$; RMSEA = 0.028). Second, the longitudinal gradients obtained for the four variables were consistent with the expectations. Namely, knowledge and distant vision had the less pronounced longitudinal time gradient, while speed and close vision showed much sharper average decline over time. Third, the static links among the four variables were relatively strong for both level and change components. For example, level of speed correlated .65 with level in knowledge and .46 with level in distant vision. Level in knowledge correlated .35 with level in distant vision. Analogously, change in speed correlated .76 with change in knowledge and .41 with change in distant vision. Change in knowledge correlated .81 with change in distant vision. Fourth, the dynamic links among the four variables were also non-negligible for the system considered. Specifically, changes in perceptual speed were predicted by chronological age and by close vision, changes in verbal knowledge by age and itself, changes in close vision by distant vision and by itself; and changes in distant vision by perceptual speed and verbal knowledge. In short, the reliable dynamic links among the four variables were essential, above and beyond the traditionally modeled static links mentioned above, to account for the structure of the data. Indeed, neglecting to account for the dynamic resulted in a highly significant decrease in fit of 46 χ^2 points for 12 degrees of freedom.

The two more specific results appeared from nested statistical comparisons. First, perceptual speed emerged as the most important but not as the only predictor of change in the other variables. In addition, for the four variables considered, dynamic links across sensory and intellectual domains were more pronounced than dynamic links within sensory and intellectual domains. This pattern of results extends the cross-sectional link observed in earlier studies to reliable longitudinal lead-lag couplings across domains while controlling for chronological age and task differences in measurement reliability. Though alternative explanations cannot be ruled out (e.g., effects of sample heterogeneity, statistical dependence of the parameters), such a pattern does seem at odds with the notion of age-linked but functionally independent (i.e., substantively spurious) associations. Also, given that effects of age on future cognitive performance were not fully captured by sensory status and that sensory-intellectual links were bidirectional, the results also fail to support versions of the cascade and mediation hypotheses that assign precedence to the sensory domain. At the same time as the results are not fully consistent with the above accounts, the results do not reject the hypothesis that the link between sensory and intellectual functioning is due to the operation of domain-general mechanisms (cf. Christensen, this issue).

Episodic Memory Performance in Very Old Age: Limits to Plasticity

In cognitive aging research, plasticity generally refers to learning gains, levels of performance after practice on performance-enhancing techniques, or both (e.g., Baltes, 1987; Verhaeghen, Marcoen, & Goosens, 1992). With respect to episodic memory plasticity in normal aging, most studies have focused on the young–old, with ages ranging from 60 to 80 years. Most of these studies suggest that instructions and practice in mnemonic techniques may lead to pronounced improvements for healthy older adults (Baltes & Kliegl, 1992; Lindenberger & Baltes, 1995; for meta-analysis, see Verhaeghen et al., 1992). At the same time, older adults generally benefit less from training programs than young adults (e.g., Baltes & Kliegl, 1992; Kliegl, Smith, & Baltes, 1990; Verhaeghen & Marcoen, 1996).

To extend our knowledge about adult age differences in plasticity of episodic memory performance into very old age, Singer, Lindenberger, and Baltes (2003a) provided episodic memory instructions and training to a subsample of old and very old BASE participants who had completed the T4 Intensive Protocol ($n = 96$, mean age = 84 years, age range = 75–101 years). For comparative purposes, a young reference sample ($n = 20$, mean age = 24.7 years, age range = 21–29 years) was included in the study. Both age groups were instructed and trained in a simplified version of the Method of Loci (MoL), an imagery-based mnemonic strategy for the serial recall of word lists.

In the group of very old individuals, performance gains following instructions were modest (1.62 *SD* from baseline performance; direct comparison to the young group were rendered impossible by ceiling effects in that group). Furthermore, according to a relatively lenient criterion, 85% of the BASE participants failed to improve their recall performance during four sessions of practice. Given the positive selection bias of the BASE sample, these results are likely to overestimate the average amount of available plasticity in the population. Thus, compared to earlier results based on healthy samples of young-old individuals (e.g., Kliegl et al., 1990), the amount of plasticity in episodic memory performance appears to be severely compromised at more advanced ages. In particular, old–old individuals appear to be limited in their ability to optimize the use of a newly acquired mnemonic skill through further experience.

The Singer et al. (2003a) study also allowed looking at predictive links of concurrent and longitudinal covariates to plasticity within the BASE sample. As expected, perceptual speed was more closely related to individual differences in episodic memory after training than before training, while the predictive value of verbal knowledge and socio-biographical variables decreased during this interval. This result confirmed the prediction that instructions and extensive practice induce the use of more general cognitive resources, and reduce the role of background factors such as pre-experimental familiarity with episodic memory strategies or other aspects of the test situation (cf. Kliegl et al., 1990). Finally, prior six-year longitudinal changes in perceptual speed (i.e., T4 minus T1 difference scores of the perceptual speed composite) predicted individual differences in episodic memory plasticity. At the same time, these changes did not account for any additional variance in plasticity after controlling for concurrent T4 status.

In sum, the findings reported by Singer et al. (2003a) suggest (a) that plasticity in episodic memory performance is severely compromised in old age; (b) that biological status is a more powerful source of individual differences in episodic memory plasticity in very old age than socio-biographical status; and (c) that inter-individual differences in six-year longitudinal changes and intervention-based short-term changes are connected (cf. Lindenberger & Baltes, 1995).

RETEST EFFECTS IN BASE: RESULTS FOR THE COGNITIVE BATTERY

Validity threats to developmental studies on intellectual functioning have often been discussed under the headings of selectivity, cohort effects, and retest effects, among others (e.g., Lövdén & Lindenberger, in press; Salthouse, 2000; Schaie, 1988, 1996). Out of these three commonly acknowledged validity threats, retest effects have probably received least attention, perhaps less than they deserve (but see, e.g., McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002; McArdle, Prescott, Hamagami, & Horn, 1998; Rabbitt et al., 2001; Rabbitt et al., this issue; Rönnlund, Nyberg, Bäckman, & Nilsson, 2002; Salthouse, 1991; Schaie, 1988, 1996; Wilson et al., 2002).

The term retest effects refers to the possibility that, in longitudinal designs, prior exposure to a test and to the testing situation may affect performance at retest, either through specific factors such as practice on test-relevant skills, or through more general reactive effects such as familiarization with the testing environment or changes in factors such as motivation and interest (which may vary in positive or negative directions). Retest effects, if left analyzed, may lead researchers to underestimate the amount of average age-related decline in the underlying ability dimension of interest. Also, retest effects may differ across persons, tests, abilities, and interactions thereof (e.g., Rabbitt et al., this issue).

To reduce retest effects, some longitudinal studies have used quasi-equivalent versions of tests alternating across measurement occasions (e.g., the parallel forms approach; see Giambra, Arenberg, Zonderman, Kawas, & Costa, 1995; Hultsch et al., 1998). However, this practice may introduce confounding sources of variability in change, which complicate analyses of individual differences in change. Also, the parallel forms approach has not always been fully successful in eliminating or reducing retest effects (e.g., Hultsch et al., 1998).

Typically, retest effects are considered as a validity threat that is specific to repeated-measures designs such as longitudinal panel studies of cognitive aging. However, at least some of the factors contributing to retest effects are likely to operate in cross-sectional studies as well. For example, individual differences in pre-experimental familiarity with the kind of problem structures used in tests of intellectual functioning as well as familiarity with general aspects of test taking are probably not restricted to repeated assessment but also present at the first measurement occasion. In this sense, mechanisms contributing to retest effects may also contribute to cohort effects. Also, given that influences contributing to retest effects are present in both cross-sectional and longitudinal research designs, longitudinal designs are actually advantageous because they allow, albeit not perfectly, for statistical quantification and control of such influences.

At the level of samples, one way of estimating retest effects is to compare the performance of returnees with the performance of cohort-matched but not previously tested samples of participants (Salthouse, 1991; Schaie, 1988). If attrition effects are controlled, the remaining differences between samples reflect retest effects (and error). In a recent study from the Betula study, Rönnlund et al. (2002) used this approach to explore retest effects in semantic memory performance (a composite of fluency and verbal knowledge) and episodic memory performance (a composite of cued and free recall). Two large and representative samples were compared (age range at baseline $=$ 35–80 years, total $n = 1788$, measurement interval between occasions $=$ 5 years). Average retest effects were negligible for the semantic memory composite ($SD = 0.04$), but reliable for the episodic memory composite ($SD = 0.15$). For episodic memory only, adjusting for retest effects accentuated the decline observed for the older groups of individuals (60–80 years), and converted the increments observed at younger ages into an age-invariant gradient. Thus, a failure to model retest effects may positively bias age gradients for some variables, but not necessarily for all (see also Rabbitt et al., 2001).

In the following, we examine the possible presence of retest effects in BASE. Our analyses are restricted to the eight longitudinally administered tests of the BASE cognitive battery. We employ recently developed statistical models that permit to estimate the shape and magnitude of retest effects (e.g., McArdle et al., 1998; Rabbitt et al., 2001). In contrast to most of the earlier BASE analyses on cognition, we evaluate retest effects at the level of individual tests, and not at the level of ability composites. This decision was based on the consideration that retest effects may well differ between individual tests within abilities. Aggregating across tests prior to analyzing retest effects may hide such differences.

Data analyses were based on the T1 to T4 assessment waves of BASE, and encompassed all available data points at the level of tests (see Table 1). Analyses of this sort accommodate incomplete data and unbalanced data structures, and also adjust for selectivity under the assumption that data are missing at random (MAR; Rubin, 1974). The latter feature deserves some elaboration. The MAR assumption allows for differences in prior level of performance and prior observed change to predict future participation, but the assumption is violated if the unobserved change for individuals that drop out is different from the observed change. From the results summarized above, we can infer that unobserved change in intellectual abilities would in fact be different from the observed change in BASE, but the magnitude of these effects are minor as compared to the effects of level. Furthermore, the most attrition-informative variables are included in the models that will be applied (i.e., intelligence and age; see Fig. 1 and above) and in longitudinal studies on intellectual functioning such as BASE the data are correlated across measurement occasions. Thus, even in the presence of the relatively minor selectivity effects in change, the initial level is to some degree predictive of the missing value after dropout; that is, the level data carry information about observed, and unobserved, change. Accordingly, with likelihood analyses under MAR we can, at least to some extent, account for selectivity effects in change in addition to capture the largest selectivity effects (i.e., level). Therefore, analyses under the MAR assumption offer a relatively robust approach to the present research questions (cf. Rabbitt et al., this issue).

Statistical Procedures

Analyses were performed with longitudinal multilevel models (Bryk & Raudenbush, 1987; Laird & Ware, 1982). Change was defined in terms of chronological age. As a consequence, participants contributed short time segments (mean observation period $=$ 1.98 years) to an age gradient spanning 34 years (70–104 years).

Chronological age was centered at the average longitudinal age of the sample. Specifically, intercept means, intercept variances, and slope-intercept covariances, if present, were estimated at 84.73 years. All variables were normed to the T1 cross-sectional sample, and scaled as *T* scores ($M = 50$, $SD = 10$).

Data analysis was univariate (i.e., separate analyses for each of the eight cognitive tests), and proceeded in two steps. First, each cognitive test was modeled separately without estimating retest effects. This allowed us to estimate an optimal average age gradient for each variable as well as inter-individual differences around that gradient, while neglecting retest effects. We included tests for both linear and quadratic effects of chronological age. Quadratic terms were residualized on linear terms to obtain the quadratic component of age that is independent of the linear component of age. Second, we included predictors representing various components of retest effects, and compared the resulting age gradients with the age gradients obtained without such predictors. Retest effects were coded as a function of measurement occasion, and not as a function of age. For example, at age 85, individuals may have taken a test for the first, second, or third time.

Two types of retest effects were separately specified and tested. The first type corresponds to the hypothesis that retest effects occur at the second measurement occasion, and persist constantly thereafter, without growing. Thus, participants' performance would be influenced by a constant retest component after first assessment; any additional exposure to the test would not result in further retest-induced performance changes. We designate this type of retest effect as flat. In statistical terms, this effect corresponds to a vector with a value of 0 for the first occasion of measurement, and with values of 1 for all subsequent measurement occasions.

The second type of retest effect corresponds to the hypothesis that retest influences grow linearly as a function of measurement occasion. This component reflects the assumption that participants' performance continues to benefit from further exposures to the test. We designate this retest effect as growing; the values of the corresponding vector are equal to the occasion of measurement minus one. To summarize, the flat retest effect represents a "jump" from the first to the second measurement occasions, whereas the growing retest effect represents a linear slope starting at the second measurement occasion.

The two types of retest effects were weakly correlated with linear (*r*'s ranging from -0.06 to -0.01) and quadratic (*r*'s ranging from -0.23 to -0.09) components of age. Given that age and retest effects were empirically separable (i.e., not collinear) in the present data analysis design, we also tested for age by retest component interactions, which are equivalent to the multiplication of chronological age by either flat or growing retest effects. Retest effects and age-by-retest interaction effects were initially modeled as fixed effects only, and then as random effects, the latter allowing for inter-individual differences.

Presence of the two types of retest effects was assessed in separate models. When different retest models significantly improved the corresponding baseline longitudinal models, the model with the greatest degree of improvement was retained. For two variables, Identical pictures and Word Beginnings, the two types of retest effects resulted in equivalent improvements. In these two cases, we report both results.

Results

The MLM parameters for the first set of univariate models are displayed in Table 3. All cognitive variables displayed fixed negative linear age effects, ranging in value from -0.89 and -0.66 annual *T*-score decrements for Digit Letter and Identical Pictures (perceptual speed) to -0.34 and -0.29 annual *T*-score decrements for Vocabulary and Spot-a-Word (verbal knowledge), in that order. In addition, Digit Letter, Identical Pictures, Categories, Vocabulary, and Spot-a-Word displayed negative quadratic components, indicating an acceleration of cognitive decline with increasing age. Thus, for these variables, the linear age effect should not be interpreted in isolation. Descriptively, quadratic components were largest for the two indicators of perceptual speed and smallest for the indicators of verbal knowledge.

Table 4 summarizes the results for models with retest predictors. Retest effects were significant for Identical Pictures, Categories, Word

Table 3. MLM Parameter Estimates (Standard Errors) for Cognitive Tests Without Modeling Retest Effects.

Intellectual Ability Test	Perceptual speed		Fluency		Episodic memory		Verbal knowledge	
	Digit Letter	Identical Pict.	Categories	Word Begin.	Pair Assoc.	Memory Text	Vocabulary	Spot-a-Word
Fixed Effects								
Intercept	49.64 (0.46)	50.31 (0.40)	50.98 (0.39)	50.44 (0.41)	50.27 (0.40)	50.28 (0.37)	50.35 (0.41)	50.11 (0.43)
Linear age	−0.89 (0.05)	−0.66 (0.05)	−0.52 (0.04)	−0.38 (0.05)	−0.43 (0.05)	−0.41 (0.04)	−0.34 (0.04)	−0.29 (0.05)
Res age squared	−0.03 (0.004)	−0.02 (0.005)	−0.01 (0.004)	–	–	–	−0.01 (0.004)	−0.01 (0.005)
Random Effects								
Intercept	76.74 (6.08)	52.42 (5.03)	65.62 (4.78)	58.95 (5.37)	58.68 (5.15)	36.31 (4.75)	66.33 (5.40)	67.42 (6.01)
Linear age	0.44 (0.06)	–	–	–	–	–	–	–
Covariance	1.55 (0.50)	–	–	–	–	–	–	–
Residual	9.88 (0.43)	25.77 (2.12)	24.62 (1.04)	34.68 (2.62)	33.60 (2.49)	45.26 (3.49)	23.19 (1.85)	25.14 (2.07)
−2LL (REML)	11,866	5,265	10,848	6,029	5,977	5,930	5,815	5,438

Note. Perceptual speed is measured by Digit Letter and Identical Pictures, fluency by Categories and Word Beginnings, episodic memory by Paired Associates and Memory for Text, and verbal knowledge by Vocabulary and Spot-a-Word; res age squared is the residualized squared component of chronological age after controlling for the linear component of chronological age; covariance denotes the covariance between the intercept and chronological age random effects (if both are reliably different from zero); −2LL is the −2 log likelihood fit index; REML denotes restricted maximum likelihood.

Table 4. MLM Parameter Estimates (Standard Errors) for Cognitive Tests With Modeling Retest Effects.

Intellectual Ability Test	Perceptual speed		Fluency		Episodic memory		Verbal knowledge	
	Digit Letter	Identical Pict.	Categories	Word Begin.	Pair Assoc.	Memory Text	Vocabulary	Spot-a-Word
Fixed Effects								
Intercept	49.64 (0.46)	49.69 (0.41)	50.19 (0.41)	50.09 (0.42)	50.05 (0.42)	50.28 (0.37)	49.90 (0.41)	50.11 (0.43)
Linear age	−0.89 (0.05)	−0.66 (0.05)	−0.59 (0.04)	−0.41 (0.05)	−0.46 (0.05)	−0.41 (0.04)	−0.41 (0.05)	−0.29 (0.05)
Res age squared	−0.03 (0.004)	–	–	–	–		–	−0.01 (0.005)
Flat retest	–	2.02 (0.47)	1.72 (0.32)	–	–	–	–	–
Flat retest by age	–	−0.20 (0.07)						
Growing retest	–	–	–	1.05 (0.33)	0.68 (0.33)	–	1.45 (0.28)	–
Random Effects								
Intercept	76.74 (6.08)	48.73 (4.84)	62.95 (4.60)	56.68 (5.32)	57.55 (5.13)	36.31 (4.75)	63.70 (5.25)	67.42 (6.01)
Linear age	0.44 (0.06)	–	–	–	–	–	–	–
Covariance	1.55 (0.50)	–	–	–	–	–	–	–
Residual	9.88 (0.43)	25.83 (2.15)	24.55 (1.04)	35.07 (2.67)	33.87 (2.52)	45.26 (3.49)	22.96 (1.84)	25.14 (2.07)
−2LL (REML)	11,866	5,235	10,802	6,020	5,973	5,930	5,786	5,438

Note. Perceptual speed is measured by Digit Letter and Identical Pictures, fluency by Categories and Word Beginnings, episodic memory by Paired Associates and Memory for Text, and verbal knowledge by Vocabulary and Spot-a-Word; flat retest by age denotes the interaction between the flat retest effect and linear chronological age; res age squared is the residualized squared component of chronological age after controlling for the linear component of chronological age; covariance denotes the covariance between the intercept and chronological age random effects (if both are reliably different from zero); −2LL is the −2 log likelihood fit index; REML denotes restricted maximum likelihood.

Beginnings, Paired Associates, and Vocabulary. Fixed retest effects were positive for all five variables, ranging from 0.68 *T*-score units for Paired Associates to 2.02 *T*-score units for Identical Pictures. Compared to the average annual decline rate, these effects are quite impressive. Evidence for flat fixed retest effects was found for Categories, whereas fixed retest effects for Paired Associates and Vocabulary were estimated to be growing. For Identical Pictures and Word Beginnings, both types of fixed retest effects were significant and about equal in magnitude. The alternative parameters for the latter two tests are presented in Table 5.

Most importantly, linear age gradients were more negative in models with reliable retest predictors. For instance, after modeling a flat retest effect, the annual linear decline on Categories increased from −0.52 to −0.59 *T*-score units. Analogous differences were found for Word Beginnings (−0.38 vs. −0.41), Paired Associates (−0.43 vs. −0.46), and Vocabulary (−0.34 vs. −0.41). Figure 4 illustrates the resulting differences in age gradients for Identical Pictures and Vocabulary.

Identical Pictures was the only variable with a reliable age by retest interaction. Previous to estimating retest effects, a linear age gradient of

Table 5. Alternative MLM Parameters (Standard Errors) for Identical Pictures and Word Beginnings.

Test	Identical Pict.	Word Begin.
Fixed Effects		
Intercept	49.76 (0.40)	50.11 (0.42)
Linear age	−0.66 (0.05)	−0.40 (0.05)
Res age squared	–	–
Flat retest	–	1.33 (0.51)
Growing retest	1.27 (0.31)	–
Growing retest by age	−0.16 (0.05)	–
Random Effects		
Intercept	49.18 (4.84)	56.69 (5.34)
Linear age	–	–
Covariance	–	–
Residual	25.59 (2.13)	35.16 (2.68)
−2LL (REML)	5,236	6,022

Note. Identical Pict. refers to Identical Pictures, a test of perceptual speed; Word Begin. refers to Word Beginnings, a test of fluency; res age squared is the residualized squared component of chronological age after controlling for the linear component of chronological age; growing retest by age denotes the interaction between the growing retest effect and linear chronological age; covariance denotes the covariance between the intercept and chronological age random effects (if both are reliably different from zero); −2LL is the −2 log likelihood fit index; REML denotes restricted maximum likelihood.

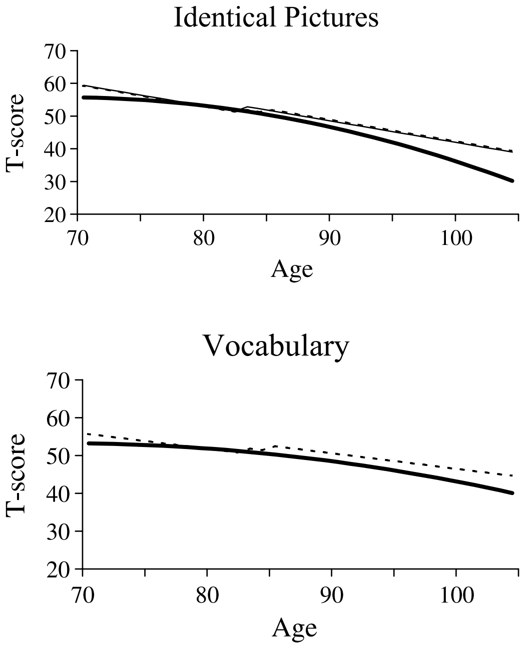

Fig. 4. Age gradients for Identical Pictures (upper panel) and Vocabulary (lower panel) before and after modeling retest effects. In both panels, the thick solid line refers to age gradients without statistical control of retest effects. Retest effects were specified as a function of measurement occasion. For illustration, they are displayed at the averaged sample age of a given measurement occasion. For Identical Pictures, both flat retest effects (solid thin line) and growing retest effects (dashed line) were statistically reliable. For Vocabulary, only a growing retest effect (dashed line) was observed.

−0.66 and a quadratic age gradient of −0.22 *T*-score units were estimated. Analyses including retest effects revealed the same linear age decrement, no quadratic age effect, and either a flat retest effect of 2.02 with an age by retest interaction of −0.20, or a growing retest effect of 1.27 with an age by retest interaction of −0.16. As was true for Categories and Vocabulary, the retest predictor attenuated the quadratic age effect. In addition, younger and older participants profited differentially from retest, the former benefiting more, the latter benefiting less. If this interaction was not considered, the estimated age gradient was either −0.70 with a main flat retest effect of 2.13, or −0.71 with a main growing retest effect of 1.33.

Finally, no reliable random effects of retest were observed for any of the eight cognitive tests. Given the lack of statistical power of the present data analysis design in detecting random retest effects, the absence of such effects is not surprising.

Discussion

The present analyses strongly suggest that retest effects, if left unanalyzed, may contribute a non-negligible positive bias to intellectual age gradients in old and very old age. The presence of retest effects is especially problematic for data-analytic schemes that arrange longitudinal data as a function of measurement occasion, for research designs that confound chronological age and measurements occasion (e.g., when all individuals enter the study at the same age and are tested at identical points in time), or combinations of both. In these cases, retest and age effects are confounded, and separate estimates of retest effects cannot be obtained. However, the descriptive quantification of retest effects poses fewer problems with multiple cohort research designs such as BASE, especially when the resulting data are analyzed as a function of age, and with versatile statistical tools such as MLM.

The specification of retest effects led to substantial reductions of quadratic contributions to age gradients. This result raises the more general question whether findings of accelerating decline (e.g., Colsher & Wallace, 1991; Wilson et al., 2002) might, at least in part and for some of the measures, be due to retest effects. This observation also applies to previous analyses of the BASE data that did not consider retest effects. For example, the degree of similarity between purely cross-sectional and cross-sectional/longitudinal convergence data may increase when retest effects are taken into account (cf. Salthouse, 1991, 2000). However, note that these considerations rely on not always transparent statistical assumptions such as sample homogeneity and data missing at random. To expand the current knowledge concerning the robustness of the statistical models, particularly in the presence of violations of these assumptions, deserves high priority. It should also be noted that the analyses make use of relatively short segments of longitudinal (i.e., within-person) information. Thus, whatever the statistical procedure, we cannot exclude that the resulting age gradients, retest-adjusted or not, also contain influences due to cohort, rather than age, or interactions of the two (cf. Baltes, 1968).

The magnitude of retest effects varied considerably across cognitive tests, ranging from non-significant (Spot-a-Word, Memory for Text, and Digit Letter) to quite impressive (e.g., Identical Pictures). At the same time, tests indicating the same intellectual ability, such as Digit Letter and Identical Pictures, differed markedly in retest effects. Thus, the present analyses confirm the expectation that retest effects operate at the indicator and not at the ability level. Hence, the decision to analyze retest effects at the level of individual tests seems justified. However, substantively, researchers clearly would like to know the extent to which age gradients at the ability level are altered after modeling retest effects at the indicator level. Accordingly, most of the past cognitive work BASE has focused on ability composites, and not on indicators. Future data-analytic efforts need to integrate indicator and ability levels of analysis. Specifically, one may envision second-order LGMs and MLMs to estimate age gradients at the ability level on the basis of retest-adjusted indicator age gradients.

A significant age-by-retest interaction was observed for Identical Pictures, revealing that older individuals profited less from repeated exposure to this test than younger individuals, perhaps reflecting late-life age differences in implicit learning (Salthouse, McGuthry, & Hambrick, 1999). This finding appears to be inconsistent with evidence reported by Rabbitt et al. (this issue), who noted an increase in retest effects with age. Note that the two studies differ in many respects, such as the measures used and the age ranges considered (i.e., 70–104 years in the present analyses compared to 49–92 years in Rabbitt et al., this issue). The substantial reduction in cognitive plasticity (learning potential) with advancing age (e.g., Singer et al., 2003a; see above) appears to be more consistent with negative, rather than positive, associations between age and retest effects.

Finally, and as suggested by others (e.g., Baltes et al., 1988), the conceptual status of retest effects needs further thought. Retest effects are often assumed to reflect ability-extraneous performance factors. According to this line of reasoning, controlling for retest-related variance helps to purify the variance represented at the ability level. At the same time, at least some of the substantive interpretations of retest effects, such as those alluding to cognitive plasticity, learning potential, implicit learning, and the like, may be intimately related to some of the intellectual abilities under study, such as perceptual speed or episodic memory. Thus, it needs to be kept in mind that portions of the variance captured by retest effects may be conceptually related to variance at the level of intellectual abilities.

To summarize, our analyses demonstrate that retest effects influence the shape of age gradients in old and very old age. Two types of retest effects, flat and growing, were separately assessed and compared in magnitude. Retest effects differed in magnitude from test to test, even between highly correlated indicators of the same ability. Future work needs to explore the implications of these findings for analyses located at the ability level, and to identify the mechanisms that contribute to retest effects, and inter-individual differences therein.

OUTLOOK

The present article has focused on themes that figured prominently in BASE research on cognition within the last 10 years: selection, age gradients, dedifferentiation, and limits to plasticity. In addition, we presented new analyses to estimate the contribution of retest effects to performance on eight longitudinally administered cognitive tasks. We would like to emphasize once more that our review of past work on cognition in BASE was far from being exhaustive. Due to our focus on a small number of inter-connected key themes, we neglected a large range of other cognition-related research in BASE. For example, we did not report empirical studies that have examined the links of cognition to domains such as health (Steinhagen-Thiessen & Borchelt, 1999), personality and social functioning (Smith & Baltes, 1997), control beliefs (Kunzmann, Little, & Smith, 2002), or everyday competence (e.g., Lang, Rieckmann, & Baltes, 2002; Marsiske, Klumb, & Baltes, 1997). Moreover, we did not discuss recent re-analyses of the factorial structure of the BASE cognitive battery using nested group factor models that are particularly suitable for disentangling general and specific effects of aging (Schmiedek & Li, in press).

Though prescriptions for future research directions are always risky, we would like to end this article by suggesting three possible routes for future research, both within and beyond the immediate research context of BASE. First, inter-disciplinary links within BASE may be strengthened further to better understand and predict inter-individual differences in patterns of intellectual functioning and change. For instance, biological indicators of sample heterogeneity such as various indicators of somatic health as well as genetic markers may help to discriminate between and perhaps predict different patterns and profiles of behavioral aging (cf. Bäckman, this issue; Nilsson, Sikström, Adolfsson, Erngrund, & Nylander, 1996; Rabbitt et al., this issue; for relevant work within BASE, see Hillen et al., 2000). Compared to the classification of individuals into subgroups on the basis of retrospective criteria such as proximity to death (Singer et al., 2003b; see Fig. 3) or later dementia diagnosis (e.g., Sliwinski, Lipton, Buschke, & Stewart, 1996), biological markers may serve to prospectively classify individuals into differently aging groups (e.g., Deary et al., 2002; Hillen et al., 2000). If collections of aging individuals such as the one followed in BASE represent mixtures of differently aging groups of individuals, some of the aging patterns characterizing such collections at the level of inter-individual differences may not generalize to all or not even to the majority of individuals (cf. Borsboom, Mellenbergh, & van Heerden, 2003; Molenaar, Huizenga, & Nesselroade, 2003; for an application of mixture distribution analysis to cross-sectional cognitive data in BASE, see Reischies, Schaub, & Schlattmann, 1996).

Second, the statistical properties of methods that are commonly used to analyze data with longitudinal information, such as LGM, MLM, and related methods, need to be explored in more detail (Hertzog & Nesselroade, in press). Critical properties of these methods, such as efficiency and lack of bias, have not been sufficiently explored under realistic conditions (but see Hamagami & McArdle, 2001). Clearly, Monte-Carlo simulations and related techniques are needed to overcome this problem. At the same time, such simulations can be used to examine the implications of sample heterogeneity, and other violations of assumptions, for findings about behavioral aging that are based on inter-individual differences in change (e.g., Molenaar et al., 2003).

These observations lead to our third and final consideration. In brief, we echo the longstanding plea of some methodologists (e.g., Nesselroade, 2001) to invest more time and effort into the description and explanation of individual change trajectories (see also Lövdén & Lindenberger, in press; Nesselroade & Schmidt McCollam, 2000). Longitudinal panel studies of old and very old age such as BASE, with few measurement occasions per subject but large and heterogeneous samples, have been indispensable in charting the terrain of behavioral aging. Specifically, and together with a few others studies of this sort (Bäckman, this issue; for review, see Bäckman et al., 2000), BASE has helped to extend our knowledge about cognitive aging into the very old segment of the elderly population, or the Fourth Age (Baltes & Smith, 1999, 2003; Singer et al., in press-a, in press-b). Future longitudinal studies need to follow up on this chart by extensive multivariate observations of within-person change and variability (Hamaker, Dolan, & Molenaar, 2003).

ACKNOWLEDGEMENTS

The work summarized in this article was conducted in the context of the Berlin Aging Study (BASE), which is co-chaired by Paul B. Baltes and Karl Ulrich Mayer at the Max Planck Institute for Human Development, Berlin, Germany. The authors express their gratitude to all colleagues in the Berlin Aging Study who contributed to the work summarized in this article. Martin Lövdén and Ulman Lindenberger are now at the Max Planck Institute for Human Development, Berlin, Germany. Paolo Ghisletta is at the Center for Interdisciplinary Gerontology, University of Geneva, Switzerland, and at the Center for Life Course and Life Style Study, Universities of Lausanne and Geneva, Switzerland.

REFERENCES

Ackerman, P.L. (1996). A theory of adult intellectual development: Process, personality, interest, and knowledge. *Intelligence*, *22*, 227–257.

Aitkin, A.C. (1934). Note on selection from a multivariate normal population. *Proceedings of the Edinburgh Mathematical Society*, *4*, 106–110.

Anstey, K.J. (1999). Sensorimotor variables and forced expiratory volume as correlates of speed, accuracy, and variability in reaction time performance in late adulthood. *Aging, Neuropsychology, and Cognition*, *6*, 84–95.

Anstey, K.J., Lord, S.R., & Williams, P. (1997). Strength in the lower limbs, visual contrast sensitivity, and simple reaction time predict cognition in older women. *Psychology and Aging*, *12*, 137–144.

Anstey, K.J., & Smith, G.A. (1999). Interrelationships among biological markers of aging, health, activity, acculturation, and cognitive performance in late adulthood. *Psychology and Aging*, *14*, 605–618.

Arbuckle, J.L. (1996). Full information estimation in the presence of incomplete data. In G.A. Marcoulides & R.E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). Mahwah, NJ: Lawrence Erlbaum Associates.

Bäckman, L., & Nilsson, L.-G. (1996). Semantic memory functioning across the adult life span. *European Psychologist*, *1*, 27–33.

Bäckman, L., Small, B.J., Wahlin, A., & Larsson, M. (2000). Cognitive functioning in very old age. In F.I.M. Craik & T.A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 499–558). Mahwah, NJ: Lawrence Erlbaum Associates.

Baltes, P.B. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development*, *11*, 145–171.

Baltes, P.B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*, *23*, 611–626.

Baltes, P.B., & Kliegl, R. (1992). Further testing of limits of cognitive plasticity: Negative age differences in a mnemonic skill are robust. *Developmental Psychology*, *28*, 121–125.

Baltes, P.B., & Labouvie, G.V. (1973). Adult development of intellectual performance: Description, explanation, modification. In C. Eisdorfer & M.P. Lawton (Eds.), *The psychology of adult development and aging* (pp. 157–219). Washington, DC: American Psychological Association.

Baltes, P.B., & Lindenberger, U. (1997). Emergency of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging*, *12*, 12–21.

Baltes, P.B., Lindenberger, U., & Staudinger, U.M. (1998). Life-span theory in developmental psychology. In R.M. Lerner (Ed.), *Handbook of child psychology: Theoretical models of human development* (Vol. 1, pp. 1029–1143). New York: Wiley.

Baltes, P.B., & Mayer, K.U. (Eds.). (1999). *The Berlin Aging Study: Aging from 70 to 100.* Cambridge, UK: Cambridge University Press.

Baltes, P.B., Mayer, K.U., Helmchen, H., & Steinhagen-Thiessen, E. (1999). The Berlin Aging Study (BASE): Sample, design, and overview of measures. In P.B. Baltes & K.U. Mayer (Eds.), *The Berlin Aging Study: Aging from 70 to 100* (pp. 15–55). Cambridge, UK: Cambridge University Press.

Baltes, P.B., Reese, H.W., & Nesselroade, J.R. (1988). *Life-span developmental psychology: Introduction to research methods* (reprint of 1977 ed.). Hillsdale, NJ: Erlbaum.

Baltes, P.B., & Smith, J. (1999). Multilevel and systemic analyses of old age: Theoretical and empirical evidence for a fourth age. In V.L. Bengtson & K.W. Schaie (Eds.), *Handbook of theories of aging* (pp. 153–173). New York: Springer.

Baltes, P.B., & Smith, J. (2003). New frontiers in the future of aging: From successful aging to the dilemmas of the fourth age. *Gerontology*, *49*, 123–135.

Birren, J.E. (1964). *The psychology of aging.* Englewood Cliffs, NJ: Prentice-Hall.

Borsboom, D., Mellenbergh, G.J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *90*, 105–126.

Bosworth, H.B., & Schaie, K.W. (1999). Survival effects in cognitive function, cognitive style, and sociodemographic variables in the Seattle Longitudinal Study. *Experimental Aging Research*, *25*, 121–139.

Botwinick, J. (1977). Intellectual abilities. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 580–605). New York: Van Nostrand Reinhold.

Bryk, A.S., & Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*, 147–158.

Cattell, R.B. (1971). *Abilities: Their structure, growth, and action*. Boston, MA: Houghton Mifflin.

Christensen, H. (2001). What cognitive changes can be expected with normal ageing? *Australian and New Zealand Journal of Psychiatry*, *35*, 768–775.

Colsher, P.L., & Wallace, R.B. (1991). Longitudinal application of cognitive function measures in a defined population of community-dwelling elders. *Annals of Epidemiology*, *1*, 215–230.

Cunningham, W.R. (1980). Age comparative factor analysis of ability variables in adulthood and old age. *Intelligence*, *4*, 133–149.

Deary, I.J., Whiteman, M.C., Pattie, A., Starr, J.M., Hayward, C., Wright, A.F., Carothers, A., & Whalley, L.J. (2002). Cognitive change and the APOE 4 allele. *Nature*, *418*, 932.

Ekstrom, R.B., French, J.W., Harman, H.H., & Derman, D. (1976). *Manual for kit of factor-referenced cognitive tests*. NJ: Educational Testing Service.

Ghisletta, P., & de Ribaupierre, A. (2003). *Support for two-component theories of lifespan cognition from the Swiss Interdisciplinary Longitudinal Study on the Oldest-Old*. Unpublished manuscript.

Ghisletta, P., & Lindenberger, U. (2003). *Exploring the structural dynamics of the link between sensory and cognitive functioning in old age: Longitudinal evidence from the Berlin Aging Study.* Manuscript submitted for publication.

Ghisletta, P., & Lindenberger, U. (2003). Age-based structural dynamics between perceptual speed and knowledge in the Berlin Aging Study: Direct evidence for ability dedifferentiation in old age. *Psychology and Aging*, *18*, 696–713.

Ghisletta, P., & Lindenberger, U. (2004). Static and dynamic longitudinal structural analyses of cognitive changes in old age. *Gerontology*, *50*, 12–16.

Giambra, L.M., Arenberg, D., Zonderman, A.B., Kawas, C., & Costa, P.T., Jr. (1995). Adult life span changes in immediate visual memory and verbal intelligence. *Psychology and Aging*, *10*, 123–139.

Hamagami, F., & McArdle, J.J. (2001). Advanced studies of new linear dynamic systems models including group and individual differences. In G. Marcoulides & R. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 203–246). Hillsdale, NJ: Erlbaum.

Hamaker, E.L., Dolan, C.V., & Molenaar, P.C.M. (2003). *Statistical modeling of the individual: Rationale and application of multivariate time series analysis*. Manuscript submitted for publication.

Herlitz, A., Nilsson, L.-G., & Bäckman, L. (1997). Gender differences in episodic memory. *Memory and Cognition*, *25*, 801–811.

Heron, A., & Chown, S. (1967). *Age and function*. Boston: Little, Brown.

Hertzog, C.A. (1989). Influences of cognitive slowing on age differences in intelligence. *Developmental Psychology*, *25*, 636–651.

Hertzog, C.A. (1996). Research design in studies of aging and cognition. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 24–37). San Diego: Academic Press.

Hertzog, C.A., & Nesselroade, J.R. (2003). Assessing psychological change in adulthood. An overview of methodological issues. *Psychology and Aging*, *18*, 639–657.

Hillen, T., Lun, A., Reischies, F.M., Borchelt, M., Steinhagen-Thiessen, E., & Schaub, R.T. (2000). DHEA-S plasma levels and incidence of Alzheimer's disease. *Biological Psychiatry*, *47*, 161–163.

Hofer, S.M., Berg, S., & Era, P. (2003). Evaluating the interdependence of aging-related changes in visual and auditory acuity, balance, and cognitive functioning. *Psychology and Aging*, *18*, 285–305.

Hofer, S.M., & Sliwinski, M.J. (2001). Understanding ageing: An evaluation of research designs for assessing the interdependence of ageing-related changes. *Gerontology*, *47*, 341–352.

Horn, J.L. (1982). The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood. In F.I.M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 137–178). New York: Plenum Press.

Horn, J.L. (1989). Models of intelligence. In R.L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 29–73). Urbana, IL: University of Illinois Press.

Hultsch, D.F., Hertzog, C., Dixon, R.A., & Small, B.J. (1998). *Memory change in the aged*. New York: Cambridge University Press.

Hunt, W.L. (1949). The relative rates of decline of Wechsler-Bellevue "hold" and "don't hold" tests. *Journal of Consulting Psychology*, *13*, 440–443.

Jones, H.E., & Conrad, H.S. (1933). The growth and decline of intelligence: A study of a homogeneous group between the ages of ten and sixty. *Genetic Psychology Monographs*, *13*, 223–275.

Juan-Espinosa, M., Garcia, L.F., Escorial, S., Rebollo, I., Colom, R., & Abad, F.J. (2002). Age dedifferentiation hypothesis: Evidence from the WAIS III. *Intelligence*, *30*, 395–408.

Kalveram, K.T. (1965). Die Veränderung von Faktorenstrukturen durch "simultane Überlagerung" [Alterations in factorial structure through "simultaneous superimposition"]. *Archiv für die Gesamte Psychologie*, *117*, 296–305.

Kleemeier, R.W. (1962). Intellectual changes in the senium. *Proceedings of the American Statistical Association*, *1*, 290–295.

Kliegl, R., Smith, J., & Baltes, P.B. (1990). On the locus and process of magnification of age differences during mnemonic training. *Developmental Psychology*, *26*, 894–904.

Kunzmann, U., Little, T., & Smith, J. (2002). Perceiving control: A double-edged sword in old age. *Journals of Gerontology: Psychological Sciences*, *57B*, P484–P491.

Laird, N.M., & Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963–974.

Lang, F.R., Rieckmann, N., & Baltes, M.M. (2002). Adapting to aging losses: Do resources facilitate strategies of selection, compensation, and optimization in everyday functioning? *Journals of Gerontology: Psychological Sciences*, *57B*, P501–P509.

Lawley, D.N. (1943). A note on Karl Pearson's selection formulae. *Proceeding of the Royal Society of Edinburgh*, *62*, 28–30.

Li, K.Z.H., Lindenberger, U., Freund, A.M., & Baltes, P.B. (2001). Walking while memorizing: Age-related differences in compensatory behavior. *Psychological Science*, *12*, 230–237.

Li, S.-C., Jordanova, M., & Lindenberger, U. (1998). From good senses to good sense: A link between tactile information processing and intelligence. *Intelligence*, *26*, 99–122.

Li, S.-C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P.B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science*, *15*, 155–163.

Lindenberger, U. (2001). Lifespan theories of cognitive development. In N.J. Smelser & P.B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 8848–8854). Oxford, UK: Elsevier.

Lindenberger, U., & Baltes, P.B. (1994). Sensory functioning and intelligence in old age: A strong connection. *Psychology and Aging*, *9*, 339–355.

Lindenberger, U., & Baltes, P.B. (1995). Testing-the-limits and experimental simulation: Two methods to explicate the role of learning in development. *Human Development*, *38*, 349–360.

Lindenberger, U., & Baltes, P.B. (1997). Intellectual functioning in old and very old age: Cross-sectional results from the Berlin Aging Study. *Psychology and Aging*, *12*, 410–432.

Lindenberger, U., & Ghisletta, P. (2004). Modeling longitudinal changes in old age: From covariance structures to dynamic systems. In R.A. Dixon, L. Bäckman, & L.-G. Nilsson (Eds.), *New frontiers in cognitive aging.* Oxford: Oxford University Press.

Lindenberger, U., Gilbert, R., Little, T.D., Nuthman, R., Pötter, U., & Baltes, P.B. (1999). Sample selectivity and generalizability of the results from the Berlin Aging Study. In P.B. Baltes & K.U. Mayer (Eds.), *The Berlin Aging Study: Aging from 70 to 100* (pp. 56–82). Cambridge, UK: Cambridge University Press.

Lindenberger, U., Marsiske, M., & Baltes, P.B. (2000). Memorizing while walking: Increase in dual costs from young adulthood to old age. *Psychology and Aging*, *3*, 417–436.

Lindenberger, U., Mayr, U., & Kliegl, R. (1993). Speed and intelligence in old age. *Psychology and Aging*, *8*, 207–220.

Lindenberger, U., & Pötter, U. (1998). The complex nature of unique and shared effects in hierarchical linear regression: Implications for developmental psychology. *Psychological Methods*, *3*, 218–230.

Lindenberger, U., & Reischies, F.M. (1999). Limits and potentials of intellectual functioning in old age. In P.B. Baltes & K.U. Mayer (Eds.), *The Berlin Aging Study: Aging from 70 to 100* (pp. 329–359). Cambridge, UK: Cambridge University Press.

Lindenberger, U., Scherer, H., & Baltes, P.B. (2001). The strong connection between sensory and cognitive performance in old age: Not due to sensory acuity reductions operating during cognitive assessment. *Psychology and Aging*, *16*, 196–205.

Lindenberger, U., Singer, T., & Baltes, P.B. (2002). Longitudinal selectivity in aging populations: Separating mortality-associated versus experimental components in the Berlin Aging Study (BASE). *Journal of Gerontology: Psychological Sciences*, *57B*, P474–P482.

Lövdén, M., & Lindenberger, U. (in press). Development of intellectual abilities in old age: From age gradients to individuals. In O. Wilhelm & R.W. Engle (Eds.), *Understanding and measuring intelligence*. Thousand Oaks, CA: Sage.

MacDonald, S.W.S., Hultsch, D.F., Strauss, E., & Dixon, R.A. (2003). Age-related slowing of digit symbol substitution revisited: What do longitudinal age changes reflect? *Journals of Gerontology: Psychological Sciences*, *58B*, P187–P194.

MacFarland, R.A. (1968). The sensory and perceptual processes in aging. In K.W. Schaie (Ed.), *Theory and methods in research on aging* (pp. 9–52). Morgantown: West Virginia University.

Maier, H., & Smith, J. (1999). Psychological predictors of mortality in old age. *Journals of Gerontology: Psychological Sciences*, *54B*, P44–P54.

Marsiske, M., Klumb, P., & Baltes, M.M. (1997). Everyday activity patterns and sensory functioning in old age. *Psychology and Aging*, *12*, 444–457.

Mayer, K.U., Maas, I., & Wagner, M. (1999). Socioeconomic conditions and social inequalities in old age. In P.B. Baltes & K.U. Mayer (Eds.), *The Berlin Aging Study: Aging from 70 to 100* (pp. 227–255). Cambridge, UK: Cambridge University Press.

Mayr, U., & Kliegl, R. (2000). Complex semantic processing in old age: Does it stay or does it go? *Psychology and Aging*, *15*, 29–43.

McArdle, J.J. (2001). A latent difference score approach to longitudinal dynamic structural equation analyses. In R. Cudeck, S. duToit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 342–380). Lincolnwood, IL: Scientific Software International.

McArdle, J.J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R.W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*, 115–142.

McArdle, J.J., Hamagami, F., Elias, M.F., & Robbins, M.A. (1991). Structural modeling of mixed longitudinal and cross-sectional data. *Experimental Aging Research*, *17*, 29–51.

McArdle, J.J., Hamagami, F., Meredith, W., & Bradway, K.P. (2000). Modeling the dynamic hypotheses of Gf-Gc theory using longitudinal life-span data. *Learning and Individual Differences*, *12*, 53–80.

McArdle, J.J., Prescott, C.A., Hamagami, F., & Horn, J.L. (1998). A contemporary method for developmental-genetic analyses of age changes in intellectual abilities. *Developmental Neuropsychology*, *14*, 69–114.

Mitrushina, M., & Satz, P. (1991). Stability of cognitive functions in young-old versus old–old individuals. *Brain Dysfunction*, *4*, 174–181.

Molenaar, P.C.M., Huizenga, H.M., & Nesselroade, J.R. (2003). The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In U.M. Staudinger & U. Lindenberger (Eds.), *Understanding human development. Dialogues with lifespan psychology* (pp. 339–360). Dordrecht, NL: Kluwer Academic Publishers.

Nesselroade, C.S., & Schmidt McCollam, K.M. (2000). Putting the process in development processes. *International Journal of Behavioral Development*, *24*, 295–300.

Nesselroade, J.R. (2001). Intraindividual variability in development within and between individuals. *European Psychologist*, *6*, 187–193.

Nilsson, L.-G., Bäckman, L., Erngrund, K., Nyberg, L., Adolfsson, R., Bucht, G., Karlsson, S., Widing, M., & Winblad, B. (1997). The Betula prospective cohort study: Memory, health, and aging. *Aging, Neuropsychology, and Cognition*, *4*, 1–32.

Nilsson, L.-G., Sikström, C., Adolfsson, R., Erngrund, K., & Nylander, P.-O. (1996). Genetic markers associated with high versus low performance on episodic memory tasks. *Behavior Genetics*, *26*, 555–562.

Nyberg, L., Maitland, S.B., Rönnlund, M., Bäckman, L., Dixon, R.A., Wahlin, Å., & Nilsson, L.-G. (2003). Selective adult age differences in an age-invariant multi-factor model of declarative memory. *Psychology and Aging*, *18*, 149–160.

Park, D.C., Lautenschlager, G., Hedden, T., Davidson, N.S., Smith, A.D., & Smith, P.K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*, 299–320.

Pearson, K. (1903). Mathematical contributions to the theory of evolution: XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society of London (Series A)*, *200*, 1–66.

Powers, E.A., & Bultena, G.L. (1972). Characteristics of deceased dropouts in longitudinal research. *Journal of Gerontology*, *27*, 350–353.

Rabbitt, P.M.A., Diggle, P., Smith, D., Holland, F., & McInnes, L. (2001). Identifying and separating the effects of practice and of cognitive ageing during a large longitudinal study of elderly community residents. *Neuropsychologia*, *39*, 532–543.

Reinert, G. (1970). Comparative factor analytic studies of intelligence throughout the life span. In L.R. Goulet & P.B. Baltes (Eds.), *Life-span developmental psychology: Research and theory* (pp. 476–484). New York: Academic Press.

Reinert, G., Baltes, P.B., & Schmidt, L.R. (1966). Kritik einer Kritik der Differenzierungshypothese der Intelligenz [Critique of a critique of the differentiation hypothesis of intelligence]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *13*, 602–610.

Reischies, F.M., Schaub, R.T., & Schlattmann, P. (1996). Normal ageing, impaired functioning and senile dementia: A mixture distribution analysis. *Psychological Medicine*, *26*, 785–790.

Riegel, K.F., & Riegel, R.M. (1972). Development, drop, and death. *Developmental Psychology*, *6*, 306–319.

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2002). *Stability, improvement, and decline in adult life-span development of declarative memory: Cross-sectional and longitudinal data from a population-based study.* Manuscript submitted for publication.

Rubin, D. (1974). Characterizing the estimation of parameters in incomplete-data problems. *Journal of American Statistical Association*, *69*, 467–474.

Salthouse, T.A. (1991). *Theoretical perspectives on cognitive aging*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Salthouse, T.A. (1993). Speed and knowledge as determinants of adult age differences in verbal tasks. *Journal of Gerontology*, *48*, 29–36.

Salthouse, T.A. (2000). Methodological assumptions in cognitive aging research. In F.I.M. Craik & T.A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 467–498). Mahwah, NJ: Lawrence Erlbaum Associates.

Salthouse, T.A. (2003). Interrelations of aging, knowledge, and cognitive performance. In U.M. Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with lifespan psychology* (pp. 265–288). Kluwer Academic Publishers.

Salthouse, T.A., Hambrick, D.Z., & McGuthry, K.E. (1998). Shared age-related influences on cognitive and noncognitive variables. *Psychology and Aging*, *13*, 486–500.

Salthouse, T.A., McGuthry, K.E., & Hambrick, D.Z. (1999). A framework for analyzing and interpreting differential aging patterns: Application to three measures of implicit learning. *Aging, Neuropsychology, and Cognition*, *6*, 1–18.

Schaie, K.W. (1988). Internal validity threats in studies of adult cognitive development. In M.L. Howe & C.J. Brainard (Eds.), *Cognitive development in adulthood: Progress in cognitive development research* (pp. 241–272). New York: Springer.

Schaie, K.W. (1994). The course of adult intellectual development. *American Psychologist*, *49*, 304–313.

Schaie, K.W. (1996). *Intellectual development in adulthood. The Seattle Longitudinal Study.* New York: Cambridge University Press.

Schaie, K.W., Baltes, P., & Strother, C.R. (1964). A study of auditory sensitivity in advanced age. *Journal of Gerontology*, *19*, 453–457.

Schaie, K.W., Maitland, S.B., Willis, S.L., & Intrieri, R.C. (1998). Longitudinal invariance of adult psychometric ability factor structures across 7 years. *Psychology and Aging*, *13*, 8–20.

Schmiedek, F., & Li, S.-C. (2004). Toward an alternative representation for disentangling age-associated differences in general and specific cognitive abilities. *Psychology and Aging*, *19*, 40–56.

Schneider, B.A., & Pichora-Fuller, M.K. (2000). Implications of perceptual deterioration for cognitive aging research. In F.I.M. Craik & T.A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 155–219). Mahwah: Lawrence Erlbaum Associates.

Singer, T., Lindenberger, U., & Baltes, P.B. (2003a). Plasticity of memory for new learning in very old age: A story of major loss? *Psychology and Aging*, *18*, 306–317.

Singer, T., Verhaeghen, P., Ghisletta, P., Lindenberger, U., & Baltes, P.B. (2003b). The fate of cognition in very old age: Six-year longitudinal findings in the Berlin Aging Study (BASE). *Psychology and Aging*, *18*, 318–331.

Sliwinski, M., Lipton, R.B., Buschke, H., & Stewart, W. (1996). The effects of preclinical dementia on estimates of normal cognitive functioning in aging. *Journals of Gerontology: Psychological Science*, *50B*, P162–P170.

Small, B.J., Fratiglioni, L., von Strauss, E., & Bäckman, L. (2003). Terminal decline and cognitive performance in very old age: Does cause of death matter? *Psychology and Aging*, *18*, 193–202.

Smith, J., & Baltes, P.B. (1997). Profiles of psychological functioning in the old and oldest old. *Psychology and Aging*, *12*, 458–472.

Steinhagen-Thiessen, E., & Borchelt, M. (1999). Morbidity, medication, and functional limitations in very old age. In P.B. Baltes & K.U. Mayer (Eds.), *The Berlin Aging Study: Aging from 70 to 100* (pp. 131–166). Cambridge, UK: Cambridge University Press.

Verhaeghen, P., & Marcoen, A. (1996). On the mechanisms of plasticity in young and older adults after instruction in the method of loci: Evidence for an amplification model. *Psychology and Aging*, *11*, 164–178.

Verhaeghen, P., Marcoen, A., & Goosens, L. (1992). Improving memory performance in the aged through mnemonic training: A meta-analytic study. *Psychology and Aging*, *7*, 242–251.

Verhaeghen, P., & Salthouse, T.A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and non-linear age effects and structural models. *Psychological Bulletin*, *122*, 231–249.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York: The Psychological Corporation.

Wechsler, D. (1982). Handanweisung zum Hamburg-Wechsler Intelligenztest für erwachsene (HAWIE) (Manual for the Hamburg-Wechsler Intelligence Test for adults). Bern, Switzerland: Huber.

Wechsler, D. (1997). *WAIS-III: Wechsler Adult Intelligence Scale. Administration and scoring manual* (3rd ed.). San Antonio: The Psychological Corporation.

Wilson, R.S., Beckett, L.A., Barnes, L.L., Schneider, J.A., Bach, J., Evans, D.A., & Bennett, D.A. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging*, *17*, 179–193.

Zelinski, E.M., & Burnight, K.P. (1997). Sixteen-year longitudinal and time lag changes in memory and cognition in older adults. *Psychology and Aging*, *3*, 503–513.