2. EMPIRICAL INVESTIGATIONS

2.1 Overview

The empirical aspect of this thesis has two major goals. One is to examine the applicability and usefulness of the NF model in representing multivariate individual differences in cognitive abilities in old and very old age. This is pursued by applying the model to the empirical data from the Berlin Aging Study (Baltes & Mayer, 1999) and by a Monte Carlo simulation study that investigates the performance of the NF model under various controlled conditions of simulated data sets with known true parameter values.

The second goal is to explore and evaluate the importance of specific group factors within the dedifferentiated factor space observed in old age. This evaluation includes comparisons of different factor models in terms of model fit, communalities of the factors, ability of the models to capture the dimensionality of age-related differences using different conceptualizations of the age relations, and predictive relations of the multivariate cognitive space to selected external criterion variables from other domains. Before a detailed description of the variables and analyses, a brief review of BASE and related published analyses is given.

2.1.1 About the Berlin Aging Study

The Berlin Aging Study (BASE; Baltes & Mayer, 1999) is a multidisciplinary longitudinal study of the old and oldest old. It involved a multidisciplinary collaboration of the scientific fields of internal medicine and geriatrics, psychiatry, psychology, and sociology. The measures included in this study were selected to represent a systemic–wholistic view of psychological functioning in old and very old age (Baltes & Smith, 1997). They range from detailed descriptions of family and employment history, current economic situation and activities, medical anamnesis, the inclusion of careful and detailed physical, psychiatric, neuropsychological, and dental examinations, to a broad selection of psychological variables from the domains of social relationships, self and personality, and intellectual functioning. The study started with a large cross-sectional sample, stratified by age and gender, with an age range of 70 to over 100 years. At present, the longitudinal follow-ups include five intake assessments and five intensive protocols over a period of about eight years.

2.1.1.1 Previous Analyses of BASE Cognitive Ability Data With Hierarchical Models

In earlier publications on the structure of the cognitive abilities measured in BASE, hierarchical models have been used (Lindenberger & Baltes, 1997). Results showed that due to high correlations of the first-order factors of reasoning, speed, memory, fluency, and knowledge, a strong second-order general factor could be included in the model. The loadings of the first-order factors on this general factor were very high, with magnitudes above .90. The fact that most of the variance in cognitive functioning could be explained by one general factor obviously can be seen as cross-sectional support for the general process accounts. However, it has also been demonstrated that the loadings on the general factor still were significantly smaller than 1.00. It was acknowledged that this implied that each of the first-order factors contained reliable specific variance—independent of the general factor[2]. To this end, Lindenberger and Baltes (1997) offered the interpretation that age-based dedifferentiation operates at the second-order level while preserving the existence and configuration of first-order abilities (Lindenberger & Baltes, 1997, p. 419).

Relating the hierarchical model structure to the age variable indicated a strong age relation of the general factor and a weaker additional age relation of the first-order perceptual speed factor, both mediated by sensory–sensorimotor functioning (Lindenberger & Baltes, 1997). No additional direct or indirect paths from age on any of the other cognitive first-order factors were found to be significant. Analyses of six-year longitudinal information from BASE on a subset of the cognitive tasks indicated multidirectionality of the age changes, with perceptual speed declining faster than memory and fluency, and knowledge remaining fairly stable before exhibiting a dramatic decline above age 95 (Singer, Verhaeghen, Ghisletta, Lindenberger, & Baltes, 2003).

The cognitive battery of BASE has also been applied to a younger adult sample (i.e., 25–69 years; Baltes & Lindenberger, 1997). In this study, Baltes and

---

[2] It is important to note here that the residual variance in the first-order factors—after accounting for the general factor—is not error variance. Error variance is already removed at the level of indicator variables by their variable-specific residual terms. It follows that about one fifth of the first-order factor variance is reliable specific variance ($1 - .90^2 = .19$).

Lindenberger showed that the correlations among the five ability constructs and sensory functioning were higher in the BASE sample than in the younger adult sample. These analyses, however, did not include structural models providing more detailed descriptions of the relative importance of general versus specific sources of variance at different ages.

So far, there has not been a systematic examination on the importance that specific group factors might have for the description of individual differences in cognitive functioning in old age. Therefore, the central theme of the empirical investigation presented here is to use the NF model to examine the role specific factors play within the structure of cognitive abilities in an unbiased way. Furthermore, the importance of specific factors for relating the multivariate space of cognitive abilities to age and to theoretically important criterion variables is also examined.

2.1.2 Motivations for the Selection of External Criterion Variables

The following section introduces the external criterion variables that were chosen from BASE to analyze the relative contributions of general and specific factors. These variables were selected because relations to general and specific intelligence factors could be expected based on the following rationales derived from theoretical considerations and previous findings.

2.1.2.1 Education and Socioeconomic Status

Analyses of cross-sectional BASE data showed a strong effect of education and other indexes of socioeconomic status on mean level of general intelligence, with no differences of the observed age gradients between groups of high versus low socio–economic status (SES) (Lindenberger & Baltes, 1997). It was of interest to examine the relation of SES to intelligence more closely by using the NF model. It could be expected that SES, and especially education, would show significant relations to the specific knowledge factor—in addition to an effect of the general factor—because of the major role that schooling and professional education play in the acquisition of knowledge.

2.1.2.2 *Personality Factors*

Research on the interrelations and interfaces between personality traits and cognitive abilities has a long tradition (e.g., Ackerman, 1996; Cattell, 1971; Eysenck, 1994; see also Sternberg & Ruzgis, 1994, for a collection of theoretical approaches). In BASE, short scales of the three major personality constructs of neuroticism, extraversion, and openness to experience (Costa & McCrae, 1985; McCrae & Costa, 1997) were included. In a meta-analysis on the overlap of intelligence, personality, and interests, Ackerman and Heggestad (1997) reported only a weak average positive correlation of intelligence measures with extraversion and weak negative correlation with neuroticism. However, there was a medium-sized correlation between general intelligence and openness (r = .33), which seemed to be predominantly due to a correlation with the crystallized aspects of general intelligence (r = .30 with *Gc*; r = .08, n.s., with *Gf*). The openness trait describes a tendency of "proactive seeking and appreciation of experience for its own sake; toleration for and exploration of the unfamiliar" (Costa & McCrae, 1985) and can be subdivided into the facets of openness to fantasy, aesthetics, feelings, actions, ideas, and values. Why does this construct correlate with ability measures of crystallized intelligence?

A key to explaining this relation lies in the finding that openness is strongly related to the *Typical Intellectual Engagement* scale (TIE; Goff and Ackerman, 1992; Rocklin, 1994), which is a composite measure of dispositions to engage in abstract thinking, problem-oriented thinking, and reading. TIE has repeatedly been shown to correlate more highly with *Gc* than *Gf*. Ackerman and Goff (1994) showed that the relation of TIE and Openness is mainly produced by a strong correlation of TIE with the O*penness to ideas* facet. This facet mainly describes intellectual curiosity and the tendency to think theoretically and analytically about issues (Costa & McCrae, 1985). The *Understanding* subscale of the *Personality Research Form* (Jackson, 1984), another measure of intellectual curiosity, is also strongly related to *Gc*, and only moderately to *Gf* (Ashton, Lee, & Vernon, 2000). It seems, therefore, that intellectual interest, which has been found to be a stable trait characteristic, manifests itself in the acquisition of crystallized abilities. Based on the investment theory of intelligence (Cattell, 1971) and supported by previous findings on younger adult samples (Ackerman & Heggestad, 1997), it could be predicted that openness to ideas should be related to general intelligence as well as to the specific factor of knowledge in old age. Elderly people have lived through a lifetime of opportunities to accumulate knowledge. Therefore, the moderating effect of intellectual curiosity for the investment of *Gf* into *Gc* should manifest in a correlation of knowledge and openness to ideas in the BASE sample.

Cross-sectional and longitudinal evidence has consistently shown declines in openness to experience across adulthood (e.g., Costa, McCrae, Zonderman, Barabano, Lebowitz, & Larson, 1986; Costa & McCrae, 1988, 1997; McCrae et al., 1999) and moderate correlations of openness with years of education ($r = .28$ in a large national sample; see Costa et al., 1986). Therefore, the mutual relations among age, openness, and education also have to be considered in the analyses.

In BASE, the openness factor was measured by six items, of which only one represented the openness to ideas facet ("*I have a lot of intellectual curiosity*"). This marker item could be used, in comparison to the other openness items, to examine the relation of this construct and the intelligence factors of the NF model.

2.1.2.3 Sensory and Sensorimotor Functioning

The finding of a strong correlation between sensory and cognitive functioning (Baltes & Lindenberger, 1997; Lindenberger & Baltes, 1994, 1997) and the formulation of the common cause hypothesis have led to the application of SEM methods to test this hypothesis with mediational models that incorporate a *common cause factor* (Christensen et al., 2001). Christensen et al.'s finding of specific negative age-associated effects for visual functioning and grip strength—beyond a general effect through the common cause factor—was interpreted by these authors as evidence suggesting that more than one process operate concurrently. As discussed in the theoretical section, however, the method used by Christensen et al. is still prone to produce biased results. Therefore, it was of interest to conduct similar analyses with BASE data, using a version of the NF model that includes sensory and sensorimotor functioning as indicators of the general factor. To represent the common cause hypothesis, the NF model had to be adjusted to account for the theoretical three-stratum hierarchical structure as shown in Figure 4, panel A. This was accomplished by including a common cause factor with loadings of all variables—in addition to a general mechanic cognitive ability factor with loadings of all cognitive variables—and specific factors for each cognitive ability and for the three sensorimotor variables (Figure 4, panel B). This model allowed the simultaneous estimation of the amount of age-associated variance in each of the six latent factors.

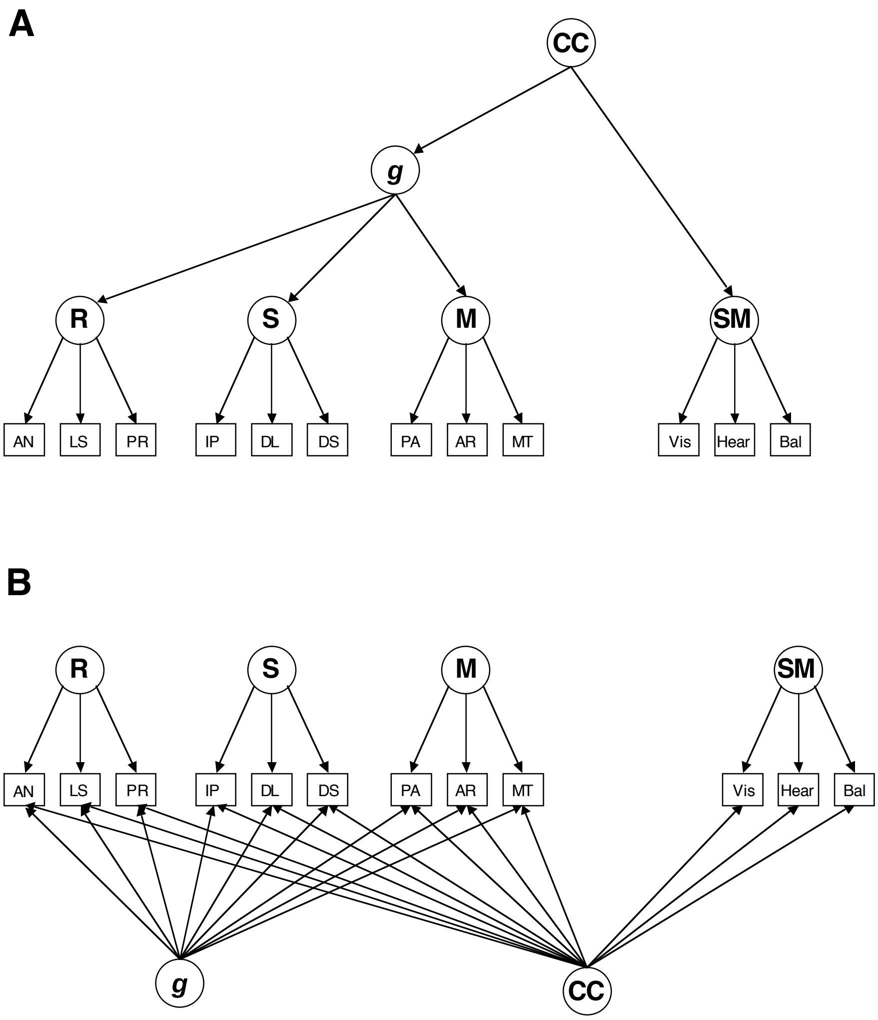

Figure 4. A: Model with a common cause (CC) factor on top of a hierarchical structure of the cognitive mechanics and variables of sensorimotor functioning. B: Nested factor representation of the hierarchical common cause factor model. *g* = general intelligence; R = reasoning; S = perceptual speed; M = memory; Hear = auditory acuity; Vis = visual acuity; Bal = balance/gait; SM = sensorimotor factor; AN = Analogies; LS = Letter Series; PR = Practical Problems; IP = Identical Pictures; DL = Digit Letter; DS = Digit Symbol; PA = Paired Associates; AR = Activity Recall; MT = Memory for Text.

2.1.2.4 *Everyday Activities*

In BASE, a broad assessment of everyday activities and capacities was also included. An index of basic competence (BaCo) could be derived from items of the activities of daily living (ADL) and instrumental activities of daily living (IADL) scales. In addition, a measure of expanded activities (ExCo) was build from information of the *Yesterday Interview* (M. M. Baltes, Mayr, Borchelt, Maas, & Wilms, 1993). M. M. Baltes et al. (1993) found BaCo to be more strongly related to health-related resources like mobility, while the psychological resources could better predict the expanded activity index, especially by using a fluid intelligence factor as a predictor. The large number of activities that have been coded in the *Yesterday Interview* (see Appendix D) allowed a further examination of this relationship between cognitive functioning and everyday activities. This was done by analyzing the relations of general and specific cognitive ability factors to a selected subset of activities that could be expected to be cognitively more challenging.

2.2 Research Hypotheses

Research questions and hypotheses concerning the empirical analyses of BASE data could be subdivided into three sets. Table 1 gives details of the hypotheses in these three complexes.

Table 1

Overview of Research Questions and Hypotheses

Existence of Specific Group Factors in a NF Structure

1. Given that is has been shown that specific variance of the first-order factors in a hierarchical model exists (Lindenberger & Baltes, 1997), the application of the NF model to the cognitive battery of BASE will show evidence of a strong general factor, but also significant and substantial contributions of specific group factors corresponding to the different ability constructs.

General and Specific Age Relations

2. Relating the NF structure to the chronological age variable will result in significant specific age-associated influences. It will not be possible to sufficiently explain the observed age correlations for all ability constructs by just an age relation of the general factor. Rather, specific factors will show a pattern of age relations that contrasts abilities that are differentially affected by cultural experiential factors or by biological factors.

Relations of the General and Specific Factors to External Criterion Variables

3. The specific factors will show relations to the external criterion variables of socioeconomic status, education, openness to experience, and expanded activities of daily living, independent of the general factor.
4. Does a general common cause factor mediate all age-related variance of the cognitive and sensorimotor variables?

2.3 Method

2.3.1 Sample

Analyses in this article are based on data from the first cross-sectional measurement occasion of BASE with the full intensive data protocol. This protocol comprised 14 sessions with data collection in the disciplines of internal medicine, psychiatry, sociology, and psychology. The sample was stratified by age and gender, with 43 women and 43 men in each of six age brackets (70–74, 75–79, 80–84, 85–89, 90–94, 95+). Sample size, therefore, was $N = 516$ (Age range: 70–103 years, M(Age) = 84.9 years). Detailed information regarding BASE in terms of study design (Baltes, Mayer, Helmchen, & Steinhagen-Thiessen, 1999), selectivity effects and generalizability issues (Lindenberger, Gilberg, Little, Nuthmann, Pötter, & Baltes, 1999), and on prevalence and effects of dementia in the sample (Lindenberger & Reischies, 1999) have been well documented elsewhere (Baltes & Mayer, 1999). For a broad collection of results from analyses of the cognitive battery in BASE, see Lindenberger and Baltes (1997).

To extend the analyses comparing the nested factor structure of two subsamples of BASE split at the age median into the context of covariance dedifferentiation across the adult lifespan, a comparison sample of younger adults (Age range: 25–69 years, M(Age) = 48.2 years, $SD = 14.7$, 58% women) was also used (for further information on this sample, see Baltes & Lindenberger, 1997).

2.3.2 Cognitive Tasks

The BASE cognitive battery comprised 14 tasks measuring the five ability constructs of: (a) reasoning (Analogies [AN], Letter series [LS], and Practical Problems [PR]); (b) perceptual speed (Digit Letter [DL], Digit Symbol Substitution [DS], and Identical Pictures [IP]); (c) memory (Activity Recall [AR], Memory for Text [MT], and Paired Associates [PA]); (d) fluency (Category: *Animals* [CA], and Word Beginnings: *Letter S* [WB]); (e) knowledge (Practical Knowledge [PK], Spot-a-Word [SW], and Vocabulary [VC]). A detailed description of the tasks can be found in Lindenberger, Mayr, and Kliegl (1993), while information on procedures, task administration, and psychometric properties for these tasks in the BASE sample is provided in Lindenberger and Baltes (1997).

2.3.3 External Criterion Variables

2.3.3.1 Socioeconomic Status

Socioeconomic status (SES) was modeled as a latent variable with the indicator variables income, occupational prestige, social class index, and education (see Lindenberger & Baltes, 1997; Mayer, Maas, & Wagner, 1999, for more information on these variables). The latent factor model had very good fit with $\chi^2[2] = 2.26$, RMSEA = .016, CFI = 1.00; the standardized factor loadings were .70 for social class, .38 for income, .78 for occupational prestige, and .67 for education.

2.3.3.2 Openness for Experience

Openness for experience was measured by a six-item scale with items taken from the NEO (Costa & McCrae, 1985). The items could be classified as belonging to the openness facets *fantasy*, *actions*, *aesthetics*, and *ideas* (see Table 4, for the items and facets). The total scale including all these items, however, had only low internal consistency with Cronbach's $\alpha = .54$.

2.3.3.3 Activity Competence and Engagement

A basic activity competency index (BaCo) was created as an aggregate of theparticipants' subjective ratings of the activities of daily living items (ADL; Katz, Ford, Moskowitz, Jackson, & Jaffee, 1963), and two items from the instrumental activities of daily living (IADL; Lawton & Brody, 1969) index (see M. M. Baltes et al., 1993). Additional information about an expanded scope of engagement in activities came from the *Yesterday Interview*, a modified version of the original version by Moss and Lawton (1982; see M. M. Baltes et al., 1993, for details on the interview procedure). Responses of this interview were coded into 44 specific activity categories (see Appendix D for a list of those categories). For the analyses of these activity codings, data of only 485 participants were used. The remaining 31 subjects were excluded because their recollection of the activities from the previous day was questionable, which seemed mainly to be due to the presence of dementia.

Of the 44 coded activity durations, a subset had to be chosen for the analyses relating activity to the NF model, because only a small number of activities showed reliable correlation patterns to the cognitive ability tasks. If no such pattern could be observed, it would not make sense to apply SEM models to try to explain observed covariance patterns of these variables. The following, empirically driven, selection procedure was used. Of all 44 activities, only those that had positive correlations with at least three of the 14 cognitive tasks at a significance level of $p < .01$ were chosen (Table D1 in Appendix D provides all correlations between the activity categories and the cognitive tasks). Of the ten tasks that met this criterion, two additional ones had to be excluded (*Regular paid work* and *Helping other person*), because they had extremely little variance, i.e., of the 485 subjects, more than 470 subjects had a value of zero. The eight remaining activity duration categories referred to basic care, shopping, craft/needlework, self-treatment, writing, phone calls and active and passive locomotion.

2.3.3.4 Sensory and Motor Performance Indicators

Sensorimotor performance was measured by three composite variables: an aggregate of auditory acuity in several frequency bands in both ears, an aggregate of close and distance visual acuity averaged over left and right eye, and a composite of the *Romberg stance* and the *Turn 360* tasks (see Lindenberger & Baltes, 1997, for additional information on these measures).

2.3.4 Analyses and Structural Equation Models

2.3.4.1 Overview of Analyses

In a first step, the NF model was fitted to the whole BASE sample from the first cross-sectional measurement occasion. Then, multiple group models for the NF structure were conducted to compare (a) two subsamples of BASE (i.e., split at the age median), and (b) the full BASE sample with the younger adult sample. A second series of analyses was carried out to relate the chronological age variable to the factors of the NF model. In these analyses, age was conceptualized either as an independent variable, correlate, or dependent variable. To better understand the complex pattern of age relations that resulted from these analyses, an oblique factor model was also used to examine the age correlations of correlated first order factors and their regression weights in predicting chronological age. A third set of analyses examined the relations of the general and specific factors to the external criterion variables.

2.3.4.2 *Structural Equation Models*

Most of the presented models were estimated with PROC CALIS (SAS Institute, 1989a). The multiple group models were run with EQS 5.7b (Bentler & Wu, 1995), and bootstrap control analyses were conducted with AMOS 4 (J. Arbuckle, 2000). For all SEM models, the Maximum Likelihood minimization criterion was used, together with the Levenberg-Marquardt optimization technique of SAS PROC CALIS for most analyses. This optimization method, though computationally extensive, is usually more reliable than the quasi-Newton or Newton–Raphson procedures used by other SEM programs (SAS Institute, 1989a).

For the models using age as a correlate or dependent variable, the factor variances of the NF model were fixed to one, and all factor loadings were freely estimated. Because the factor variances could not be fixed in the models with age as a predictor variable, one of each factors' loadings was set to one. The two loadings of the specific fluency factor were constrained to be equal[3] to ensure identification of the factor.

For the analyses relating the NF model to external criterion variables, the measurement model was fixed, using the unstandardized solution of the NF model without external variables. Similar to the extension analysis in exploratory factor analysis, this approach constrained the factors to be the same across analyses with different criterion variables. If the measurement model were not fixed, the factor loadings might change considerably from one model to the other, because of differences in the relations of the indicator variables to the external criterion. This would make the interpretation and comparison difficult, and would provide a less conservative test of the predictive validities of the nested factors. However, in evaluating model fit, one has to adjust for the degrees of freedom gained by fixing parameters to values estimated from empirical data in preceding models (R. S. Burt, 1976).

---

[3] Constraining the unstandardized loadings of the two fluency tasks on their specific factor to be equal in order to achieve identification of the factor was seen as appropriate because the residual variance of both tasks after accounting for the general factor were of similar size. In separate analyses, constraining the two loadings to explain an equal portion of variance independent of the general factor led to very similar results. Nonetheless, it should be noted that interpretations of the results for the specific fluency factor should be more cautious than for the other factors.

2.4 Results

2.4.1 Measurement Model With the Nested Factor Structure

As a first step, the NF model for the five theoretical constructs of the psychometric battery was fitted to the total sample. For each of the 14 task variables, one path to a general, and one to its respective specific factor was specified. Figure 5 shows the measurement structure and standardized factor loadings for this model. Loadings on the general factor were all high, ranging from .61 to .84. The specific loadings were smaller, but still substantial with values from .20 to .49. The fit of this model was acceptable, with $\chi^2[64] = 145.3$, $p < .001$, RMSEA = .05, CFI = .98.

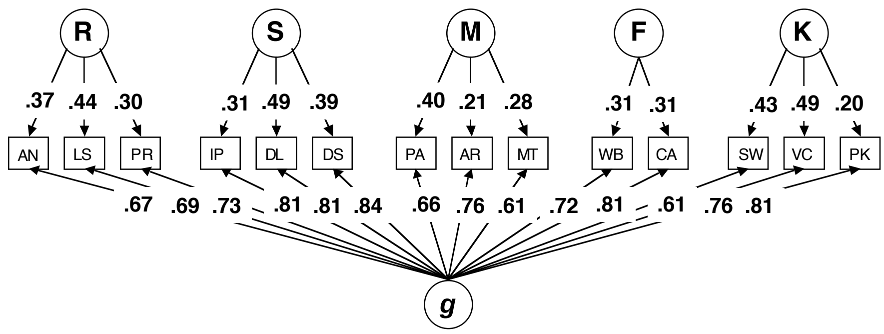

Figure 5. Nested factor model with BASE data (N = 516). g = General factor. R = Specific reasoning factor. S = Specific perceptual speed factor. M = Specific memory factor. F = Specific fluency factor. K = Specific knowledge factor. AN = Analogies; LS = Letter Series; PR = Practical Problems; SW = Spot-a-Word; VC = Vocabulary; PK = Practical Knowledge; IP = Identical Pictures; DL = Digit Letter; DS = Digit Symbol; WB = Word Beginnings; CA = Categories; PA = Paired Associates; AR = Activity Recall; MT = Memory for Text. All factor loadings are standardized estimates.

To test whether the specific factors were significant, separate analyses were conducted, leaving out one of the specific factors at a time. Each of these models with one general and the four remaining specific factors are nested within the full model, and therefore χ^2-difference tests could be calculated. These tests resulted in χ^2-differences of $\Delta\chi^2[3] = 77.6$ for the specific reasoning, $\Delta\chi^2[3] = 164.2$ for the specific speed, $\Delta\chi^2[3] = 28.6$ for the specific memory, $\Delta\chi^2[1] = 21.0$ for the specific fluency, and $\Delta\chi^2[3] = 97.1$ for the specific knowledge factors, respectively. All these tests were significant with $p < .001$. Furthermore, each of the factor loadings on the specific factors was significant with $p < .001$. Specifying a model with only a general and no specific factors resulted in unacceptable model fit (χ^2[77] = 687.4, $p < .001$, RMSEA = .12, CFI = .89), as did a model specifying only orthogonal group factors for the five theoretical constructs (χ^2[78] = 2002.4, $p < .001$, RMSEA = .22, CFI = .64).

A comparison of these results with the hierarchical models reported in Lindenberger and Baltes (1997) showed that with the NF model, it was not necessary to include correlated error terms for Digit Symbol and Digit Letter, and for Spot-a-word and Vocabulary, respectively[4]. The proportionality constraints implied by the hierarchical model necessitated these adjustments to the measurement model. The NF model is more flexible in accounting for such similarities at the indicator variable level, such that the empirical relations could be captured by the general and specific factors.

It was informative to calculate communalities for the general and specific factors to estimate their overall importance for the description of the multivariate space defined by the 14 cognitive tasks. Communalities were computed by squaring and summing the standardized loadings on the different factors, and dividing the resulting sums by the number of variables. The communality for the general factor was h^2 = 54.8% and for all five specific together it was h^2 = 13.3%. This means that about one fifth of the total variance that could be explained by the six factors of the model was accounted for by the specific factors (cf. Lindenberger & Baltes, 1997).

The total communality of the NF model was also compared to the communality of an oblique factor model to examine whether the inclusion of a general factor did increase the total amount of variance explained. The total communality of the oblique factor model was 68.8%, indicating that this was not the case. Furthermore, an oblique factor model also showed acceptable fit (χ^2[68] = 184.5, $p < .001$, RMSEA = .06, CFI = .98), which could be further improved by allowing the same correlated residual terms as for the hierarchical model (χ^2[66] = 144.8, $p < .001$, RMSEA = .05, CFI = .99). The benefit of using a NF model, therefore, did not lie in explaining additional variance of the indicator variables, but rather in representing the structure in a way that allowed separating general from specific variance.

---

[4] Lindenberger and Baltes (1997) did include correlations of the error terms for these two pairs of variables to significantly improve model fit.

2.4.2 Age Differences in the Factor Structure

To test the assumption of the NF model structure being invariant across the age range of BASE (i.e., 70 to 103), the BASE sample was split into two halves (N = 258 each) at the age median of 85.15. Multiple group SEM methodology (Lee & Tsui, 1982) was used to test the NF model with parameters estimated separately for the two groups against a model with equality constraints across groups. The approach used here deviated from the standard procedure for testing invariance of factor structures, which proposes to run a sequence of models specifying configural, incomplete, and complete measurement invariance (Meredith, 1993). These models put an increasing number of constraints on the factor-loading matrix, the factor variance–covariance matrix, and the variance–covariance matrix of the unique or error terms. In the NF model, however, factor covariances are set to zero to begin with, as the factors are defined to be orthogonal to each other. In an oblique factor model, information about dedifferentiation is contained in the covariances of the factors. In the NF model, this information is carried by the relative importance of the general factor, as compared to the specific factors. Fixing the factor variances to one, the communalities of the general and specific factors, which are implied by the standardized loadings, provide information about the potential dedifferentiation of the structure.

The fit of the model with equality constraints on factor loadings across groups was good (χ^2[155] = 255.7; p < .001; CFI = .98; RMSEA = .04), and could only be slightly improved by releasing the equality constraints (χ^2[128] = 202.5; p < .001; CFI = .98; RMSEA = .03; Appendix E provides parameter estimates for both groups). A χ^2-difference test of the two models indicated that this difference was significant ($\Delta\chi^2$[27] = 53.15, p < .01). However, rescaling this difference to the metric of an RMSEA index (Browne & DuToit, 1992) resulted in a value of .04, which indicated that the differences between the NF structures of the two age groups were only minor.

Was the observed difference due to a general dedifferentiation of the structure or to some more complex pattern of differences in single factor loadings? This question was approached by calculating communalities of the general and specific factors in the different age groups. Table 2 contains this information. The communality of the general factor was 46.8% in the young-old, and 49.1% in the old-old group, while the total communality of the specific factors was 16.2% in the young-old, and 15.3% in the old-old group, an indication of only very slight dedifferentiation within the BASE sample. Comparing these results to the communalities of the total BASE sample, which were 54.8% for the general, and 13.3% for the sum of the specific factors, showed how a reduction of age variance reduces the relative strength of the general factor.

Why did there seem to be relative stability of the factor structure, as compared to other studies examining this issue? One explanation could be that most of the "neo-integration" of the factor space takes place from adulthood to old age (Lindenberger & Baltes, 1997), so that the factor space above the age of 70 is already very much condensed and further dedifferentiation is hard to detect. One way to address this hypothesis was to compare the results for the BASE sample with a NF representation of the same cognitive battery applied to the younger adult sample of the Baltes and Lindenberger (1997) study. As shown in Table 2, the communality of the general factor in the younger sample was only 24.9%, as compared to the total specific communality of 27.1%[5]. On average, the specific factors, therefore, explained more variance than the general factor in this sample, a clear demonstration of dedifferentiation from the second to the third and fourth age (see Appendix E for standardized factor loadings).

Table 2

Communalities of the General and Specific Factors in the Nested Factor Model for Different Age Groups

Age Group	*N*	h^2 (*g*)	h^2 (ΣSpecific)	h^2 (R)	h^2 (K)	h^2 (S)	h^2 (F)	h^2 (M)
20-69	171	.25	.27	.06	.07	.07	.04	.04
70-103	516	.55	.13	.03	.03	.04	.01	.02
70-85.1	258	.47	.16	.03	.03	.05	.03	.03
85.2-103	258	.49	.15	.04	.04	.05	.01	.03

Notes. h^2 = Communality; *g* = General factor; ΣSpecific = Sum of the specific factors; R = Specific factor of reasoning; K = Specific factor of knowledge; S = Specific factor of perceptual speed; F = Specific factor of fluency; M = Specific factor of memory.

Because the statistical test for measurement invariance of the NF model indicated only a slight difference in fit of the constrained and unconstrained models, and the

[5] Model fit of this model was very good with $\chi^2[64] = 76.3$; $p = .14$; CFI = .98; RMSEA = .03.

comparison of communalities did not indicate general dedifferentiation of the structure, it was concluded that the assumption of invariance of the NF model structure holds within the BASE sample to a satisfying degree. Analyses based on the whole sample were therefore possible.

2.4.3 General and Specific Age Relations of the Nested Factor Structure

2.4.3.1 Age-as-a-Predictor Model

Including the age variable into the model by specifying statistical causal paths from age to the general and all specific factors led to good model fit ($\chi^2[72] = 158.7$, $p < .001$, RMSEA = .05, CFI = .98). There was a strong age-associated effect on the general factor ($\gamma = -.69$, $Z = 8.97$). Additional significant effects were found for the specific factors of knowledge: ($\gamma = .56$, $Z = 3.96$), fluency ($\gamma = .48$, $Z = 2.78$), and memory ($\gamma = .31$, $Z = 2.01$)[6]. The positive effects for knowledge, fluency, and memory were of medium to strong effect sizes and counterbalanced the strong negative general age effect. Figure 6 shows the age-associated effects and the standardized factor loadings of this model.

2.4.3.2 Age-as-a-Covariate Model

A model with age included as a covariate of the orthogonal general and specific factors also showed good fit ($\chi^2[72] = 161.1$, $p < .001$, RMSEA = .05, CFI = .98). Significant correlations with age were $r = -.67$ ($Z = 13.73$) for the General factor, and $r = .33$ ($Z = 4.84$) and $r = .24$ ($Z = 2.46$) for the specific factors of knowledge and fluency[7], respectively. Figure 7 shows these correlations, together with the standardized factor loadings. While the pattern of a strong negative age correlation of the general factor and positive effects on the specific factors was very similar to

---

[6] A more appropriate test of the significance of individual parameters in a SEM model than that given by the approximate Z values is given by separate likelihood ratio (LR) test for each parameter. Such LR tests are invariant over different parameterizations of a model (Gonzalez & Griffin, 2001). Calculating those tests by comparing models with and without the parameter of interest included resulted in highly significant LR-tests with $\Delta\chi^2[1] = 31.0$ for the specific knowledge and $\Delta\chi^2[1] = 19.2$ for the specific fluency factor, but an only marginally significant test of $\Delta\chi^2[1] = 3.3$ for the specific memory factor.

[7] LR-tests for the specific age relations of the specific knowledge and fluency factors resulted in significant values of $\Delta\chi^2[1] = 29.6$ and $\Delta\chi^2[1] = 7.5$, respectively.

that observed with the age-as-a-predictor model, the specific effects were of smaller magnitude.

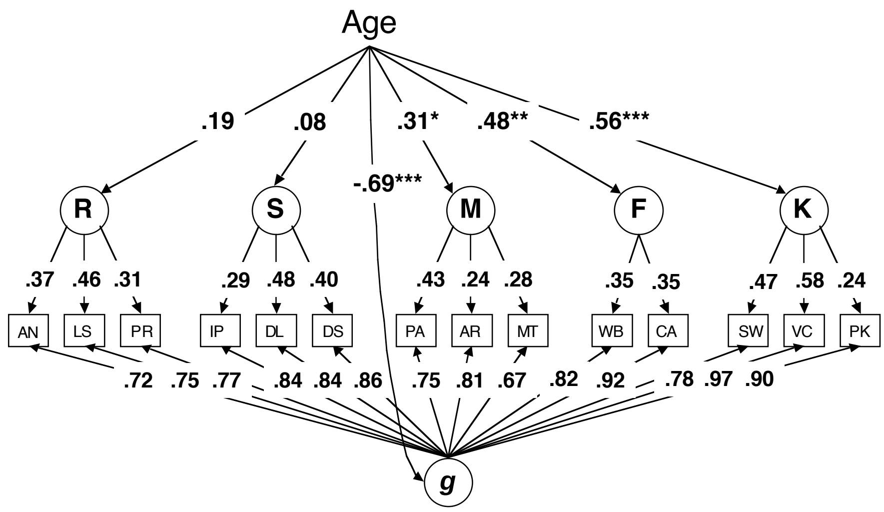

Figure 6. Directed age-associated effects on the factors of the nested factor model structure ($N = 516$). All factor loadings and regression weights are standardized estimates. g = General factor. R = Specific reasoning factor. S = Specific perceptual speed factor. M = Specific memory factor. F = Specific fluency factor. K = Specific knowledge factor. AN = Analogies; LS = Letter Series; PR = Practical Problems; SW = Spot-a-Word; VC = Vocabulary; PK = Practical Knowledge; IP = Identical Pictures; DL = Digit Letter; DS = Digit Symbol; WB = Word Beginnings; CA = Categories; PA = Paired Associates; AR = Activity Recall; MT = Memory for Text.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Appendix F provides detailed information on comparing the unstandardized factor loadings of the age-as-a-predictor model, the age-as-a-covariate model, and a measurement model based on age-partialled covariances. This comparison showed that, indeed, the factor patterns for the age-partialled and the age-as-a-predictor model were almost exactly the same, and that the difference between the age-as-a-covariate and the age-as-a-predictor model lay in different loadings on the general, but not on the specific factors.

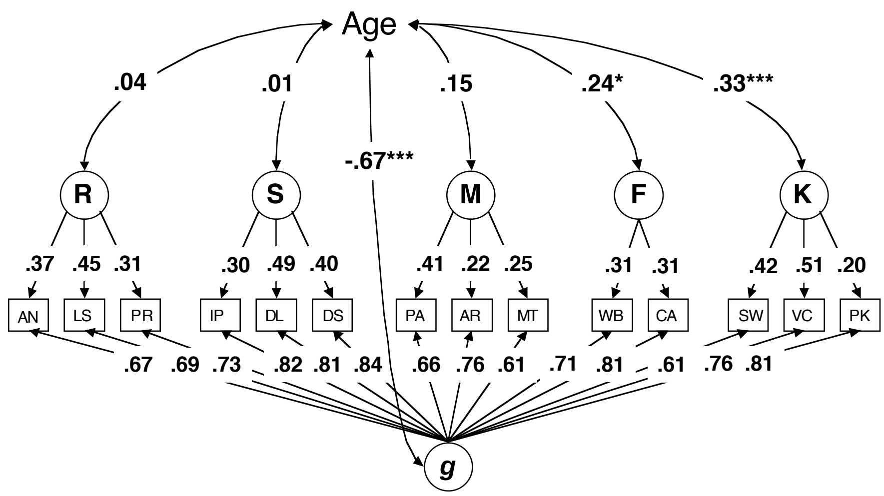

Figure 7. Age correlations of factors in the nested factor model (N = 516). All factor loadings and correlations are standardized estimates. g = General factor. R = Specific reasoning factor. S = Specific perceptual speed factor. M = Specific memory factor. F = Specific fluency factor. K = Specific knowledge factor. AN = Analogies; LS = Letter Series; PR = Practical Problems; SW = Spot-a-Word; VC = Vocabulary; PK = Practical Knowledge; IP = Identical Pictures; DL = Digit Letter; DS = Digit Symbol; WB = Word Beginnings; CA = Categories; PA = Paired Associates; AR = Activity Recall; MT = Memory for Text.
* $p < .05$, ** $p < .01$, *** $p < .001$.

2.4.3.3 *Age as a Dependent Variable*

Building on the last model with age as a covariate of the orthogonal general and specific factors, one can turn around the much-used equation with age as an independent variable and treat chronological age as the dependent variable in a regression analysis. In such an approach, age differences are predicted by the factors of the NF model. Using only the general factor as a predictor of age accounted for 36% of the observed age variance, while a model with the general and the five specific factors as predictors could explain 61%, an almost doubled amount of variance! To understand why the specific factors added such an impressive amount of predictive power, one has to consider that because they are orthogonal to the general factor and to each other, the square of their age correlations directly equals the amount of incremental variance explained. These results do not preclude the possibility that other predictors might serve equally well in predicting age differences. However, they strongly emphasize the importance of considering specific factors in addition to a general factor when the goal is to relate multivariate individual differences in cognitive functioning to individual differences in chronological age.

2.4.3.4 *Comparison With Oblique Factor Model*

The age correlations of the specific factors in the NF model can be viewed as necessary adjustments to an overly parsimonious model with only the general factor being related to age. Because the general factor in these models derives from the finding that the first-order factors in an oblique factor model were highly correlated, it is instructive to also use such an oblique factor model to predict chronological age as a comparison to the results for the NF model presented in the last section.

If correlated factors are used to predict age in a multiple regression analysis, the regression weights assigned to the factors differ considerably from the age correlations of these factors (see Figure 8). This is because multiple regression analysis provides the best linear combination of the factors by taking into account and adjusting for the pattern of common variances among all possible pairs of factors. To achieve this, the age-related portions of the common variance for all pairs of factors have to be distributed among the factors. Figure 8 shows that this resulted in a pattern of regression weights that was similar to the pattern obtained for the age correlations of the specific factors in a NF model. While in the oblique factor model the adjustments for common variance are reflected by the differences between age correlations and regression weights, in the NF model the age correlations of the specific factors adjust for the common variance that is captured by the general factor. However, a comparison of the amount of age variance that could be explained by the different models revealed that with the five oblique factors, only 40% of the age variance was explained. Furthermore, this model fitted worse than the NF model (χ^2[77] = 223.0, p < .001, RMSEA = .06, CFI = .97). This showed that the NF model allows for a representation of age-associated variance within a structure that captures the multidimensionality and multidirectionality of cognitive age differences substantially better than an oblique factor solution.

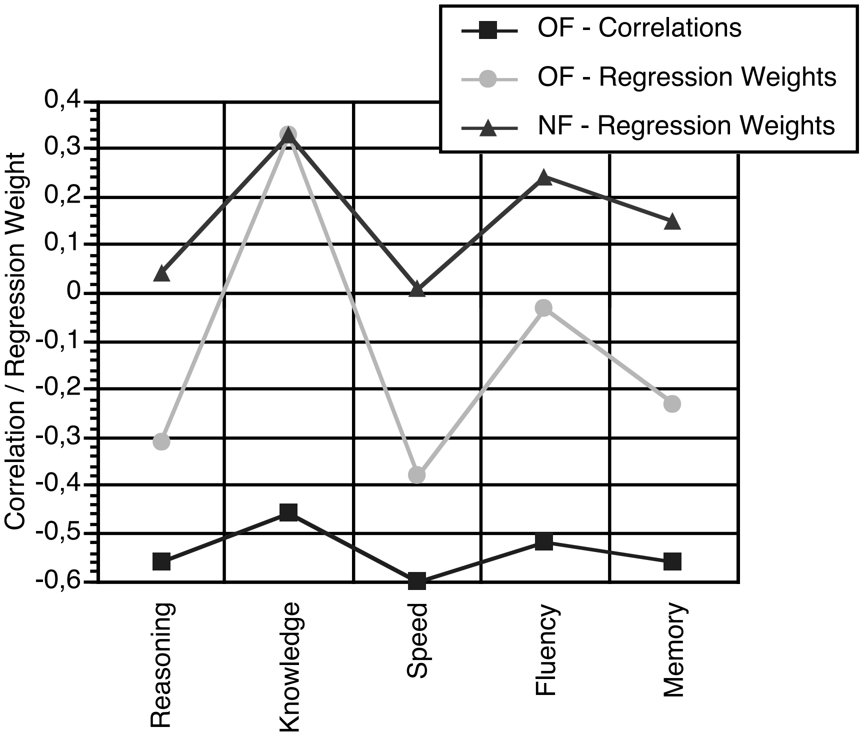

Figure 8. Age correlations of the factors in an oblique factor (OF) model, regression weights for a regression of age on oblique factors and for a regression on the specific factors in a nested factor (NF) model.

2.4.4 Summary of Nested Factor Models and Their Relations to Age

The results presented so far afford three major conclusions. First, there was reliable specific variance contained in each ability construct that was independent of a general factor. Second, the amount of specific variance was less in older adults (age 70–103) than in younger adults (age 25–69). However, there was no indication of further dedifferentiation within the age range of the BASE sample. These results obtained with the NF model replicate the findings of Baltes and Lindenberger (1997) and Lindenberger and Baltes (1997), who also reported the existence of specific first-order factor variance that accounted for about one fifth of the total reliable variance in a hierarchical model. Third, the effects of age clearly had to be conceptualized as multidirectional. This could be shown by conceptualizing age as a predictor variable as well as by correlating age with the general and specific factors, and by predicting age differences with the orthogonal factors of the NF structure. In each case, a multidimensional representation captured important additional information about age-related variance and was superior to an oblique factor representation. Analyses reported in the next section were conducted to examine whether this multi-dimensional representation provides an improved frame for examining relations to additional external variables other than age.

2.4.5 Relations of General and Specific Factors to External Criterion Variables

2.4.5.1 Socioeconomic Status and Education

As shown in Table 3, the general factor of the NF model was significantly related to all SES index variables. This led to a strong correlation of the general factor with the latent SES construct. Independent of this general relation, the SES factor and its indicators were also substantially correlated with the orthogonal specific factor of knowledge. Interestingly, years of education also had a significant positive relation to the specific reasoning factor that was independent of the correlation with the general factor. Comparisons of models including correlations of the specific factors with models that only contained a general factor correlation yielded significant χ^2-difference tests (see Table 3). Analyses with age partialled out of the NF model and the SES indicators also resulted in practically the same pattern of significant correlations (see Appendix G).

Table 3

Correlations of General and Specific Factors in the Nested Factor Model With Indicators and a Latent Factor of Socioeconomic Status

	r_{Age}	g	R	K	S	F	M	$\Delta\chi^2[5]^a$
Income	-.04	**.29*****	-.06	**.19***	.00	-.10	-.01	15.2
Social Class	-.01	**.28*****	-.01	**.26****	.00	.01	-.09	25.7
Occupational Prestige	-.08	**.38*****	.02	**.25****	-.04	.08	.00	23.1
Education	**-.14****	**.35*****	**.20***	**.28*****	.03	.11	-.04	34.6
Latent Construct of SES	-.09	**.48*****	.08	**.37*****	-.01	.07	-.05	41.4

Notes. [a] χ^2-difference test of model comparisons with models that only specified a correlation with the general factor. g = General factor; R = Specific factor of reasoning; K = Specific factor of knowledge; S = Specific factor of perceptual speed; F = Specific factor of fluency; M = Specific factor of memory. Correlations in boldface were significant with * $p < .05$, ** $p < .01$, *** $p < .001$.

2.1.1.2 *Openness for Experience*

The Openness scale had significant relations to the general factor and the specific factor of memory (Table 4). The relation to the specific knowledge factor, which could be expected based on the metaanalytic findings of Ackerman and Heggestad (1997), did not reach significance. The analysis of the relations of the NF structure to the individual items that went into the openness scale, however, provided further insight into the relation of openness to the cognitive ability factors. The *Intellectual curiosity* item, indeed, correlated with the general as well as the specific knowledge factor. The relation of the openness sum scale to the memory factor mostly reflects the significant correlation between this factor and the *Daydreaming* and *Poetry* items (Table 4). While it is difficult to explain the relation to the *Daydreaming* item, the finding of a positive attitude towards poetry being related to the specific factors of knowledge, fluency, and memory—and not to the general factor—is an example of how focusing only on general factor variance might miss interesting and meaningful relations.

Table 4

Correlations of General and Specific Factors in the Nested Factor Model With Items and Scale Score of Openness for Experience From the NEO

Item content (Facet)	r_{Age}	g	R	K	S	F	M
„I don't like to waste my time daydreaming" (Fantasy) [a]	**.12****	**-.15****	-.06	-.11	-.03	-.07	**.19***
„Once I find the right way to do something, I stick to it" (Actions) [b]	.05	**-.16****	.07	-.10	-.11	-.17	.03
„Poetry has little or no effect on me" (Aesthetics) [a]	.03	-.03	-.03	**.16***	.14	**.29****	**.29****
„I often try new and foreign foods" (Actions)	**-.22*****	**.25*****	.02	.03	.15	.18	.11
„I have a very active imagination" (Fantasy)	**-.15****	**.31*****	-.12	.13	-.05	.02	-.01
„I have a lot of intellectual curiosity" (Ideas)	**-.14****	**.33*****	.03	**.27*****	-.03	.04	-.01
Openness scale	**-.10***	**.19*****	-.04	.13	.04	.13	**.22***

Notes. [a] Item wording was changed to a positive statement in the German version. [b] Item was reflected in the total scale score. g = General factor; R = Specific factor of reasoning; K = Specific factor of knowledge; S = Specific factor of perceptual speed; F = Specific factor of fluency; M = Specific factor of memory. Correlations in boldface were significant with * $p < .05$, ** $p < .01$, *** $p < .001$.

Because openness has been found to be related to education in other studies (McCrae & Costa, 1997), it was of interest to examine to what degree the observed relations of education and openness with the cognitive factors were shared among these two constructs. The correlation of years of education with the intellectual curiosity item was $r = .23$ ($p < .001$). Therefore, both variables accounted for largely independent portions of variance in the general and the specific knowledge factor. The finding of a significant relation of education, but not of the intellectual curiosity item, to the specific reasoning factor provided further evidence of both variables capturing different aspects of antecedent influence on the development of individual differences in cognitive functioning. The pattern of significant correlations presented in Table 4 was not changed when age was partialled out by using a NF model with age as a predictor (see Appendix H).

2.1.1.3 Sensory and Sensorimotor Functioning

The NF model for the cognitive mechanics and the sensorimotor variables, with one common cause factor, one general factor of cognitive mechanics, and the four specific factors (see Figure 9) had good fit ($\chi^2[33] = 55.8$, $p < .01$, RMSEA = .04, CFI = .99). The sensorimotor variables had high loadings on the general, but also substantial loadings on their specific factor. Introducing age as a correlate of all factors into the model resulted in a strong and significant correlation of $r = -.60$ ($Z = 15.30$) with the common cause factor, and an even stronger correlation of $r = -.70$ ($Z = 9.68$) with the specific sensorimotor factor, while the age correlations of the general and specific cognitive factors were not significant. Model fit was very good ($\chi^2[40] = 51.3$, $p = .11$, RMSEA = .02, CFI = 1.00). This showed that, even though the common cause factor could not account for all age-associated variance in the sensorimotor variables, it captured all age-associated variance of the cognitive variables. The strong loadings of the sensorimotor variables on this common cause factor support the notion of a coupling of cognitive and sensory functioning in old age.

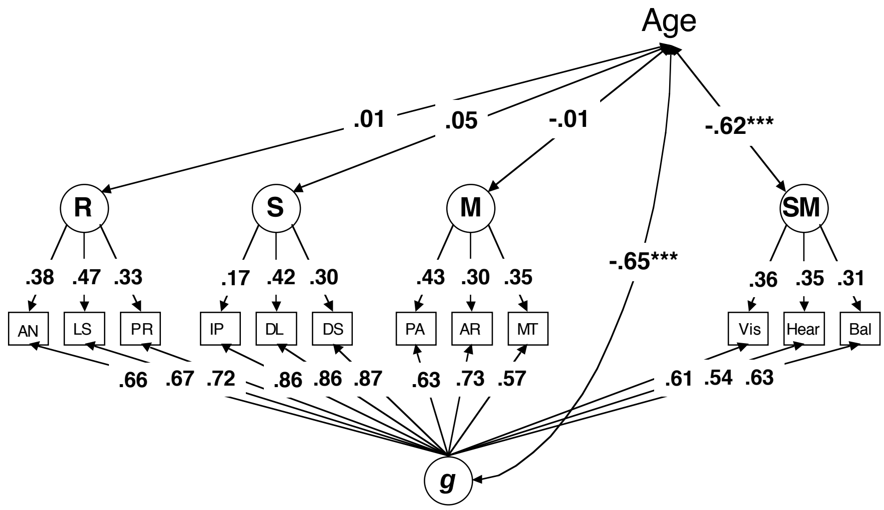

Figure 9. Common cause factor model of reasoning (R), speed (S), memory (M), and visual acuity (Vis), auditory acuity (Hear), and Balance/Gait (Bal) as indicators of a sensorimotor (SM) factor. All parameter values are standardized estimates. AN = Analogies; LS = Letter Series; PR = Practical Problems; IP = Identical Pictures; DL = Digit Letter; DS = Digit Symbol; PA = Paired Associates; AR = Activity Recall; MT = Memory for Text.
* $p < .05$, ** $p < .01$, *** $p < .001$.

2.1.1.4 *Basic and Expanded Activities*

The seven selected activities were related to the NF model by predicting these activities with the general and specific factors, while controlling for an index of basic competency (BaCo), which is a composite of the six ADL and two IADL items (see M. M. Baltes et al., 1993). The BaCo index was included in the prediction model because previous analyses have shown that BaCo was a good predictor of the expanded competency (ExCo) sum score (M. M. Baltes et al., 1993; Marsiske, Klumb, & M. M. Baltes, 1997). Therefore, it was of interest to show the predictive validity of the cognitive factors beyond the part of variance that could be explained by the basic constraints in functioning captured by the BaCo index.

Results of the SEM models showed that for the activities of basic care, shopping, and craft/needlework, none of the general and specific cognition factors had a significant latent regression weight. For the duration of self-treatment activities, however, a significant correlation with the general factor of $r = .30$ ($Z = 4.07$) was found. The activity of making phone calls was only significantly related the specific memory factor ($r = .25$, $Z = 2.38$). Passive and active locomotion also could—to some degree—be predicted by the cognitive ability factors. Active locomotion—walking, driving, or riding a bicycle—had a significant relation to the general factor (r = .18, Z = 2.49) and to the specific factor of reasoning (r = .19, Z = 2.28). Passive locomotion—public transport or being a passenger in someone else's car—was positively related to BaCo (r = .16, Z = 2.61) and had a negative correlation with the specific knowledge factor (r = -.22, Z = 2.81). While the ability to drive a car or bicycle could be expected to be related to the attentional capacities that are captured by the factors of general cognitive ability and reasoning, these results cannot be unambiguously interpreted without further information about which specific ways of transportation were used. This also applies to passive locomotion, where the need to take public transport might be a consequence of lower social status, which, in turn, has already been shown to be related to the specific factor of knowledge.

Very interesting were the results for the activity category of *Writing*. As can be seen in Figure 10, this activity was not significantly related to the BaCo index. While BaCo could be well predicted by the general factor, this factor did not contribute significantly to the prediction of writing activities. However, there were significant positive prediction weights for the specific factors of speed (β = .30, Z = 3.64), fluency (β = .31, Z = 2.68), and reasoning (β = .17, Z = 2.01). Together, these three specific factors accounted for 24% of variance. Omitting these specific effects resulted in a significant reduction of model fit ($\Delta\chi^2[5]$ = 19.9, p < .01) and greatly attenuated the total variance explained to R^2 = .04, indicating that the specific effects amount to 20% of explained variance.

2.1.6 Summary of Analyses of External Criterion Variables

Results from the analyses relating the NF model to external criterion variables have shown significant and meaningful relations of the general and specific factors to variables of socioeconomic background, personality constructs, and everyday activities. These analyses illustrate that moving away from aggregate variables on the criterion side, such as the SES sum index, the openness scale score, or the ExCo activity composite, opens the possibility to detect relations of specific cognitive factors independent of the general factor, which has its strongest relations to aggregated criterion variables.

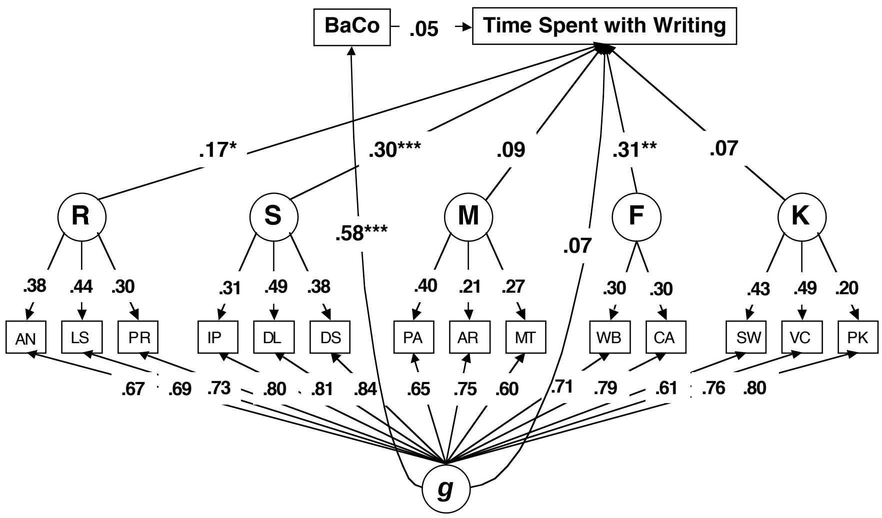

Figure 10. Prediction of the duration of writing activities in the *Yesterday Interview* with basic competency (BaCo) and the nested factor model. All factor loadings and regression weights are standardized estimates. g = General factor. R = Specific reasoning factor. S = Specific perceptual speed factor. M = Specific memory factor. F = Specific fluency factor. K = Specific knowledge factor. AN = Analogies; LS = Letter Series; PR = Practical Problems; SW = Spot-a-Word; VC = Vocabulary; PK = Practical Knowledge; IP = Identical Pictures; DL = Digit Letter; DS = Digit Symbol; WB = Word Beginnings; CA = Categories; PA = Paired Associates; AR = Activity Recall; MT = Memory for Text.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Using the NF model to represent the hypothesis of a common cause factor showed that all age-associated variance of the cognitive mechanics and a large portion of age-associated variance of the sensorimotor variables were contained in the common factor. However, the results also indicated the need to incorporate a negative age relation of the specific sensorimotor factor.

2.1.7 Control Analyses

All results in the preceding section were based on parameter estimates derived from the application of a theoretical model with latent unobservable variables to empirical data. The validity of these models and their parameter estimates depends on whether the assumptions of these models are met. Before interpreting the results, it is therefore necessary to examine the validity of these assumptions. This can be done either by testing the assumptions directly or by running control analyses—that are based on different assumptions—and comparing these results to the initial findings. The following control analyses will focus on the issues of missing data, distributional assumptions, and parameter interdependency and sensitivity.

2.1.7.1 *Potential Influence of Imputed Missing Values*

In the original raw data of the BASE cognitive battery at the first measurement occasion, 494 of the total 7224 data points (i.e., 516 subjects times 14 tests), or 6.8% were missing. These missing values were imputed with regression estimates obtained by predicting missing values with a regression of observed scores on other tests within the same ability constructs (e.g., predicting a missing value on *Letter Series* by observed values on *Analogies* and *Practical problems*, weighted by regression weights derived from the sample with complete observations; for a description of this procedure, see Lindenberger and Baltes, 1997). Analyses based on the imputed data are less likely to lead to biased results than listwise deletion and comparisons of analyses with and without imputed values obtained virtually identical results in the analyses of Lindenberger and Baltes (1997). However, the question of how much potential influence this procedure might have for analyses using the NF model is still an important issue to be investigated. Specific factors might have been artificially enhanced by the regression estimates for missing values, which could lead to increased correlations among the respective indicator variables. Therefore, two kinds of control analyses were conducted to address this issue.

First, the NF model with age correlations was re-estimated for the sub-sample that had complete observations for all 14 tests ($N = 416$). This resulted in a NF measurement model that was very much the same as in the analyses with the imputed missing values (see Appendix J). However, the age correlation of the specific memory now was also significant ($r = .20$, $Z = 2.14$). The age correlations of the other factors were also very similar to the ones reported before ($r = -.66$, $Z = 12.17$ for the general factor; $r = .39$, $Z = 5.35$ for the specific knowledge factor, and $r = .29$, $Z = 2.97$ for the specific fluency factor). The fit of the model was good with $\chi^2[72] = 147.11$, $p < .001$, RMSEA = .05, CFI = .98.

Secondly, a different procedure for estimating missing data was used: the EM algorithm (Dempster, Laird, & Rubin, 1977) as implemented in the SPSS 10.0 Missing Values module. This algorithm used all the available information on the 14 tests and the age variable as well as the interrelations among these variables to get best estimates of the missing values. With this data set, the age correlation of the specific memory factor was again significant ($r = .16$, $Z = 2.00$). The age correlations of the other factors were $r = -.67$ ($Z = 14.76$) for the general factor, $r = .36$ ($Z = 5.80$) for the specific knowledge factor, and $r = .27$ ($Z = 3.01$) for the specific fluency factor. The fit of the model was still acceptable, though slightly worse than with the regression imputation data ($\chi^2[72] = 190.95$, $p < .001$, RMSEA = .06, CFI = .98; see Appendix J, for the measurement model parameters). Taken together, the results of the control analyses with different ways of dealing with missing data did not indicate a potential influence of the regression imputation procedure on the interpretation of the general and specific factors in the analyses presented before.

2.1.7.2 Multivariate Normality Assumption of SEM Models

An important assumption of the maximum likelihood estimation techniques used in SEM is that the data are drawn from a multivariate normal distribution. One way to test this assumption is to use Mardia's estimate of multivariate kurtosis (Mardia, 1970), which is implemented in standard SEM programs. The normalized value of this index for the 14 cognitive tasks was 6.27. While this value could be considered relatively high, there are no established cut-off values and it is not clear how strongly violation of the normality assumption—as indexed by this estimate of kurtosis—does distort the results. Therefore, bootstrap control analyses were used to explore this issue further. Bootstrapping (Efron & Tibshirani, 1993; Nevitt & Hancock, 2001) is a method based on repeatedly re-running the analyses of interest on samples drawn from the original sample. It allows deriving standard errors for parameter estimates when assumptions about the population distribution are violated. In addition, the comparison of the bootstrapped parameter estimates to the ones estimated for the original sample provides an estimate of the bias that is produced by violation of the distributional assumptions. The only requirement to perform bootstrap analyses is that the sample is representative for the population, an assumption quite tenable for the first cross-sectional data set of BASE.

Bootstrap analyses were conducted with AMOS 4.0 (J. Arbuckle, 2000). For the NF model with age correlations, one thousand bootstrap samples were drawn. Appendix K provides detailed results for this model. The main finding was that factor loadings did not change considerably and all loadings were highly significant. The covariances of the nested factors with age were slightly stronger, in the positive as well as in the negative direction, while the corresponding standard errors were a little larger than in the original analyses. In general, the measurement model as well as the pattern of age relations and significant results was very similar to the analyses based on the normality assumption. Therefore, possible violation of the distributional properties in the current data set did not seem to influence results in any important way.

Because the activity variable *Writing* was highly skewed, the model with the nested factors predicting BaCo and *Writing* was also re-analyzed with a bootstrap procedure. As shown in Appendix L, again, the results were remarkably stable. The only slight difference was that the path from specific reasoning to *Writing* was only marginally significant ($p = .05$) based on the bootstrap standard errors.

2.1.7.3 Sensitivity Analysis of the Estimates of General and Specific Age-Associated Effects

To explore parameter interdependencies and the sensitivity of overall model fit to changes in the parameter estimates of general and specific age-associated effects (Li, Lewandowsky, & DeBrunner, 1996; see Appendix B for an introduction), a sensitivity analysis was conducted. Based on the NF model with age as covariate, a series of 31 models were estimated in which the value of the age correlation of the general factor was fixed to standardized values ranging from -.50 to -.80 in steps of .01. All other parameters of the model were freely estimated. The dependent variables of interest in this manipulation were the resulting estimates for the age correlations of the specific factors—indicating the degree of parameter interdependence—and the model fit as indexed by the RMSEA—indicating sensitivity of the model to biased values of the general factor correlation. Figure 11 shows that there was some degree of parameter interdependence among estimates of the age correlation of the general and specific factors. The higher the negative values for the general factor, the higher were the positive values for the specific factors. This inverse relation of the general and specific age correlations was reflected in the correlations of the corresponding parameter estimates, which ranged from -.41 to -.48 between the estimated age correlations of the specific and the general factors. Figure 11 also shows that the specific factor correlations even changed from positive to negative as the general factor correlation was fixed to smaller negative values.

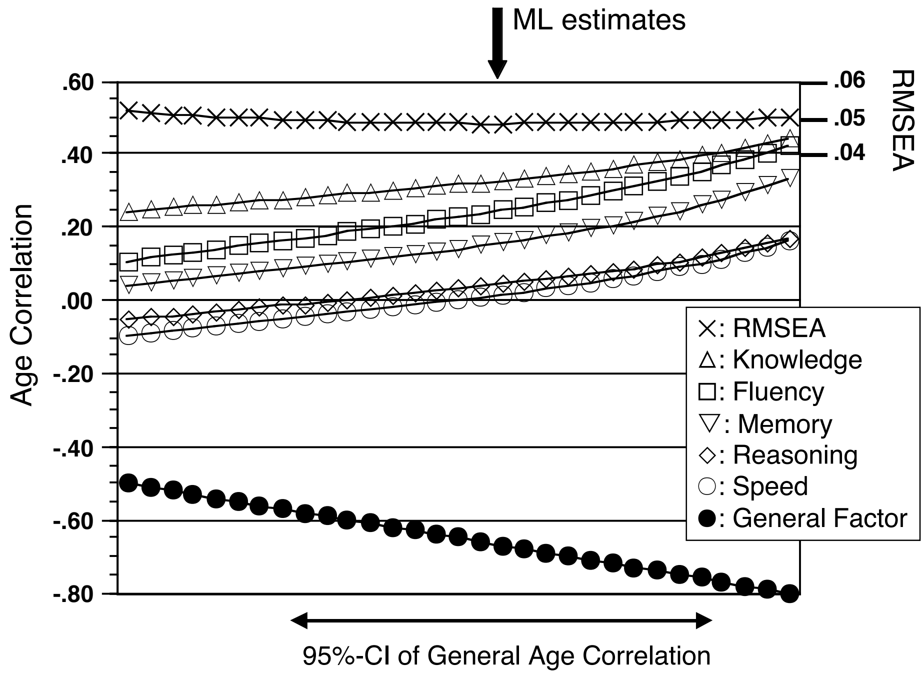

Figure 11. Sensitivity analysis of the age correlations in the nested factor model. The bold arrow denotes the point of the maximum likelihood estimates. Around this point, the value of the general factor age correlation was fixed to different values, and the estimates for the specific age correlations changed accordingly. The double-headed arrow denotes the range of the 95% confidence interval (CI) of the general factor age correlation estimate.

The most important observation of this sensitivity analysis, however, was that model fit was not very sensitive to changes in the estimates of the correlation between the general factor and age. If RMSEA values rounded to .05 are taken as indication of good model fit, this correlation could take on any value between -.50 and -.80. If RMSEA values smaller than .07 or .08 were interpreted as acceptable fit, the model could even be driven, e.g., to a combination of a general correlation of -.25, and negative age correlations of the specific factor ranging from .00 to -.38 (RMSEA = .066). These results show the flexibility of a model with simultaneous general and specific effects acting on the same variables in accounting for different assumptions about the general effect. What remained unchanged for all these possible combinations of general and specific effects, however, is that a general effect alone was not sufficient.

To examine this issue of parameter interdependencies and parameter sensitivity further, a Monte Carlo simulation study was conducted. This study had the aim of investigating which characteristics of the study design and the empirical composition of the indicator variables' variance have an influence on how accurately general and specific effects can be disentangled with the NF model.

2.5 Monte Carlo Simulation Study

To examine the hypothesis that the heterogeneity of the relative proportions of general and specific variance in the indicator variables has an influence on the parameter sensitivity and interdependence of the general and specific effect estimates, a Monte Carlo simulation study was conducted. This study was designed to address the following questions. First, how strongly does the heterogeneity of the variance composition influence the standard errors of the estimates of the general and specific effects? Second, will the true age-associated effects be recovered with the NF model if age correlations are specified for all specific factors in the model, but not all of them do have true relations to age? Third, how sensitive are common SEM fit indices to the omission of specific effects of varying size from the model and does the heterogeneity of the variance composition in the indicator variables have an influence on the sensitivity of these fit indices? Other variables, such as sample size or number of indicator variables per factor, were also manipulated to compare their relative impact on the dependent variables and to simulate a broad range of practically relevant conditions.

2.5.1 Study Design and Method

The study was set up to manipulate sample size, number of indicator variables, relative proportion of variance explained by the general factor, amount of heterogeneity of general and specific variance composition, and effect size of the specific effect as experimental factors. As shown in Figure 12, the basic model was a NF model with one general and three specific factors, which were each defined by three indicator variables. The age correlation of the general factor was set to -.71, which means that half of the general factor's variance was age-associated. The specific effect was simulated by an age correlation of the first specific factor, which was varied across different experimental conditions in steps of .10 from +.30 to -.50. This allowed the examination of how sensitive SEM fit indices are to the omission of specific effects.

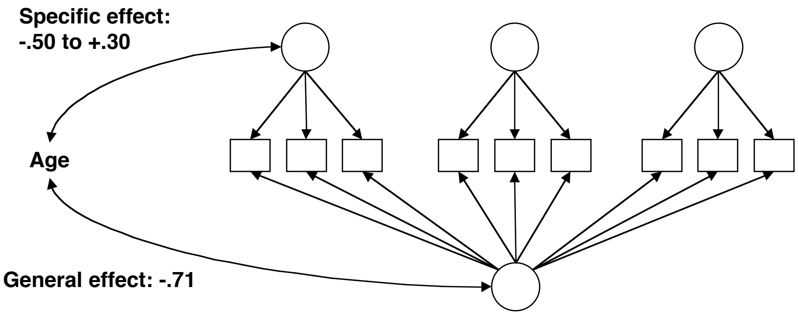

Figure 12. Basic design of the Monte Carlo simulation study.

Sample size was set to 125, 250, 500, and 1000 cases, respectively, to cover the whole range from the minimum sample size necessary for SEM analyses up to an ideal number not reached in the usual multivariate correlational studies on cognitive aging. The number of indicator variables per factor was set to either three or four. This was done because it has been shown that larger numbers of indicators are beneficial for the reliability of parameter estimates (Marsh, Hau, Balla, & Grayson, 1998). Therefore, the number of indicators also could be an important factor for the sensitivity and interdependence of the general and specific effect estimates.

The factor loadings were set up in a way that 80 percent of each variable's variance was reliable variance explained by the general and specific factors. An average of 40, 50, or 60 percent of the total variance was explained by the general factor to simulate different degrees of dedifferentiation of the factor structure, i.e., the general factor accounted for 50, 63.5, or 75 percent of the reliable variance. Around these average amounts of explained variance, disproportionality in the variance composition was introduced by varying the relative magnitudes of general and specific variance. In the low disproportionality (LD) condition, the amount of general variance for the first indicator of a factor was reduced by 2 percent, and increased by 2 percent for the specific factor. For the last indicator, a reversed manipulation of increasing general and decreasing specific variance was done. For the one or two indicators in between, variance proportions were chosen as linear interpolations between the two extremes (e.g., 48, 49.3, 50.7, and 52 percent for the variance explained by the general factor in four indicator variables, accounting for an average of 50 percent, and 32, 30.7, 29.3, and 28 percent for the variance explained by the specific factor, accounting for an average of 30 percent). In the medium and high disproportionality (MD and HD) conditions, the same manipulations were conducted, but with a maximum decrease and increase of 6 and 10 percent for the general and specific factor variances of the first indicator, respectively.

Population covariance matrices were calculated for each condition with a program written in SAS IML (SAS Institute, 1989b) and SAS Macro Language (SAS Institute, 1990). Based on this population covariance matrix, samples of varying size were repeatedly drawn from a multivariate normal distribution. Model parameters were estimated with SAS PROC CALIS (SAS Institute, 1989a). Model estimation was conducted twice for each sample. Once with the specific effect included, and once omitting the specific effect. This within-sample manipulation tested the model's sensitivity to misspecifications as well as the precision of parameter estimation when the correct model was used.

The experimental factors were completely crossed, resulting in a 4 (sample size) x 2 (number of indicators) x 3 (general factor variance) x 3 (disproportionality) x 9 (magnitude of specific effect) x 2 (specification of specific factor in model) design. For each of the 1296 cells, 50 replication samples were used, resulting in a total of 64800 single runs. This study design followed a recent recommendation by Skrondal (2000) to put more weight on a broad sampling of conditions than on replicating each condition many times. Appendix M provides the SAS program code that was used to simulate the data files and to estimate the models.

To analyze the effects of the experimental manipulations on dependent variables, the design was split into the 32400 runs with a correctly specified model, and the other 32400 runs where the specific effects were omitted. The dependent variables for the correctly specified models were the number of Heywood cases (negative error variance terms of the indicator variables), the number of improper solutions (age correlation estimates outside the range of -1.00 to +1.00), and the standard errors of the estimates for the general and specific age correlations. For the models with no specific effects specified, the dependent variable was the RMSEA fit index. This procedure tested whether the misspecification of models—i.e., specifying only a general effect—could be detected by using the conventional cut-off criteria for this fit index.

2.5.2 Results

2.5.2.1 *Heywood Cases and Improper Solutions*

First, it is of interest to understand how much disproportionality of general and specific variance across indicator variables influences the number of Heywood cases and improper solutions encountered. In practical applications such problems do complicate, and sometimes even prevent, the interpretation of results. As expected, the results provided in Appendix N show that sample size had a significant impact on the number of Heywood cases and improper solutions. Number of indicators only had a comparatively weak effect on the number of Heywood cases and no significant effect on number of improper solutions. Most interestingly, the disproportionality of general and specific variance in the indicator variables showed a strong negative correlation with the number of improper solutions and a weak positive correlation with the number of Heywood cases (see Table 5). The beneficial effect of reducing the number of improper solutions shows that the parameter space was much better behaved under conditions of more disproportionality. While the negative effect on the number of Heywood cases was difficult to explain, it did not pose a problem because the relative frequency of such cases was less than 1 percent even in the condition with high disproportionality.

Table 5

Influence of Disproportionality on the Number of Heywood Cases and Improper Solutions

	Disproportionality			χ^2 [*df*]
	Low	Medium	High	
Heywood cases	35	46	82	22.4 [2]
Improper solutions	2338	372	41	3674.8 [2]
Total Number of Runs	10800	10800	10800	

2.5.2.2 *Standard Errors of the Effect Estimates*

The next set of analyses targeted the effects of sample size, number of indicators, and disproportionality on the standard errors of the estimates of general and specific effects. For these analyses, the runs in which the specific effects were specified were selected, but excluding the runs in which Heywood cases and improper solutions were encountered. Appendix O provides ANOVA tables of these analyses. Because of the very large number of simulation runs, almost all of the possible main effects and interactions were significant, although most of them had only small effect sizes.

Therefore, only the most important findings are summarized here. The disproportionality factor had the strongest effect on the standard errors, followed by sample size, and a weak effect of number of indicators (the effects of the average variance explained by the general factor and the size of the specific effect are not considered here, because they were only included to represent a broad variety of empirical constellations). Figure 13 shows that the main effect of increasing the sample size from 125 to 1000 on the standard error of the specific effect was even smaller than the contrast of low and high disproportionality.

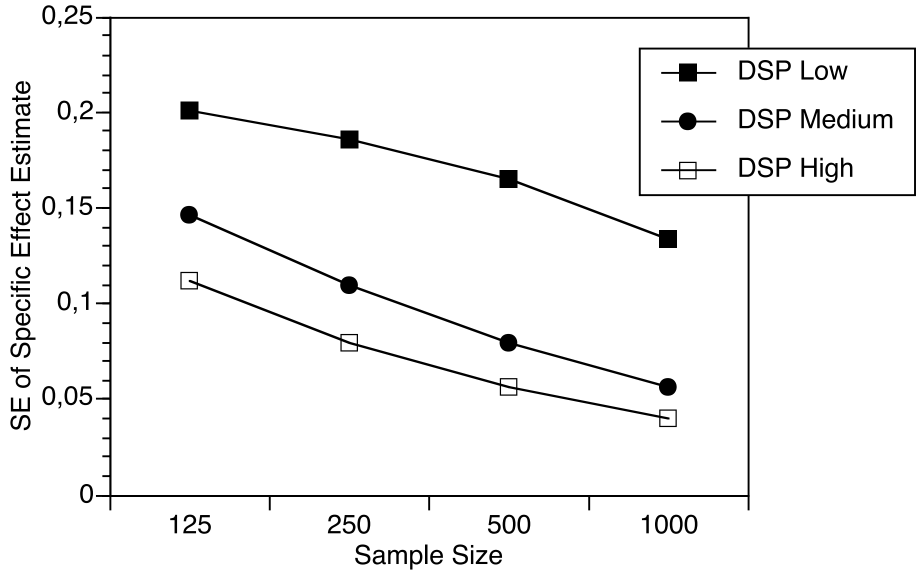

Figure 13. Standard error of the estimates of specific effects as a function of sample size and disproportionality of general and specific variance components in the indicator variables. Number of indicator variables per factor were three. DSP = Disproportionality.

2.5.2.3 *Sensitivity of Model Fit to Omission of Specific Effects*

The 32400 simulation runs in which the specific effect was not specified could be used to examine how sensitive were the measures of overall model fit to the omission of existing specific effects. Figure 14 shows the RMSEA index as a function of the size of the omitted specific effect, and the disproportionality of general and specific variance in the indicator variables (under the condition of three indicator variables per factor and a sample size of 500 to mimic the empirical situation in the BASE sample). One can see that, under the condition of low disproportionality, only omitting specific effects that are stronger than about -.40, which was larger than the empirical effects found in the BASE data, would indicate misfit of RMSEA > .05. Under the condition of high disproportionality, omitted specific effects stronger than

-.30 would lead to a RMSEA grater than .05. These results clearly showed that overall model fit criteria used with conventional cut-off values were not sensitive enough to detect small to medium sized specific effects. This underlines again the importance of using simultaneous estimation procedures that allow including all possible specific effects. Otherwise, an acceptable fit of a model that only specifies a general effect could lead to the conclusion that there is no need to further examine specific effects.

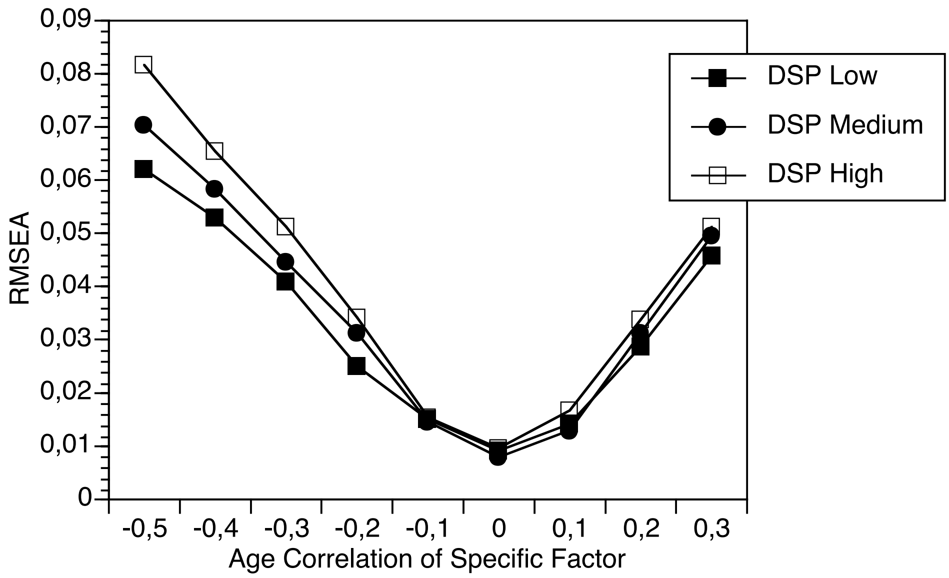

Figure 14. Model fit (RMSEA) as a function of the age correlation of the specific factor and the disproportionality of the model in conditions with a sample size of 500, three indicators per factor, and specific factor age correlations not included in the model. DSP = Disproportionality.

2.5.3 Summary of Monte Carlo Study Results

Results from the Monte Carlo simulations gave evidence that the NF model is capable of recovering the true general and specific effects—if a simultaneous estimation procedure is used. Parameters of specific effects that were not present did yield estimates within the range of expected random variability around zero. Results from the analyses of misspecified models with omitted specific effects showed that specifying only a general effect and taking overall model fit as a criterion can easily lead to missing small or even medium sized specific effects. Furthermore, it has been demonstrated that higher disproportionality of the composition of general and specific variance across the indicators of one specific factor leads to remarkably smaller standard errors of the estimates of the general and specific effects and, as a related phenomenon, to fewer improper solutions. Empirical disproportionality, therefore, is an important property of the data that the SEM models have to be able to capture. As discussed, this is achieved by the NF model and precluded by the constraints implied in the hierarchical model. Finally, the simulation results also indicated that it is highly desirable to have large sample sizes and several indicators per latent construct.