Mark L. Howe and Charles J. Brainerd
Editors

Cognitive Development in Adulthood

Progress in Cognitive Development Research

With 20 Illustrations

1988 Springer-Verlag

8. Internal Validity Threats in Studies of Adult Cognitive Development

K. Warner Schaie

Introduction

If development is viewed as a process that implicitly requires time to elapse to observe quantitative change or qualitative transformations, it then follows that age-comparative studies or other experimental manipulations can at best simulate development. It is obviously impossible to assign experimental subjects at random to different ages or different measurement occasions. The formal investigation of developmental processes, therefore, typically involves quasi-experiments in which carefully selected population samples are followed over time to observe whether or not hypothesized transformations can indeed be observed.

The concept of the quasi-experiment was popularized by the classical Campbell and Stanley (1967) research design monograph. These authors collectively denote as quasi-experiments "many natural social settings in which the research person can introduce something like experimental design . . . even though he lacks the full control over the scheduling of experimental stimuli . . . which make a true experiment possible" (p. 34). One example of a quasi-experiment for the study of cognitive aging, is the typical cross-sectional study in which aging effects are modeled by comparing subsamples that differ in age but which are assumed to be matched in all other aspects (Salthouse, 1982). A more complex example of a quasi-experiment in this substantive field would be a longitudinal study in which subjects are tested at specified ages, and re-examined at regular time intervals, where amount of practice is controlled for by adding matched controls groups at successive measurement points (Schaie, 1983a).

Because of the fact that quasi-experiments do not permit random assignment of study participants to different conditions with respect to which hypotheses are to be examined, the issue of the validity of such studies looms large. Campbell and Stanley (1967) distinguish between internal and external validity threats. The former refer to alternative interpretations which arise as a consequence of design flaws, the latter reflect the degree to which findings from internally valid studies can be generalized to other populations. Both of these classes of validity threats have substantial effects on the interpretability of studies of adult cognitive development (cf. Schaie, 1977). Specifically, these validity threats force us to examine the possibility that design flaws may be equally or more parsimonious in explaining results than the experimenter's stated hypothesis. Interestingly enough, at least some studies of adult cognitive aging have collected data that allow the investigator either to control for or to assess the validity threats inherent in their study designs. It is the unusual research report, however, that provides adequate documentation to show whether or not the reported findings might be more parsimoniously attributable to a detectable design flaw.

In this chapter I call attention to those threats to the internal validity of quasi-experiments that should be of concern to developmentalists at all stages of the life span, but that have been most directly addressed thus far in studies of adult cognitive development. All of the validity threats discussed here are directly amenable to experimental control or empirical assessment, given appropriate design strategies (cf. Schaie, 1973, 1977). Explicit designs derived from the longitudinal-sequential approach (Schaie, 1965) are examined to show how internal validity threats can be effectively handled in longitudinal studies. Finally, some empirical data are provided that show applications of these designs to the assessment of the significance of internal validity threats for data on adult cognitive behavior.

Threats to the Internal Validity of Developmental Studies

Eight different threats to the internal validity of quasi-experiments have been described by Campbell and Stanley (1967). One of these, maturation, represents no threat to the validity of developmental studies, if their intent is to test hypotheses about effects of aging. The remaining seven threats to be considered in order represent rival hypotheses to the effect of aging (i.e., maturational or age-specific changes). They involve the rival hypotheses of the effects of history, testing (reactivity), instrumentation, statistical regression, experimental mortality (attrition), selection, and certain interactions thereof. I briefly review the implications of these threats as they apply to the interpretability of findings in aging studies. A representative paradigm in such studies would be the natural experiment format of the typical pretest-treatment-posttest variety, in which the "treatment" is assumed to be the aging of the organism under study. In this design I compare the behavior of a sample of individuals at two points in time (T_1 and T_2) and infer that a change observed in the dependent variable has occurred as a function of the treatment; that is, the lapse of time during which the organism has matured. I now examine specific potential flaws to this inference.

History

With the exception of well-controlled animal studies, it is inevitable that events may occur in the environment between T_1 and T_2 that could account for the observed behavioral change, whether or not a maturationally determined change had occurred. We have no assurance that the unknown environmental event that resulted in behavioral change will recur with similar impact if we were to replicate our study over another time interval (Schaie, 1982). For example, age differences in life satisfaction for black and while elderly have been shown to exhibit different temporal patterns as a consequence of environmental changes that had more dramatic effects on the black than the white elderly (Schaie, Orshowsky, & Parham, 1982). We must conclude then that the single-cohort longitudinal study fails to control for history. This flaw might not be unduly bothersome in studies of physiological development during childhood, except during periods of intense political turbulence that create special hazards (e.g., malnutrition, exposure to excess radiation). In adults, however, changing conditions of health care and or stress may quite readily suppress or exacerbate maturational phenomena. In behavioral studies, there are so many possible sources of events at all age levels that can affect the dependent variable differentially across time that the single-cohert longitudinal study will rarely be useful to detect reliably generalizable behavioral changes (cf. Schaie, 1972; Nesselroade & Baltes, 1974).

Reactivity (Testing or Practice Effects)

Administering a test or survey instrument or introducing an observer into a behavioral situation may also result in effects that might erroneously be interpreted to be maturational. In the cognitive literature in particular, practice effects may be particularly serious over short periods of time but long-term reactivity effects may also prevail (cf. Schaie & Parham, 1974). For example, it is not clear whether differential aging patterns in fluid and crystallized intelligence (Horn, 1982) may not be attributable to the fact that greater practice opportunities are available for crystallized abilities. The reactivity threat to the validity of a longitudinal study can be controlled for by carrying a randomly assigned untested control group to be examined at T_2 only. However, this simple approach is also flawed because reactivity will be confounded with effects of experimental mortality that in this design cannot be assessed for the control group.

Instrumentation

Changes in measurement instruments, experimenters, or protocol between T_1 and T_2 are the most obvious sources of artifactual changes that could be caused by instrumentation effects. Other more subtle sources reflect changes in the projection of observed (measured) variables on the latent constructs of interest (unmeasured variables), changes that may well be substantial if comparisons are made over different developmental stages or extensive time periods (cf. Schaie & Hertzog, 1982, 1985). The latter effects can, of course, be examined only if a study contains multiple markers of the constructs at all measurement points (an empirical example testing the construct equivalence for a cognitive data set in adults is provided in Schaie, Willis, Hertzog, & Schulenberg, 1987). Instrumentation effects for single variables can be untangled from practice effect, but will remain confounded with the effect of history.

Statistical Regression

The concept of statistical regression implies that when the position of an individual at T_2 is predicted from that person's score at T_1, the best prediction would be that the score at T_2 will be closer to the mean of the group. The degree of such movement will be directly influenced by the reliability of the measurement operations; the lower the reliability, the greater the regression to the mean. It should be noted, however, that the concept of statistical regression is a function of the system of linear predictions (cf. Hays, 1963, pp. 500–501). In the typical research situation, regression to the mean may occur for empirical as well as statistical reasons. For example, it has been argued that age may be more kind to the more able, but empirical studies do not necessarily support this argument (Baltes et al., 1972).

If the regression effect results from unreliability of the measures, then one would expect the variance at T_2 to be lower than that at T_1, but such reduction in variance is not guaranteed when regression to the mean occurs as a consequence of interventions that have differential effects at different portions of the range of talent. In any event, it is likely that mean scores above the population average at T_1 will be lower at T_2 while low means will increase. Such regression effects would serve to enhance or obscure true maturational effects (Furby, 1973). There have been some arguments to support the contention that regression effects should not be cumulative across successive occasions (Nesselroade, Stigler, & Baltes, 1980). Regression effects must, nevertheless, be taken seriously in any longitudinal two-point comparisons. Examples of intuitive methods using a time-reversal paradigm that will differentiate maturational and regression effects have been described elsewhere (Baltes et al., 1972; Campbell & Stanley, 1967; Schaie and Willis, 1986b), but will not be considered further in this chapter.

Experimental Mortality (Attrition)

A major threat to the internal validity of a pretest-treatment-posttest design occurs when all subjects tested at T_1 are not available for retest at T_2. In studies of cognitive aging, experimental mortality would include death, disability, disappearance, and simple failure to cooperate for the second test. There is a substantial literature that reports differences in base performance between those who appear and those who fail to appear for the second or subsequent tests. Typically the dropouts at base score lower on ability variables and describe themselves as possessing less socially desirable traits than to the retest survivors (e.g., Baltes, Schaie, & Nardi, 1971; Cooney, Schaie, & Willis, 1988; Riegel, Riegel, & Meyer, 1967; Schaie, Labouvie, & Barrett, 1973). As a consequence it has been argued that longitudinal studies may represent successively more elite groups, and may no longer be sufficiently generalizable. This proposition can be tested, however, and suitable adjustments for attrition are readily available.

Selection

Differential selection per se is not a problem in single-cohort longitudinal studies. However, it becomes the major internal validity threat in cross-sectional and other nonequivalent control group designs. It has long been known that it is virtually impossible to rule out the likelihood that differences observed across groups selected at different ages may well be a function of differential recruitment, particularly when snowball or other voluntary panels of subjects are utilized (Schaie, 1959). Selections effects in developmentally oriented studies have typically been considered under the rubric of cohort effects (cf. Baltes, Cornelius, & Nesselroade, 1979; Schaie, 1965, 1977, 1986). More complex designs are required to control for or assess these effects (see following).

Interactions

A number of interactions of the threats discussed may also be problematic for the design of studies of adult cognitive development.

Maturation by History

This interaction is of particular concern in cross-sectional studies in which one must consider the possibility that effects of environmental impact could have differential import depending on the life stage of the individual when the impact occurred. For example, the effects of the Vietnam era made a much greater impact on the cohort then liable to be drafted than on younger or older persons. In longitudinal studies, the possibility of differential maturation by history effects must also be considered when individuals are followed over several measurement points.

Maturation by Experimental Mortality

Drop-out effects do not necessarily occur in a symmetrical fashion. In particular, drop out due to death or disability may be systematically related to the age of the sample studied. For example, among the young-old, drop-out effects are greatest for individuals lost by illness, while for the old-old the effects are greatest for those lost due to death (Cooney et al., 1988). Experimental mortality by age effects can be estimated only when more than one cohort is followed longitudinally.

Maturation by Reactivity

Again, practice or other reactivity effects may well differ by age. This interaction may reflect prior differential exposure to the measurement operations at different ages. For example, college students are likely to have attained an asymptotic level with respect to many ability measures because of similar recent experiences during the educational process. Older individuals, by contrast, may not have taken educational tests for several decades. Assessment of the significance of this interaction requires the assessment of practice effects at two or more age levels.

Maturation by Selection

Maturational effects can have differential magnitudes depending on the selection conditions prevailing when sampling occurs. For example, changes in rate of cognitive decline will lead to differential representation of well-functioning older persons for successive generations. The disaggregation of aging and selection (cohort) effects requires the study of two or more cohorts at two or more ages.

Controlling for any of these complex validity threats or disaggregating the effects attributable to maturation from the various confounded sources of variance typically requires multiple test occasions and replications of cross-sectional or longitudinal studies or their combination in a single design. We next examine appropriate design complications that have been proposed for use in the study of adult cognitive development. Later, empirical examples of an application of each of these designs are presented.

The Longitudinal-Sequential Approach As a Method for the Control or Assessment of Internal Validity Threats

In this section, I discuss a number of designs that can be utilized to control or assess most of the internal validity threats discussed above. To begin, a number of rather simple design variations that control for a single rival hypothesis are described. Designs that control for multiple threats are then presented. The reader should note that the most carefully controlled or complex design is not necessarily the best. In fact, the first step in design selection must be to reason out whether or not a given threat has a high probability of occurrence for the dependent variable of interest. All controls are expensive in experimenter effort and resources, and controls should therefore be imposed only when absolutey required to protect the integrity of a given study. Good research paradigms are therefore always driven by meta-theoretical context (cf. Costa & McCrae, 1982; Hultsch & Hickey, 1978; Labouvie, 1982; Nesselroade & Labouvie, 1985; Schaie & Hertzog, 1982). The following designs have been developed in a dialectic process that involved conducting studies that were discovered to have design flaws, designing paradigms that would remedy these flaws, and then discovering that the new designs required acquisition of further data sets (cf. also Schaie, 1959; 1973; 1977; 1986; Schaie & Willis, 1986, pp. 23–34).

Direct Assessment of Attrition Effects

Attrition effects can be assessed directly in a pretest-treatment-posttest design by testing the null hypothesis for the dependent variable at T_1, comparing participants who have returned for the T_2 measurement with those who have not (cf. Baltes, Schaie, & Nardi, 1971; Schaie et al., 1973). Because of the cross-sectional nature of the comparison, however, this most direct analysis will only provide information on the changing nature of our longitudinal sample; it will not provide an estimate of the effects of attrition on the measurement of age changes. Nor can this level of attrition analysis account for the possible confounding of attrition with practice, or the interaction with maturation, and selection. More complex analyses, as are described later in this chapter are required to resolve these confounds.

Direct Assessment of Practice Effects

The most direct assessment of practice (reactivity) effects is accomplished either by carrying a control group that has not been examined at T_1, or by drawing a new random sample from the cohort tested at T_1 for assessment at T_2. Since experimental and control groups differ at T_2 only in that the control has not received previous practice, differences in performance at T_2 can then be treated as a reasonable estimate of the magnitude of practice effects. An important caveat for this design, however, must be that attrition effects are equivalent for the experimental and control groups. But attrition effect estimates require that T_1 scores be available for both experimental and control groups. We obviously do not have this information for the control group, and must conclude that the most direct estimate will necessarily be confounded with attrition effects. Nevertheless, if attrition is relatively limited, the direct assessment of practice as suggested here may be the most viable approach. Moreover, if we are willing to assume that attrition effects in the experimental and control group are comparable, it is then possible to adjust estimates of practice effects by the differentials between dropouts and controls at the T_1 assessment (see examples described later in this chapter).

It has been shown that nonrandom effects of experimental mortality are most pronounced from T_1 to T_2; thereafter, attrition continues but tends to become somewhat more random with respect to the dependent variable until advanced ages are reached (cf. Gribbin & Schaie, 1979). An alternative, somewhat more complex, design for estimating practice effects (that would also control for nonpractice-related reactivity to the first measurement) would require carrying both experimental and control groups for one more measurement occasion. The estimate of practice would then be the difference between groups of equivalent maturational level at T_3, at which time we would be testing for the effect of one additional unit of practice. In this case it would, incidentially, be possible to assess the equivalence of attrition effects from T_2 to T_3 across experimental and control groups (see above).

Designs That Combine Controls for the Effects of History or Selection With the Assessment of Attrition Effects and/or Practice

The more complex designs required to handle more than one internal validity threat or those involving interactions are derived from what has been called "a general model for the study of developmental problems" (Schaie, 1965, 1977, 1986) and typically involve a further complication of one of the longitudinal-sequential strategies. These strategies allow three basic options: The first approach crosses maturational and selection effects (cohort-sequential design) by sampling two or more cohorts over the same age range, thus conducting a multicohort longitudinal study. The second option crosses maturational and history effects (time-sequential design) by sampling two or more age levels at two or more measurement points; this is a replicated cross-sectional study. The third approach crosses maturational and selection effects (cross-sequential design) by sampling two cohorts across the same time interval—a mixed cross-sectional and longitudinal approach.

History by Attrition Designs

We can assess the effects of history and attrition by considering two longitudinal samples carried over the same age range: that is, sample 1 is followed from T_1 to T_2 and sample 2 from T_2 to T_3. Because only the first measurement points are considered for either sample, this design is cross-sectional and thus controls for testing and reactivity. This design can be analyzed by a two-way ANOVA that crosses attrition and time of measurement. Note, however, that this approach does not control for selection effects.

The time-sequential strategy allows a more general (and interesting) format for this design in that it allows estimating the components of variance associated with maturation, history, and attrition. The sampling plan and ANOVA model for this general approach have previously been described (Schaie, 1977, p. 53, Table 10; also see following example). This design requires a minimum of two age levels from two samples initially assessed at T_1 or T_2. The attrition information is obtained by assessing sample 1 also at T_2 and sample 2 at T_3. The last assessment point does not enter the analysis.

Selection by Attrition Designs

A more complex approach is required to cross maturational, selection and attrition effects. This can be accomplished by extending the standard cohort-sequential design (Schaie, 1977, p. 44, Table 2). Two cohorts are followed over a minimum of two ages. The first cohort is assessed at T_1, T_2, and T_3, while the second cohort is assessed at T_2, T_3, and T_4. In each case the final assessment is for the purpose of identifying dropouts and does not enter the analysis. Note that this design does not control for testing and reactivity. That is, attrition and practice effects must necessarily be confounded. However, it is possible to include the necessary controls by utilizing the cohort-sequential design with independent measurements (Schaie, 1977, p. 45, Table 3). In this instance, we would need to carry a minimum of two samples from each of two cohorts at two age levels. Again four assessment points are required: $C_1A_1S_1$ is tested at T_1 and T_2; $C_1A_2S_2$ and $C_2A_1S_1$ are tested at T_2 and T_3; and $C_2A_2S_2$ is tested at T_3 and T_4. Only data at first time of test are included in this analysis to control for testing and reactivity.

History by Selection by Attrition Designs

During those periods of adulthood when maturational change is minimal, it is also possible to specify designs that allow crossing attrition effects, with both history and selection. These designs involve extensions of the cross-sequential approach (Schaie, 1977, p. 49, Tables 7 and 8). These designs follow a minimum of two cohorts over at least two times of measurement, with an additional measurement point at T_3 to determine dropouts following the T_2 assessment. The repeated measures design requires classification into dropouts and survivors after the first and second assessment. Experimental mortality at T_2 consequently is confounded with practice effects. In the independent measurement variant, practice is controlled for, but separate samples must be assessed twice for each level of the design.

History by Practice Designs

Crossing history and practice for a particular age level requires a minimum of four samples and three measurement occasions. Data are compared at T_2 and T_3 for samples of equivalent age that were either previously assessed at T_1 and T_2, or that are assessed at T_2 and T_3 for the first time. A more general format of this design, which balances ages at pretest, is an extension of the time-sequential design that crosses maturation, history and practice (see Schaie, 1977, p. 47, Table 6). This design requires four samples at each of two age levels, half of which are pretested at an earlier occasion before generating the data that enter this analysis. Note that this design confounds practice with attrition.

Selection by Practice Designs

The cohort-sequential design with independent measurements can be expanded to permit crossing of maturation, selection, and practice. A minimum of two samples are required at each age level for each cohort, half of whom have received practice at a previous data point that does not enter analysis. As a consequence, a minimum of four assessment points will be required for this design, which also confounds practice with attrition (see Schaie, 1977, p. 46, Table 4).

History by Selection by Practice Designs

Once again, for those adult age levels when maturation can be assumed to be trivial, it would be possible to cross history, selection, and practice by an extended version of the cross-sequential design with independent measurements. The minimum design here requires three assessment points. That is, each of the four possible cohort/time of measurement combinations (at T_2 and T_3) would require two samples, one of which received practice at T_1 or T_2 respectively (see Schaie, 1977, p. 50, Table 9).

Designs Crossing Practice and Attrition

Using the independent measurement designs allows assessing the effects of experimental mortality while controlling for practice. The converse approach (assessing effects of practice while controlling for experimental mortality) is not feasible because study participants returning for a second

assessment represent, by definition, only the group of retest survivors. It is possible, however, to cross attrition and practice, if rather than comparing only survivors and dropouts after T_1, we also consider survivors and dropouts after T_2. In this case a prior occasion not entering analysis is required for half of our groups, and all individuals must be followed to the occasion beyond the last analysis point, to determine dropouts and survivors for each subset. All the designs described can be treated in this manner, but an additional assessment occasion is required. Thus designs crossing practice, attrition, maturation, and history, or those crossing practice, attrition, selection, and history will require four assessment occasions. The design that crosses practice, attrition, selection, and history requires a minimum of five occasions (see Schaie, 1977, pp. 54–55, Table 11).

Empirical Data on the Significance of Internal Validity Threats for Data on Adult Cognitive Development

Description of the Data Base

The data to be discussed come from the Seattle Longitudinal Study (SLS), a multiwave panel study that uses as its population frame the membership of a metropolitan health maintenance organization (cf. Schaie, 1983b). All 3442 participants were, at first test, community-dwelling adults who were randomly selected from each 7-year age stratum included in each panel. These data were collected in 1956 (N = 500; ages 22–70 years), 1963 (N = 997; ages 22–77), 1970 (N = 705; ages 22–84); 1977 (N = 612; ages 22–84), and 1984 (N = 628; ages 22–84 years). At each successive data point as many survivors of the previous wave as possible were reexamined. Thus we have 1357 participants for whom 7-year longitudinal data at T_2 are available for four data sets: 1963 (N = 303; ages 29–77), 1970 (N = 420; ages 29–84), 1977 (N = 340; ages 29–91), and 1984 (N = 294; ages 29–91). Fourteen-year longitudinal data at T_3 are available for 723 participants in three data sets: 1970 (N = 162; ages 36–84), 1977 (N = 337; ages 36–91), and 1984 (N = 224; ages 36–94). Twenty-one-year longitudinal data at T_4 exist for 355 participants in two data sets: 1977 (N = 130; ages 43–91), and 1984 (N = 225; ages 43–94). Finally, there is one 28-year longitudinal data set in 1984 at T_5 (N = 97; ages 50–94 years).

All participants were in good health when tested, and were representative of the upper 75% of the socioeconomic stratum. For the total data base educational levels averaged 13.27 years (range, 4–20 years), and occupational status averaged 6.25 on a 10-point scale using census classifications ranging from unskilled labor to professional.

Throughout the study, subjects were assessed with the first five primary mental abilities (Schaie, 1985; Thurstone & Thurstone, 1941), the Test of Behavioral Rigidity (Schaie & Parham, 1975), and a demographic information form. All subjects were tested in small groups in sessions, which for the first three waves lasted about 2 hours, for the fourth wave about 3 hours, and for the fifth wave in two sessions of 2.5 hours each (necessary because multiple markers of the abilities and other additional measures were added). For the examples in this chapter we will limit our discussion to the five primary abilities: Verbal Meaning, the ability to comprehend words, a measure of recognition vocabulary; Spatial Orientation, the ability to mentally rotate objects in two-dimensional space; Inductive Reasoning, the ability to infer rules from examples that contain regular progressions of information; Number, the ability to manipulate number concepts, as measured by checking simple addition problems; and Word Fluency, a measure of recall vocabulary.

Some Basic Findings

The foregoing data base by now contains sufficient information to allow us to conduct virtually all of the analyses that I have described, allowing either controls for or assessment of the magnitude of effects of the internal validity threats described previously. I begin my discussion by presenting some basic data and then proceeding through some further analyses investigating the effects of the internal validity threats. For ease of comparison across the different measurement variables, all raw scores have been scaled to T-score form ($M = 50, SD = 10$) on the basis of the total sample of 3442 individuals at first test across all five test occasions. All data have been organized in subsets with equal interval boundaries; that is, all age and cohort groupings are expressed in 7-year intervals to conform to the 7-year intervals between assessment occasions.

The Cross-Sectional Sequence

The initial test data can be conceptualized as a sequence of five cross-sectional studies. Table 8.1 provides the scaled means for each separate sample for all available age groups. Weighted means across the five data sets are also provided, and the significance level (from one-way ANOVAs) is indicated for those age levels at which the means for the separate samples differ in a nonchance manner. Significant findings here provide evidence for the presence of selection by history effects. Significant variation across time (time lags) are characteristic for most age levels of all the Primary Mental Ability (PMA) variables, with the exception of Spatial Orientation.

These data of course do not allow direct inferences on age change in cognitive behavior within individuals. They are, however, relevant for a description of age differences among groups of individuals, as well as the identification of peak performance ages, at particular points in time. Over the age range from 25 to 67 years (data available for all five test occasions), for example, peak performance ages for Verbal Meaning have shifted upwards from age 32 in 1956 to age 46 in 1984, and the range of mean age differences has

Table 8.1. Scaled Means for the Primary Mental Abilities for Five Cross-Sectional Samples Assessed at 7-Year Intervals

Mean Age	1956	1963	1970	1977	1984	Average
Verbal Meaning						
25	52.64	53.30	53.84	54.68	55.46	53.94
32	54.87	54.05	53.80	56.22	54.83	54.64
39	51.90	54.20	53.95	54.96	56.86	54.83**
46	53.36	51.73	54.86	52.49	57.36	53.54***
53	49.10	48.35	54.45	52.87	54.79	51.43***
60	44.45	46.84	52.30	50.64	52.46	49.16***
67	42.56	42.57	45.26	46.44	48.68	44.86***
74	—	39.66	39.85	40.88	44.32	41.11**
81	—	—	37.92	35.72	40.60	37.92**
Spatial Orientation						
25	54.00	53.30	53.84	54.68	55.46	53.94*
32	54.95	54.16	57.28	55.98	54.02	55.14
39	51.96	53.16	53.84	54.78	53.10	53.38
46	51.12	51.76	54.73	52.72	53.82	52.71
53	47.28	48.99	50.82	51.00	49.51	49.52
60	46.16	48.14	48.85	47.65	48.08	47.87
67	44.10	44.22	43.77	46.98	43.97	44.55
74	—	41.97	42.16	41.68	41.72	41.88
81	—	—	40.70	39.44	39.82	39.92
Inductive Reasoning						
25	55.19	58.60	59.84	59.02	60.01	58.23***
32	55.67	56.02	58.14	57.72	58.86	56.91*
39	51.07	53.84	54.13	56.60	57.83	54.53***
46	51.68	50.07	53.56	52.61	56.18	52.29***
53	48.41	46.45	51.09	51.55	53.72	49.56***
60	42.63	44.83	49.52	48.31	50.97	47.10***
67	42.04	40.53	42.80	45.02	46.83	43.13***
74	—	39.86	39.51	40.82	44.16	41.03***
81	—	—	38.81	38.91	38.86	38.86
Number						
25	48.79	50.73	51.29	49.25	48.22	49.69
32	51.67	53.78	53.49	50.72	50.55	52.35
39	51.25	54.31	52.37	51.17	48.83	52.10**
46	53.58	53.01	55.46	48.07	51.52	52.57***
53	52.80	50.66	54.74	51.80	48.99	51.77**
60	47.58	49.64	55.20	50.42	49.50	50.45***
67	47.67	46.55	48.24	49.94	48.79	48.06*
74	—	44.33	44.90	44.64	46.91	45.19
81	—	—	41.35	42.07	41.73	41.74

TABLE 8.1 *Continued*

Mean Age	1956	1963	1970	1977	1984	Average
Word Fluency						
25	53.96	52.34	53.36	53.50	55.19	53.63
32	56.65	52.54	50.78	54.71	54.90	53.71**
39	54.30	51.91	50.36	52.00	54.71	52.45*
46	56.49	50.80	52.66	49.55	53.47	52.20***
53	55.63	47.68	52.82	50.91	51.98	51.15***
60	50.00	49.15	50.42	50.09	47.70	49.43
67	47.95	44.64	44.12	47.68	46.95	46.01*
74	—	44.66	41.54	43.07	44.78	43.45*
81	—	—	42.46	41.82	41.24	41.84

*$p < .05$; **$p < .01$; ***$p < .001$.

been reduced from approximately 1 *SD* to 0.6 *SD*. For Inductive Reasoning, the range of age differences has remained the same, but the peak performance age has shifted upwards from age 32 to age 46. For Word Fluency the range of age difference has increased slightly, and the peak performance age has shifted downward from 32 to 25. Age difference patterns have remained fairly stable for Spatial Orientation, while substantial nonsystematic fluctuation occurred for Number skill.

It should be evident that any particular set of age differences will be confounded with cohort (selection) differences, and that an assessment of reasonably stable age differences will require aggregation of data for several cross-sectional studies covering the same age ranges. Such average data are given in the last column of Table 8.1.

THE LONGITUDINAL SEQUENCES

The estimates of mean level obtained at various ages from the longitudinal studies are not directly interpretable since they reflect data from attrited samples. Direct estimates of within-individual change over time are consequently more informative. Table 8.2 provides data on longitudinal change for the PMA variables over a 7-year period aggregated across the four longitudinal sequences described. The last column in this table aggregates 7-year changes regardless of level of prior practice to obtain the largest possible sample sizes. It will be noticed that on average small gains are experienced into the thirties, and that decremental age changes are quite small and only reach Cohen's (1977) criterion of a minimally interesting effect size (0.2 *SD*) for some variables by age 67. Such changes, increasing in magnitude, are then seen for successive 7-year intervals.

It is of interest further to note the cumulative effect of longitudinal age changes as derived from the overall estimates in the last column of Table 8.2.

TABLE 8.2. Magnitude of Longitudinal Age Changes for the Primary Mental Abilities Across a 7-Year Interval[a]

	T_1 to T_2		T_2 to T_3		T_3 to T_4		T_4 to T_5		Average	
Mean Ages	*N*	*d*	*N*	*d*	*N*	*d*	*N*	*d*	*N*	*d*
Verbal Meaning										
25–32	135	+1.42		—		—		—	135	+1.42
32–39	169	+1.70	69	+0.08		—		—	238	+1.23
39–46	213	+1.06	96	+0.02	34	−2.26		—	343	+0.45
46–53	204	+0.54	108	−0.56	52	−0.95	13	−0.73	377	−0.03
53–60	223	−0.28	126	−0.55	56	−1.60	17	−1.67	422	−0.68
60–67	170	−1.59	117	−2.24	70	−2.82	17	−0.65	374	−2.03
67–74	163	−2.12	66	−3.17	55	−3.00	22	−3.28	306	−2.59
74–81	60	−1.79	49	−4.05	21	−2.61	14	−5.93	144	−3.43
81–88	18	−4.68	8	−9.70	10	−9.25	3	−8.67	39	−7.36
Spatial Orientation										
25–32	135	+0.99		—		—		—	135	+0.99
32–39	169	+0.27	69	+0.24		—		—	238	+0.26
39–46	212	+0.03	96	−0.45	34	+0.96		—	342	−0.02
46–53	204	+0.13	108	−0.73	52	−1.57	13	−2.89	377	−0.45
53–60	222	−0.26	126	−1.14	56	−0.40	17	−3.03	421	−0.66
60–67	168	−2.05	117	−1.41	69	−3.16	17	+0.41	371	−1.94
67–74	162	−2.81	64	−1.49	54	−2.91	21	+0.04	301	−2.35
74–81	59	−4.10	49	−2.15	21	−1.53	14	−5.47	143	−3.19
81–88	18	−4.68	7	−4.06	8	−4.96	3	−6.97	36	−4.81
Inductive Reasoning										
25–32	135	+0.48		—		—		—	135	+0.48
32–39	169	+0.68	69	−0.24		—		—	238	+0.42
39–46	213	+0.08	96	−0.34	34	−0.48		—	343	−0.09
46–53	204	−0.31	108	+0.30	52	−0.29	13	+0.72	377	−0.10
53–60	223	+0.19	126	−0.20	56	−1.75	17	−1.77	422	−0.26
60–67	170	−2.28	117	−1.49	70	−2.51	17	−2.28	374	−2.07
67–74	163	−2.07	66	−2.88	55	−3.60	22	−3.53	306	−2.63
74–81	60	−2.28	48	−2.59	21	−0.92	14	−2.68	143	−2.22
81–88	17	−2.54	8	−8.34	10	−3.44	3	−2.24	38	−3.97
Number										
25–32	135	+1.32		—		—		—	135	+1.32
32–39	169	+ .14	69	−1.36		—		—	238	−0.29
39–46	213	−0.10	96	−0.85	34	−1.14		—	343	−0.41
46–53	204	−0.48	108	−2.50	52	−0.95	13	−0.86	377	−1.14
53–60	222	−0.22	126	−1.40	56	−1.43	17	−1.43	421	−0.78
60–67	171	−1.56	117	−2.47	70	−3.09	17	−0.76	375	−2.13
67–74	163	−2.24	66	−2.34	55	−3.00	22	+1.03	306	−2.17
74–81	60	−4.41	49	−5.01	21	+0.91	14	−3.47	144	−3.75

TABLE 8.2 *Continued*

Mean Ages	T_1 to T_2 N	T_1 to T_2 d	T_2 to T_3 N	T_2 to T_3 d	T_3 to T_4 N	T_3 to T_4 d	T_4 to T_5 N	T_4 to T_5 d	Average N	Average d
Word Fluency										
25–32	135	+0.89		—		—		—	135	+0.89
32–39	169	−0.55	69	+0.75		—		—	238	−0.17
39–46	213	−0.12	96	+0.69	34	−1.24		—	343	+0.01
46–53	204	−0.79	108	−0.90	52	−0.79	13	−0.77	377	−0.82
53–60	223	−1.25	126	−1.31	56	−0.74	17	−0.40	422	−1.17
60–67	171	−2.48	117	−1.07	70	−1.17	17	−2.73	375	−1.81
67–74	163	−2.49	66	−2.08	55	−2.84	22	−2.88	306	−2.56
74–81	60	−4.07	49	−2.70	21	−2.88	14	−4.25	144	−3.44
81–88	18	−4.19	8	−1.73	10	−2.96	3	+2.29	39	−2.57

[a]Data in T-score points.

The magnitudes of these cumulative changes are given in Table 8.3. For Verbal Meaning, the most crystallized of our measures, cumulative gain remains significantly above the base age until age 60. Early cumulative gains do not reach levels of statistical significance.

In terms of Cohen's criteria, cumulative decrement reaches a small but significant level for Number and Word Fluency by age 67 and for all other variables by age 74. The effects of cumulative decrement become moderately severe (0.5 *SD*) by age 74 for Number and Word Fluency, and by age 81 for

TABLE 8.3. Cumulative Magnitude of Longitudinal Age Changes for the Primary Mental Abilities From Base Age 25 to Ages 32–88[a]

Mean Age	Verbal Meaning	Spatial Orientation	Inductive Reasoning	Number	Word Fluency
32	+1.42	+0.99	+0.48	+1.32	+0.89
39	+2.65*	+1.25	+0.90	+1.03	+0.72
46	+3.10**	+1.23	+0.81	+0.62	+0.73
53	+3.03**	+0.78	+0.71	−0.52	−0.09
60	+2.35*	+0.12	+0.45	−1.30	−1.26
67	+0.32	−1.82	−1.62	−3.43**	−3.07**
74	−2.27*	−4.17**	−4.25**	−5.60**	−5.63**
81	−6.00**	−7.36**	−6.47**	−9.35**	−9.07**
88	−13.36**	−12.17**	−10.44**	−14.96**	−11.64**

[a]Data in T-score points.
*Significant at or beyond the 5% level of confidence.
**Significant at ar beyond the 1% level of confidence.

the other variables. However, substantial effect sizes (>1 *SD*) are reached only by age 88 for all variables.

The first four columns of Table 8.2 disaggregate the longitudinal data into separate estimates for those subjects who had no prior practice before the initial point of the 7-year interval, and for those who had either one, two, or three earlier experiences with the test material. With few exceptions, the groups with no prior experience show either greater gain or less decline than those with greater prior experience. One might infer from these data that the estimates for the first longitudinal interval could be inflated by favorable practice effects. It should be noted, however, that this conclusion can be preliminary only, since different levels of practice are confounded with experimental mortality (drop-out) effects. Moreover, the disaggregated data in Table 8.2 may also be disparately affected by history; that is, given the cumulative nature of the data, T_1/T_2 data are cumulated over four 7-year periods, T_2/T_3 data over three periods, and T_3/T_4 data over two periods; T_4/T_5 data are available only for a single period. These differences are of course also reflected in disparate sample sizes as shown for each estimate in Table 8.2. Some of these problems can be directly or indirectly addressed utilizing the analytic strategies mentioned for which examples utilizing our data base are now provided.

The Estimation of Attrition and Practice Effects in Longitudinal Data

I first report complete data on well-replicated direct estimates of attrition, practice, and practice adjusted for attrition effects based on large samples covering the entire age ranges represented in the Seattle Longitudinal Study. I examine how these effects have differential impact on estimates of longitudinal change at different age levels. Finally, more limited examples controlling for the interaction of attrition and practice with other validity threats are examined.

DIRECT ESTIMATES OF ATTRITION

I begin by investigating the magnitude of attrition effects, using the direct estimates described on page 247. These effects can be examined for several longitudinal sequences to contrast the base performance of individuals for whom longitudinal data are available and those who dropped out after the initial assessment, and it is also possible to consider shifts in direction and magnitude of attrition effects that may occur after multiple assessment occasions. Because the overall means are affected by slightly different age compositions, I report attrition data as the difference in average performance between dropouts and returnees (Table 8.4). Attrition effects vary across samples originally assessed at different points in time, but after the first test generally range from 0.3 to 0.5 *SD*, and must therefore be characterized as

TABLE 8.4. Difference in Average Performance at Base Assessment Between Dropouts and Returnees[a]

	Sample 1 $N = 500$	Sample 2 $N = 997$	Sample 3 $N = 705$	Sample 4 $N = 612$
After Test 1				
Verbal meaning	4.07**	6.38**	6.27**	6.12**
Spatial orientation	2.52**	4.00**	4.08**	4.25**
Inductive reasoning	3.06**	5.28**	5.70**	6.70**
Number	1.97*	3.95**	5.16**	4.45**
Word fluency	3.06**	3.66**	4.84**	3.68**
After Test 2				
Verbal meaning	3.51**	1.97*	3.71**	
Spatial orientation	2.16**	1.35	5.09**	
Inductive reasoning	5.14**	2.54**	5.81**	
Number	2.13*	1.65*	2.87**	
Word fluency	2.41*	1.01	2.67**	
After Test 3				
Verbal meaning	4.10*	2.30*		
Spatial orientation	4.85**	0.48		
Inductive reasoning	4.35**	4.73**		
Number	0.58	1.89		
Word fluency	3.96*	1.16		
After Test 4				
Verbal meaning	4.72**			
Spatial orientation	3.45*			
Inductive reasoning	4.45*			
Number	1.35			
World fluency	4.25*			

[a]Data in T-score points.
*$p < .05$; **$p < .01$. Note: All differences are in favor of the returnees.

being of at least moderate effect size. Attrition effects after the second test and beyond are somewhat less pronounced, but remain of significant magnitude.

Before being overly impressed by the substantial differences between dropouts and returnees, we must note that the extent to which these differences will bias our projections based on the survivors only will depend on the proportion of dropouts; that is, if attrition is modest the effects will be small, or vice versa. Table 8.5 presents the net attrition effects (in T-score points) for our samples, showing different attrition patterns. As can be seen the effect is largest for those occasions when the greatest proportion of dropouts occur, and becomes smaller as the panels stabilize and remaining loss occurs primarily through death or disabilities.

The attrition data clearly indicate that parameter estimates of levels of cognitive function from longitudinal data, in the presence of significant attrition, will be substantially higher in most instances than would be the case if an entire population sample could be followed over time. It does not necessarily follow, however, that rates of change will be overestimated unless it can be shown that there is a substantial positive correlation between base level performance and age change. On the basis of the favorable attrition, we would expect modest negative correlations due to regression effects, even though the stability of the PMA variables is quite high. The relevant correlations are reported in Table 8.6, and suggest that there is no evidence for a more favorable rate of change for the higher scoring members of the panel. In fact, some of the larger negative correlations may reflect greater age changes on some variables (e.g., Number and Word Fluency) for the more able, implying that the longitudinal panel data might in some instances be overestimating the extent of age-related decline in the general population.

TABLE 8.5. Attrition Effects Calculated as Difference Between Base Means for Total Sample and Returnees[a]

	Sample 1 $N = 500$	Sample 2 $N = 997$	Sample 3 $N = 705$	Sample 4 $N = 612$
Attrition after T_1	38.2%	53.5%	51.9%	52.3%
Verbal meaning	1.52*	3.41**	2.97**	3.20**
Spatial orientation	0.96	2.14**	2.12**	1.40*
Inductive reasoning	1.17	2.82**	2.96**	2.22**
Number	0.76	2.11**	2.68**	2.46**
Word fluency	1.17	1.95**	2.51**	2.33**
Attrition after T_2	25.0%	16.0%	16.5%	
Verbal meaning	1.59	0.70	1.54	
Spatial orientation	1.02	0.48	1.40	
Inductive reasoning	2.33**	0.91	1.99	
Number	0.93	0.59	0.98	
Word fluency	1.09	0.36	0.74	
Attrition after T_3	8.0%	7.4%		
Verbal meaning	0.97	0.57		
Spatial orientation	1.12	0.12		
Inductive reasoning	1.03	1.14		
Number	0.16	0.47		
Word fluency	.94	0.29		
Attrition after T_4	6.6%			
Verbal meaning	1.21			
Spatial orientation	0.86			
Inductive reasoning	1.14			
Number	0.35			
Word fluency	1.09			

[a]Data in T-score points.
*$p < .05$; **$p < .01$. Note: All differences are in favor of the returnees.

TABLE 8.6. Correlations of Gain Scores With Base Scores

	Sample 1	Sample 2	Sample 3	Sample 4
At Test 1				
Verbal meaning	−.24	−.26	−.15	−.23
Spatial orientation	−.33	−.27	−.24	−.23
Inductive reasoning	−.18	−.26	−.20	−.17
Number	−.25	−.28	−.37	−.19
Word fluency	−.28	−.43	−.27	−.18
At Test 2				
Verbal meaning	−.17	−.10	−.15	
Spatial orientation	−.18	−.33	−.19	
Inductive reasoning	−.16	−.02	−.20	
Number	−.25	−.19	−.21	
Word fluency	−.23	−.36	−.17	
At Test 3				
Verbal meaning	−.10	−.20		
Spatial orientation	−.23	−.25		
Inductive reasoning	−.25	−.22		
Number	−.22	−.26		
Word fluency	−.08	−.22		
At Test 4				
Verbal meaning	−.08			
Spatial orientation	−.37			
Inductive reasoning	−.14			
Number	−.40			
Word fluency	−.29			

DIRECT ASSESSMENT OF PRACTICE EFFECTS

The possible inflation of longitudinal change estimates because of practice effects can be studied by comparing individuals at the same age who are retest returnees with the performance of individuals assessed for the first time at T_2. It will be seen from Table 8.7 that the apparent practice effects estimated in this manner, at first glance, appear to be impressively large. However, these comparisons involve the comparisons of attrited and random samples. The mean values for the longitudinal group must therefore be adjusted for attrition to permit a valid comparison. The appropriate values for this adjustment are not those from Table 8.4 (the differences between the survivors and returnees), but rather the mean differences between the returnees and the entire sample at base as shown in Table 8.5. Data for the raw and adjusted practice effects are provided in Table 8.7. Because we expect practice effects to be positive, all significance tests used in this instance are one-tailed. The raw practice effects appear to be statistically significant for virtually all variables and samples, but none of adjusted effects reach significance except for Verbal Meaning in sample 1.

TABLE 8.7. Raw and Attrition-Adjusted Effects of Practice by Sample and Test Occasion[a]

	Sample 1		Sample 2		Sample 3		Sample 4	
	Raw	Adj.	Raw	Adj.	Raw	Adj.	Raw	Adj.
From Test 1 to Test 2								
Verbal	2.83**	1.31*	4.02**	0.61	3.25**	0.28	1.21*	−1.99
Spatial	1.02	0.06	3.00**	0.86	2.28**	0.28	2.06**	0.14
Reasoning	2.03**	0.86	3.43**	0.61	2.77**	−0.19	1.33*	−0.69
Number	1.02	0.26	1.29*	−0.82	3.09**	0.41	1.60*	−0.86
Word fluency	1.71*	0.54	2.94**	0.99	2.97**	0.87	1.10	−1.23
From Test 2 to Test 3								
Verbal	−3.07	−2.77	3.11**	1.97*	−3.15	−4.46		
Spatial	−0.03	0.13	2.05**	0.65	1.68*	−0.06		
Reasoning	1.16	0.48	2.71**	1.34	1.93*	−0.80		
Number	−0.50	−0.08	1.84*	1.82*	1.13	−0.07		
Word fluency	0.22	−0.09	0.41	0.61	2.15**	1.23		
From Test 3 to Test 4								
Verbal	0.20	0.20	−0.78	0.79				
Spatial	0.14	−0.34	0.30	0.63				
Reasoning	−0.17	−0.97	0.04	0.09				
Number	−1.03	−0.18	0.10	0.59				
Word fluency	0.10	−0.80	−0.99	−0.34				
From Test 4 to Test 5								
Verbal	0.90	0.29						
Spatial	0.81	−0.41						
Reasoning	−0.09	−0.86						
Number	−1.83	−1.66						
Word fluency	1.74	0.05						

[a]Data in T-score points.
*$p < .05$; **$p < .01$.

Similar analyses can be conducted to assess the continuing effects of practice at additional assessment points in the study. Data are provided up to the fifth assessment occasion. It will be seen that a few raw effect estimates are significant for practice from T_2 to T_3, but attrition-adjusted effects remain significant only for Verbal Meaning and Number in sample 2. Neither raw nor adjusted effects reach significant levels for practice from T_3 to T_4 or from T_4 to T_5.

Effects of Adjustments for Practice and Attrition on Findings of Longitudinal Age Changes

The estimates for attrition-adjusted practice effects can, of course, be derived also for the age changes described in Table 8.2. To obtain maximum cell sizes, we have computed these estimates aggregated for all our subjects at each age for which all of the required 7-year longitudinal data were available. Because the primary concern here is empirical examples of the methods advocated for dealing with internal validity threats, I present data for this and the following analyses only for the variable of Verbal Meaning over the age range from 25 to 81 years. The attrition effects are obtained by finding the differences between dropouts and returnees at T_1. Raw practice effects are obtained by obtaining differences between means for returnees at T_2 and samples of the same age tested for the first time at T_2.

Again, small attrition effects are observed that reach statistical significance for the mean ages 39 to 60. Raw practice effects are statistically significant at all ages, but do not reach significance at any age when adjusted for attrition effects. Virtually all effects, however, are in a positive direction. When these values are used to adjust the raw longitudinal age changes, early increments (from 25 to 46) are no longer statistically significant, while significant decline is found to occur earlier, by age 60. I therefore plotted a comparison of data obtained from the cross-sectional averages (from Table 8.1), the unadjusted longitudinal changes estimated on the basis of all T_1/T_2 data (from Table 8.2), and longitudinal changes adjusted for practice and attrition effects (from Table 8.8). These adjusted values still represent within-individual change that is considerably less than would be suggested by the cross-sectional data, but do show steeper decline than the unadjusted longitudinal estimates.

The adjustments to the longitudinal age changes for practice and attrition shown are rather straightforward. As noted earlier, however, they used data aggregated across several samples studied over the same age range but at different time periods. The resultant adjustments consequently would be excessive if average secular trends had been negative or insufficient if such trends were positive. Likewise, positive selection (cohort) effects could result in overestimating the raw practice effects, whereas negative selection effects would result in underestimates. Data are now presented for analyses that ex-

Table 8.8. Adjustment of 7-Year Longitudinal Age Changes for Practice and Attrition Effects for Verbal Meaning[a]

Mean Age	N	Attrition Effect	Practice Effect: Raw	Practice Effect: Adjusted	Raw Age Change	Adjusted Age Change
25–32	135	+1.41	+1.96	+0.55	+1.42*	+0.87
32–39	169	+1.09	+2.39**	+1.30	+1.70**	+0.40
39–46	213	+1.77*	+2.27**	+0.50	+1.06*	+0.56
46–53	204	+3.81**	+3.61**	−0.20	+0.54	+0.74
53–60	223	+1.84*	+2.74**	+0.90	−0.28	−1.18**
60–67	170	+2.65**	+3.31**	+0.66	−1.59**	−2.25**
67–74	163	+1.48	+2.69**	+1.21	−2.12**	−3.33**
74–81	60	+3.11	+3.68**	+0.57	−1.79**	−2.36**

[a]Data in T-score points.
*$p < .05$; **$p < .01$.

emplify how practice and/or mortality effects can be disentangled when suitably crossed with the effects of history and selection.

Estimating History by Attrition Effects

Using ANOVA, overall attrition and history effects can be estimated, controlled for maturation, and estimates of history-specific and age-specific attrition effects can be provided. The 2 (attrition) × 3 (times-of-measurement) × 8 (age levels) time-sequential ANOVA employs 2205 subjects over the age

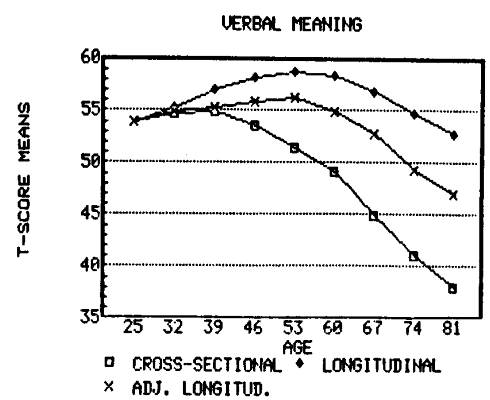

Figure 8.1. Comparison of cross-sectional age differences, raw longitudinal age changes, and longitudinal age changes adjusted for practice and attrition

range from 25 to 74 years who were first tested either in 1963, 1970, or 1977; attrition information was obtained from the second assessment of each sample 7 years later (1970, 1977, or 1984, respectively). Results of this analysis are reported in Table 8.9. All main effects are statistically significant, as is the age by history interaction. However, the age by attrition interaction is not significant, suggesting that attrition effects are randomly distributed with respect to age. The history by attrition interaction approaches statistical significance. While the overall attrition effect amounts to 3.60 T-score points, history-specific attrition effects were found to be 4.79, 2.84, and 3.14 points, respectively, for 1963, 1970, and 1977. I did not report age-specific attrition effects in this instance, because the age by attrition interaction failed to reach statistical significance.

Estimating Selection by Attrition Effects

Age-specific attrition effects can also be estimated, while controlling for the effects of maturation and selection effects, by a series of ANOVAs that cross any two adjacent ages and cohorts. The current example represents data for two 2 (attrition) × 2 (cohort) × 2 (age levels) cohort-sequential analyses with independent samples of 385 subjects for ages 53 and 60 and 366 subjects for ages 60 and 67. Data for the first cohort in each analysis come from the 1963 and 1970 data collections; the second cohort was assessed in 1970 and 1977. Attrition information on each of the four samples entering these analyses was obtained at the assessment that occurred 7 years beyond the data point entering the analysis. Results are shown in Table 8.10. The comparison at ages 53 and 60 yields significant selection and attrition effects, but no significant age effects. However, there is a significant age by selection interaction, revealing stability across age in the earlier born cohort but negative age differences in the later born cohort. In the comparison of samples at ages 60 and 67, main effects are significant for all three factors. The significant age by

Table 8.9. Time-Sequential Analysis of Variance Partitioning Effects Attributable to Age, History, and Attrition

Effect	df	Mean Square	F Ratio	p	Omega^2
Age	7	6534.87	67.89	< .001	.168
History	2	1097.83	11.41	> .001	.006
Attrition	1	8260.99	85.82	< .001	.054
Age × history	14	248.10	2.58	< .001	.014
Age × attrition	7	88.56	.92	ns[a]	—
History × attrition	2	264.45	2.75	< .06	.004
Age × history × attrition	14	79.80	.83	ns	—
Error	2157	96.26	—	—	—

[a]ns, not significant.

Table 8.10. Cohort-Sequential Analysis of Variance Partitioning Effects Attributable to Age, Selection, and Attrition

Effect	df	Mean Square	F Ratio	p	Omega^2
Comparison at Ages 53 and 60					
Age	1	5.47	.05	ns[a]	
Selection	1	622.82	6.11	<.01	.011
Attrition	1	2695.60	26.46	<.001	.087
Age × selection	1	2092.01	20.54	<.001	.134
Age × attrition	1	29.26	.29	ns	—
Selection × attrition	1	22.90	.22	ns	—
Age × selection × attrition	1	284.26	2.79	ns	—
Error	377	101.86	—	—	—
Comparison at Ages 60 and 67					
Age	1	1652.46	15.21	<.001	.048
Selection	1	788.34	7.26	<.01	.010
Attrition	1	1056.35	9.72	<.01	.034
Age × selection	1	400.96	3.69	<.05	.034
Age × attrition	1	116.58	1.07	ns	—
Selection × attrition	1	146.46	1.35	ns	—
Age × selection × attrition	1	7.27	.07	ns	—
Error	358	108.64	—	—	—

[a]ns, not significant.

selection interaction, again favors the earlier born over the later born cohort. Attrition effects in both analyses appear to be random with respect to age and selection. The magnitude of the attrition effects, however varies substantially across sets (4.77 and 3.01 T-score points, respectively).

Estimating History by Practice Effects

We can disaggregate history and practice effects while controlling for maturation by means of a time-sequential ANOVA that includes four samples per age level, two of which are assessed at first assessment and two that would have previously experienced the same measures. Data are presented for 2 (levels of practice) × 2 (times of measurement) × 7 (age levels) time-sequential analysis involving 1764 subjects aged 32 to 74 with assessments of independent random samples occurring in 1970 and 1977. Note that half of the subsamples were assessed 7 years earlier, in 1963 or 1970; the earlier data do not enter this analysis. This analysis (Table 8.11) is *not* controlled for effects of attrition, which, in this design, remain confounded with the practice effects. As seen in the earlier history by attrition analysis, no significant history effects are found between 1970 and 1977. As expected the age effect is highly significant, as is the practice effect. The magnitude of the latter (average raw practice effect) amounts to 2.61 T-score points. No significant in-

TABLE 8.11. Time-Sequential Analysis of Variance Partitioning Effects Attributable to Age, History, and Practice

Effect	*df*	Mean Square	*F* Ratio	*p*	Omega²
Age	6	7390.17	81.73	<.001	.218
History	1	4.77	.05	ns[a]	—
Practice	1	3754.57	41.52	<.001	.031
Age × history	1	158.91	1.76	ns	—
Age × practicen	1	128.85	1.43	ns	—
History × practice	1	6.36	.07	ns	—
Age × history × practice	1	1179.46	1.98	<.06	.023
Error	1761	90.42	—	—	—

[a]ns, not significant.

teractions are found between age and practice, but the three-way interaction is marginally significant. Theoretically, this interaction would reflect age-specific practice effects that differ by time of assessment. Magnitudes for these specific effects are not reported; they may occur primarily because of the failure to control for the attrition confound in this design.

ESTIMATING SELECTION BY PRACTICE EFFECTS

Age-specific practice effects, of course, can also be estimated while controlling for the effects of selection effects. A series of ANOVAs is used to cross any two adjacent ages and cohorts. This example has two 2 (levels of practice) × 2 (cohort) × 2 (age levels) cohort-sequential analyses with independent samples. The first analysis is of 625 subjects aged 53 and 60. The second analysis, with 571 subjects, compares ages 60 and 67. Data for the first cohort in each analysis come from the 1963 and 1970 data collections; the second cohort was assessed in 1970 and 1977. Two samples are involved at each data point. For exmple, one of the samples from cohort 1 tested in 1963 is at first test, while the other sample from this cohort was previously assessed in 1956. Likewise, one sample from cohort 2 tested in 1977 was tested earlier in 1970, and so on. Data from the earlier tests do not enter these analyses, and attrition is not controlled for.

Results of these analyses are shown in Table 8.12. Main effects are statistically significant for both analyses, although the selection effect is only marginally so in the comparison of ages 53 and 60. Of particular interest here are the age by practice and triple interactions. These are clearly significant in the first analysis, and marginally significant in the second. These results suggest that practice effects may indeed not be random with respect to age, particularly under specific selection conditions. Practice effects estimated in these analyses attain magnitudes of 3.16 and 3.75 *T*-score points, respectively.

TABLE 8.12. Cohort-Sequential Analysis of Variance Partitioning Effects Attributable to Age, Selection, and Practice

Effect	*df*	Mean Square	*F* Ratio	*p*	Omega²
Comparison at Ages 53 and 60					
Age	1	392.11	3.89	<.05	.010
Selection	1	295.41	2.93	<.08	.007
Practice	1	1872.14	18.56	<.001	.053
Age × selection	1	274.27	2.72	ns[a]	—
Age × practice	1	433.46	4.30	<.05	.022
Selection × practice	1	178.30	1.77	ns	—
Age × selection × practice	1	1358.81	13.47	<.001	.129
Error	617	100.84			
Comparison at Ages 60 and 67					
Age	1	916.48	8.46	<.00	.024
Selection	1	969.84	8.95	<.01	.026
Practice	1	2392.68	22.09	<.001	.068
Age × selection	1	82.26	.76	ns	—
Age × practice	1	341.60	3.15	<.07	.018
Selection × practice	1	155.58	1.44	ns	—
Age × selection × practice	1	359.25	3.32	<.07	.038
Error	563	108.33	—	—	—

[a]ns, not significant.

Note that at specific age and selection levels, practice effects range from a low of −.08 to a high of 8.33 *T*-score points.

It would have been possible in each of the two preceding examples to control for attrition effects by using in the "no practice" cells only those individuals who had been identified as having returned for further assessment of a subsequent test occasion. I do not provide examples for these further controls because data on age-specific attrition-adjusted practice effects have already been reported in the section on direct assessment of practice effects. Instead, the final example provides results of a design that directly crosses attrition and practice effects.

JOINT ESTIMATES OF HISTORY, SELECTION, ATTRITION, AND PRACTICE

The final example disaggregates the effects of attrition and practice from history and selection effects. This analysis requires subsamples examined at four different assessment points even though only data collected on two of these occasions are used; the other two occasions are required to establish two subsamples within each selection/history/attrition combination. One has been assessed on one more occasion than the other, and a subsample of each has remained in the study for one further assessment. The example represents a 2 (selection levels) × 2 (times of measurement) × 2 (practice levels)

× 2 (attrition levels) cross-sequential ANOVA with independent samples. Data from 606 subjects collected in 1970 and 1977, from two cohorts that were aged 46 and 53 in 1970 and aged 53 and 60 in 1977, respectively, are used. Each of the four history by selection combinations is divided into four further subsets. Two of these have not had prior testing, while two were previously assessed. Further, each is divided into one sample whose members left the study after the assessment included in this analysis; the other subset consists of individuals who returned for at least one other occasion. In this particular example, I am not interested in age-specific estimates, but wish to consider the relative magnitude of the four internal validity threats under study. For this reason, an age interval is selected in which earlier data suggest that maturational (age-related) change will be zero or trivial, and equal time intervals are used for all comparisons (cf. Botwinick & Arenberg, 1976). Results of this analysis are given in Table 8.13.

When we cross all four internal validity threats under these conditions, statistically significant main effects are found for history, practice, and attrition with a marginally significant effect for selection. However, none of the interactions is statistically significant. Over a 7-year interval, it appears that the largest effect is from attrition (2.62 *T*-score points in favor of the returnees), followed by history (2.48 *T*-score points in favor of the groups assessed at the earlier point in time), and practice (2.40 *T*-Score points in favor of the practiced groups). The selection effect amounts to 1.35 *T*-score

TABLE 8.13. Cross-Sequential Analysis of Variance Partitioning Effects Attributable to Time, Selection, Practice, and Attrition

Effect	*df*	Mean Square	*F* Ratio	*p*	Omega^2
Time	1	966.25	11.22	<.001	.021
Selection	1	283.91	3.30	<.07	.001
Practice	1	905.43	10.52	<.01	.026
Attrition	1	1082.07	12.57	<.001	.031
Time × selection	1	74.02	.85	ns[a]	—
Time × practice	1	77.00	.89	ns	—
Time × attrition	1	195.53	2.27	ns	—
Selection × practice	1	1.12	.01	ns	—
Selection × attrition	1	.33	.00	ns	—
Practice × attrition	1	151.69	1.76	ns	—
Time × selection × practice	1	2.24	.02	ns	—
Time × selection × attrition	1	10.72	.12	ns	—
Selection × practice × attrition	1	92.61	1.07	ns	—
Time × selection × practice × attrition	1	3.49	.04	ns	—
Error	569	86.10	—	—	—

points in favor of the later born cohort. These results might suggest that the validity threats investigated here may indeed be cumulative and, although independent of each other, must all be considered. The reader should be cautioned further that, as illustrated earlier, interactions between maturational level and the other variables considered are likely to be found both in young adulthood and advanced old age, those age levels at which positive or negative age changes are expected to occur for many cognitive variables.

Some Concluding Remarks

It is well known that cross-sectional studies of cognitive aging tend to paint unduly pessimistic pictures because positive selection (cohort) effects inflate age change estimates when modeled by age comparative studies (also see Willis, 1985, 1987). In reaction to these problems, there have been major efforts in recent years to obtain substantial longitudinal data bases that permit direct estimates of age changes. Much information can be gained from longitudinal data that is simply not available in cross-sectional studies (cf. Schaie, 1983c). What has not been given enough attention, however, is that longitudinal studies are plagued with even more complex internal validity threats.

In this chapter I have outlined the nature of these threats with specific reference to their import for studies of cognitive aging in adulthood. I have examined in detail experimental paradigms that allow the investigator to control for or assess the magnitude of effects for the validity threats of attrition, practice, history, and selection (cohort). As has been pointed out by Campbell and Stanley (1967), there is little that one can do to deal with these issues in "one-shot" studies, which should probably always be considered pilot efforts at best.

I have tried to show, however, that there are relatively simple methods available that can be applied in those situations in which at least two estimates are available for comparable groups, as is the case in many studies reported in the adult cognitive literature. Naturally, such data can only be investigated for a single threat at a time, and the investigator must use theoretical rationales to defend the selected paradigm. As the number of data points across time available to an investigator increase, so do the possibilities of investigating multiple validity threats. Again I have presented a number of alternatives and empirical examples that show why the pursuit of these matters is not esoteric at all, but may have important substantive consequences.

At least for several of the Primary Mental Abilities, attrition effects nearly always lead to an overestimate of performance levels, but not necessarily to an overestimate of rate of change. The matter of practice effects is more equivocal, because the more practiced group will always have experienced greater attrition than the less practiced group. Designs that can disaggregate both practice and attrition from aging effects are complex, however, and require multiple data points. When such disaggregation is actually accomplished, small practice effects tend to remain, suggesting that longitudinal data for some variables and over some age ranges may indeed overestimate gain in early adulthood and slightly underestimate loss in old age. However, selection effects confounding cross-sectional data remain of far greater magnitude in those instances where it has been possible to disaggregate selection and aging effects.

I conclude, then, that longitudinal parameter etimates, particularly when available for multiple data points, remain the preferred sources of our understanding of cognitive aging, if only because longitudinal designs provide the very data that can be used to apply the correctives needed to address the internal validity threats here considered.

Acknowledgments. Preparation of this chapter was partially supported by research grant AG04770-03 from the National Institute on Aging. The cooperation of staff and members of the Group Health Cooperative of Puget Sound who participated in this study is gratefully acknowledged.

References

Baltes, P.B. Cornelius, S.W., & Nesselroade, J.R. (1979). Cohort effects in developmental psychology. In J.R. Nesselroade & P.B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 61-88). New York: Academic Press.

Baltes, P.B., Nesselroade, J.R., Schaie, K.W., & Labouvie, E.W. (1972). On the dilemma of regression effects in examining ability level-related differentials in ontogenetic patterns of intelligence. *Developmental Psychology, 6,* 78-84.

Baltes, P.B., Schaie, K.W., & Nardi, A.H. (1971). Age and experimental mortality in a seven-year longitudinal study of cognitive behavior. *Developmental Psychology, 5,* 18-26.

Botwinick, J., & Arenberg, D. (1976). Disparate time spans in sequential studies of aging. *Experimental Aging Research, 2,* 55-61.

Campbell, D.T., & Stanley, J.C. (1967). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Cohen J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.

Cooney, T.M., Schaie, K.W., & Willis, S.L. (1988). The relationship between prior functioning on cognitive and personality variables and subject attrition in psychological research. *Journal of Gerontology: Psychological Sciences, 43,* 12-17.

Costa, P.T., Jr., & McCrae, R.R. (1982). An approach to the attribution of aging, period, and cohort effects. *Psychological Bulletin, 92,* 238-250.

Furby, L. (1973). Interpreting regression toward the mean in developmental research. *Developmental Psychology, 8,* 172-179.

Gribbin, K., & Schaie, K.W. (1979). Selective attrition in longitudinal studies: A cohort-sequential approach. In H. Orino, K. Shimada, M. Iriki, & D. Maeda (Ed.), *Recent advances in gerontology* (pp. 549-551). Amsterdam: Excerpta Medica.

Hays, W.L. (1963). *Statistics.* New York: Holt, Rinehart & Winston.

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

Hultsch, D.F., & Hickey, T. (1978). External validity in the study of human development: Methodological considerations. *Human Development, 21,* 76-91.

Labouvie, E.W. (1982). Issues in life-span development. In B.B. Wolman (Ed.), *Handbook of developmental psychology* (pp. 54-62). Englewood Cliffs, NJ: Prentice-Hall.

Nesselroade, J.R., & Baltes, P.B. (1974). Adolescent personality development and historical change: 1970-1972. *Monographs of the Society for Research In Child Development, 39* (1, Whole No. 154), 1-80.

Nesselroade, J.R., & Labouvie, E.W. (1985). Experimental design in research on aging. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging.* (2nd ed., pp. 35-60). New York: Van Nostrand Reinhold.

Nesselroade, J.R., Stigler, S.M., & Baltes, P.B. (1980). Regression towards the mean and the study of change. *Psychological Bulletin, 88,* 622-637.

Riegel, K.F., Riegel, R.M., & Meyer, G. (1967). A study of the dropout rates of longitudinal research on aging and the prediction of death. *Journal of Personality and Social Psychology, 5,* 342-348.

Salthouse, T.A. (1982). *Adult cognition: An experimental psychology of human aging.* New York: Springer-Verlag.

Schaie, K.W. (1959). Cross-sectional methods in the study of psychological aspects of aging. *Journal of Gerontology, 14,* 208-215.

Schaie, K.W. (1965). A general model for the study of developmental change. *Psychological Bulletin, 64,* 92-107.

Schaie, K.W. (1972). Can the longitudinal method be applied to psychological studies of human development? In F.Z. Moenks, W.W. Hartup, & J. DeWit (Eds.), *Determinants of behavioral development* (pp. 3-22). New York: Academic Press.

Schaie, K.W. (1973). Methodological problems in descriptive developmental research on adulthood and aging. In J.R. Nesselroade & H.W. Reese (Eds.), *Life-span developmental psychology: Developmental issues* (pp. 253-280). New York: Academic Press.

Schaie, K.W. (1977). Quasi-experimental research designs in the psychology of aging. In J.E. Birren and K.W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 39-58). New York: Van Nostrand Reinhold.

Schaie, K.W. (1982). Longitudinal data sets: Evidence for ontogenetic development or chronicles of cultural change? *Journal of Social Issues, 38,* 65-72.

Schaie, K.W. (1983a). *Longitudinal studies of adult psychological development.* New York: Guilford Press.

Schaie, K.W. (1983b). The Seattle Longitudinal Study: A twenty-one year exploration of psychometric intelligence in adulthood. In K.W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 64-135). New York: Guilford Press.

Schaie, K.W. (1983c). What can we learn from the longitudinal study of adult psychological development? In K.W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 1-19). New York: Guilford Press.

Schaie, K.W. (1985). *Manual for the Schaie-Thurstone Test of Adult Mental Abilities (STAMAT).* Palo Alto, CA: Consulting Psychologists Press.

Schaie, K.W. (1986). Beyond calendar definitions of age, time and cohort: The general developmental model revisited. *Developmental Review, 6,* 252-277.

Schaie, K.W., & Hertzog, C. (1982). Longitudinal methods. In B.B. Wolman (Ed.), *Handbook of developmental psychology* (pp. 91–115). Englewood Cliffs, NJ: Prentice-Hall.

Schaie, K.W., & Hertzog, C. (1985). Measurement in the psychology of aging. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 61–92). New York: Van Nostrand Reinhold.

Schaie, K.W., Labouvie, G.V., & Barrett, T. J. (1973). Selective attrition effects in a fourteen-year study of adult intelligence. *Journal of Gerontology, 28,* 328–334.

Schaie, K.W., Orshowsky, S.J., & Parham, I.A. (1982). Measuring age and sociocultural change: The case of race and life satisfaction. In R.C. Manuel (Ed.), *Minority aging: Sociological and social psychological issues* (pp. 223–230). Westport, CT: Greenwood Press.

Schaie, K.W., & Parham, I.A. (1974). Social responsibility in adulthood: Ontogenetic and sociocultural changes. *Journal of Personality & Social Psychology, 30,* 483–492.

Schaie, K.W., & Parham, I.A. (1975). *Manual for the test of behavioral rigidity.* Palo Alto, CA: Consulting Psychologists Press.

Schaie, K.W., & Willis, S.L. (1986a). *Adult development and aging* (2nd ed.). Boston: Little, Brown.

Schaie, K.W., & Willis, S.L. (1986b). Can decline in adult cognitive functioning be reversed? *Developmental Psychology, 22,* 223–232.

Schaie, K.W., Willis, S.L., Hertzog, C., & Schulenberg, J.E. (1987). Effects of cognitive training upon primary mental ability structure. *Psychology and Aging, 2,* 233–242.

Thurstone, L.L., & Thurstone, T.G. (1941). *Factorial studies of intelligence.* Chicago: University of Chicago Press.

Willis, S.L. (1985). Towards an educational psychology of the older learner: Intellectual and cognitive bases. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 818–847). New York: Van Nostrand Reinhold.

Willis, S.L. (1987). Cognitive training and everyday competence. In K.W. Schaie (Ed.), *Annual review of gerontology and geriatrics* (Vol. 7, pp. 159–188). New York: Springer.