The psychology of adult development is a branch of scientific psychology and thus shares the methodological concerns of the parent field. Adult developmental psychology uses the same kinds of data, generated by the same procedures, as do other fields of psychology. The formal experiment, the correlational study, the use of survey methodology are as much in evidence here as elsewhere. As is true for other areas of psychology, researchers in this field often have trouble finding an adequate control group to which to compare their experimental group, they have problems generalizing their results beyond the particular individuals whom they happened to observe, and they worry about statistical distortions (Birren and Birren, 1990; Birren and Schroots, 1996). There also are additional concerns, not always faced in other branches of psychology, that relate directly to the fact that in the developmental sciences we not only describe static phenomenon, but most importantly are concerned with the measurement of change over time. Finally, there are special methodological problems in doing research with older subjects (Teresi and Holmes, 1994).

CROSS-SECTIONAL AND LONGITUDINAL STUDIES

A central problem, not shared with other researchers, begins with the basic distinction between an age change and an age difference. An age change occurs in an individual as he or she grows older: At 60, for example, an individual’s reactions might not be quite as quick as they were at 25. An age difference is observed when one compares two people (or groups) of different ages. The person in our example at
age 60 might have slower reactions than another person of age 25. We are often interested in age changes (what happens as people grow older), but have data only from research on age differences (where the performance of different age groups is compared). In that case we must be quite cautious in not assuming that age differences necessarily reflect age changes. As we shall see, sometimes they do, and sometimes they don’t.

The distinction between age changes and differences corresponds to the data generated by two different experimental designs common in the study of development (see Figure 5-1). A cross-sectional design compares groups of different people varying in age at the same time. It allows us to record age differences. A longitudinal design, on the other hand, involves the observation of the same individuals at two or more different times; these kind of data represent age changes. To illustrate, two researchers might use the different designs to investigate income across the normal working life. One, using a cross-sectional design, compares the average

![Figure 5-1](https://example.com/figure5_1.png)

**Figure 5-1** Cross-Sectional Studies Test Different Groups of Different Ages, All at the Same Time. Longitudinal Studies Test the Same Group at Different Ages, Requiring Several Tests at Different Times.
income of representative samples of 25-year-olds, 30-year-olds, 35-year-olds, and so forth, up to the age of 65. The second researcher follows one sample of people, checking on their income every five years, at age 25, 30, 35, and so forth.

One major difference in the two experimental designs is immediately apparent. The cross-sectional study can be completed in a single day, whereas the longitudinal study will take 40 years! It is not surprising therefore, that many more cross-sectional studies can be found than longitudinal studies.

Age and Cohort

The differences between cross-sectional and longitudinal designs can best be understood by considering the several dimensions involved. One, of course, is age, which is a variable in both designs. A second variable is cohort, which is usually defined as the people born in the same year (or range of years). If you were born in 1980, you are in the 1980 cohort; your parents probably belong to a 1950s or 1960s cohort. Cohort is similar to the concept of generation; it is used to distinguish people by time of birth (for formal distinctions between the concepts of cohort and generation, see Riley, 1992; Riley, Foner, and Riley, 1999).

Now let us consider the two designs in more detail. The subjects in a longitudinal design all belong to the same cohort, whereas the subjects in a cross-sectional design belong to several different cohorts, each representing a different age group. As a result, the age differences produced by a cross-sectional study could be interpreted as being due to age or they may be due to cohort differences. Imagine a cross-sectional study of liking Frank Sinatra’s music. It would not be surprising to find that older people like Sinatra more than do younger people, but you might be unwilling to conclude that a liking for Frank Sinatra increases as one gets older. Most of the age difference, you would therefore assert in this case, is likely to be an effect of cohort, not of age. People who are old today were members of the “younger generation” who adored Sinatra when he was young. Today’s cohorts will grow old and, 40 years from now, perhaps will listen with rapt attention to a gray-haired Jewel!

Cohort effects often become apparent by comparing results from cross-sectional and longitudinal studies. For example, the Census Bureau often publishes statistics on average income in various age groups. Graphs of these statistics usually look like Figure 5-2: showing an increase to middle age and then decline until retirement. These statistics are often interpreted as reflecting the normal course of an individual’s earnings over the work life, but they are cross-sectional data on age differences, subject to misinterpretation. Longitudinal studies typically shows that individuals studied over their work lives almost always increase their earnings right up to the time of retirement (Schultz, 1995). The cross-sectional data mislead us because each new cohort made more money at every stage of life than did preceding cohorts. Thus, though men between 55 and 64 typically earn more than they ever did earlier in their lives, their earnings are less than those of men between 45 and 54 who started at higher salaries and received greater increases than the older men (U.S. Bureau of the Census, 1999).
Figure 5-2  Age and Cohort Effects. (a) Median income of white males by age, 1998. (b) If each generation earns ever-increasing income, but also each generation makes more at each stage of the life span, a cross-sectional study can misleadingly make it appear that income drops toward the end of one’s career.
Age and Time

Longitudinal studies observe members of the same cohort at two or more points in time. The data that they produce are age changes, and their interpretation is not clouded by possible differences between cohorts. However, longitudinal studies have their own difficulties, not the least of which is the large expense, in both time and money, necessary to mount a well-designed longitudinal investigation. Nor are longitudinal studies free from potential misinterpretation. The external validity (generalizability) of the findings is affected by the time of measurement (i.e., the period of years over which the study is conducted). Findings of change in longitudinal studies can be attributed either to a true developmental change of the sort we want to discover (a change with age in intellect or memory or personality) or to something else, some event or societal change that happened between the first and subsequent times that the subjects were observed that has nothing to do with advancing age (Schaie, 1982; 1988b, 1996b).

Suppose you had conducted a longitudinal study of attitudes toward the war in Vietnam during the late 1960s and early 1970s. You might well have found that your subjects, tested first in 1965, then in 1970, and finally in 1975, had become increasingly intolerant of our nation’s involvement in Southeast Asia. What does your finding mean? Does it mean that people become less accepting of war as they grow older? (Is it a true developmental change?) Or, as is considerably more likely...
in this case, does it mean that social-historical events occurred between surveys that resulted in a change in many individuals’ attitude toward the war?

Practice effects are also a problem in longitudinal studies. A longitudinal study involving tests (e.g., IQ tests) requires repeated testing with the same measures. What may look like an increase in the skill being assessed over time could be no more than increasing familiarity with the test items.

Cross-sectional studies are not affected by time of measurement issues, because all subjects are observed at a single time. In summary, cross-sectional studies confound age and cohort or generational effects; but longitudinal studies confound age and time-of-measurement effects.

**Time-Lag Design**

Another type of study, called a time-lag design, involves the observation of people of the same age at different times (Schaie, 1977, 1988b, 1996b). Suppose you are interested in the sex life of 20-year-olds and, specifically, in the differences between such activities today and in 1930. Assuming someone had surveyed young people on this issue in 1930, you could survey a new sample of young people today and compare the data. Age is held constant; it is not a variable. The researcher, in this instance, is typically interested in the direct study of cohort effects; that is, differences between generations. The confounded variable is now time of measurement. If you found that today’s 20-year-olds report more sexual activity, perhaps generations truly differ; but alternatively it is possible perhaps, between Time 1 and Time 2, that society has become more open and honest about reporting sexual activity.

**SAMPLING DIFFICULTIES**

One of the chief impediments to sensible conclusions about research in adult development and aging is the difficulty in gathering (and holding on to) appropriate groups of subjects. Subjects are generally presumed to be a sample of a larger population. The sample that we want to obtain is one that is representative of the larger group. But there are many ways a sample can become unrepresentative, making generalization of the results uncertain, perhaps invalid.

A representative sample will produce data that can be generalized to a larger population. The people who run opinion polls have an obvious need for representativeness in their samples, for they wish to generalize to very large groups, often the entire population of the United States, or another country. In voter surveys, the population is often considered to be the “people eligible to vote” or “people likely to vote.” The goal of such surveys is to predict the actual vote on election day or at least to determine what the vote would have been, had the election been held on the day the survey was taken. Unrepresentative samples can result in misleading data, showing one candidate ahead when in fact he or she is far behind. In some polls taken in the United States early in this century, Republicans were overrepresented because telephones were used to solicit opinions; in those days, Republicans were typically richer and were more likely to have phones.
In studies of adult development, people with greater income, more education, and better jobs are usually included in samples more often than less fortunate individuals. More affluent people are easier to find (they belong to clubs and organizations) and easier to persuade to participate (they are more likely to believe in the value of research and are proud of themselves and their lives). Thus, many of the samples providing research findings discussed in this book are not truly representative of the adult population in general, but primarily of its middle class. In addition, men also have been more frequently studied than women during young and middle adulthood, although the reverse is true in old age. The debate regarding the factual basis of the so-called midlife crisis, for example, was based almost entirely upon the life experience of middle-class males (Levinson, 1978; see Chapter 3). More recent studies of women, minorities, and the less affluent, suggest that the so-called midlife crisis may be limited to certain historical periods and specific subgroups of the population. We must therefore be cautious in generalizing research findings beyond the population of which a particular sample can reasonably be called representative.

Researchers who do longitudinal studies face particular problems in sampling. Not only do they need to find a representative sample, they also have to keep it—to recover it for each subsequent retesting. Between observations, some people die (especially if the subjects are elderly), some people move away (and may not leave a forwarding address), and some people refuse to participate in the next round of tests. Usually this subject loss makes the sample less representative, for the subjects who die, move away, or become uninterested often have less education, less income, and less prestigious jobs. In short, if the sample was not biased toward the middle class to begin with, it is likely to become so by the time the longitudinal study is completed (Cooney, Schaie, and Willis, 1988; Schaie, 1988b, 1996b; Sharma, Tobin, and Brant, 1986).

Sampling also affects studies that examine the interrelationship of different attributes and psychological constructs. If a sample is very homogeneous (restriction of range) it may be quite difficult, for example, to demonstrate a strong relationship that might readily appear if a more heterogeneous sample had been used. Likewise, the relationship of two variables may differ, depending upon the segment of the population that has been sampled (Maitland, Intrieri, Schaie, and Willis, 2000; Nesselroade, 1988a; Schaie, Maitland, Willis, and Intrieri, 1998). In longitudinal studies, it is particularly important to represent a broad variety of different genetic and environmental conditions, in order to be able to examine a broad spectrum of different aging patterns (see Kruse, Lindenberger, and Baltes, 1993).

**SPECIAL PROBLEMS IN THE MEASUREMENT OF CHANGE**

If we want to be certain that we know what it is that changes, we must of course have reliable tests. But no test can be designed to be perfectly reliable and not every person will perform reliably at all times on a given test. One of the common critiques of looking at change scores therefore is that the error of measurement for
the test given on two occasions is multiplicative, and that change scores consequently tend to be less reliable than measurements taken on any single occasion. In particular, there tends to be regression to the mean. That is, high scores on the first occasion are probably above the subjects’ true scores and low scores are probably below the subjects’ true scores. As a consequence high scores tend to become lower on the second occasion, while low scores become higher. This problem becomes less severe as change is studied over more than two occasions because observed scores will vary randomly about the true scores (Nesselroade, Stigler, and Baltes, 1980). Other methodologists argue that perhaps we should not look at group differences on two occasions, but rather study individual change profiles over multiple occasions. Such an approach is called “growth curve” measurement and allows individuals to be grouped into types that show similar patterns of change over time (Collins, 1996; Rogosa, Brandt, and Zimowsky, 1982). We can then look for predictors that inform us about the reasons why these patterns of change differ across individuals (Willett and Sayer, 1994).

In many longitudinal studies, however, there are only two data points. Also, in many instances it is indeed the average change over two occasions that needs to be assessed, particularly if we want to measure the effect of some natural or planned intervention into the developmental progress of groups of individuals. In that case, it is important to be doubly certain that our basic measurement tools are as reliable as possible.

SEQUENTIAL STUDIES

Many of the interpretational difficulties that are part and parcel of a simple cross-sectional and longitudinal designs can be alleviated with more complex approaches called sequential designs, shown in Figure 5-3 (see also Schaie, 1965, 1977, 1988a, 1996b). A cross-sectional sequence consists of two or more cross-sectional studies, covering the same age range, executed at two or more times. For example, we might compare age groups ranging in age from 20 to 80 in 1980 and then repeat the experiment in 2000 with a new sample of subjects in each age group, still from 20 to 80. A longitudinal sequence consists of two or more longitudinal studies, using two or more cohorts. Suppose we begin a longitudinal study of age changes between 20 and 80 by observing 20-year-olds in 1990 (the 1970 cohort) with the plan of observing each of these subjects again, every 10 years until they are 80. This is a simple longitudinal study. In 2000, however, we begin a second longitudinal study of the same age range, choosing a new sample of 20-year-olds from the 1980 cohort. These two longitudinal studies are the simplest case of a longitudinal sequence.

Schaie’s “Most Efficient Design”

K. Warner Schaie (1965, 1977, 1994b, 1996b), who has conducted extensive studies of research methodology in adult development, has proposed a “most efficient design,” which generates data that is useful for many informative analyses.
The most efficient design is a combination of cross-sectional and longitudinal sequences, created in a systematic way. In brief, the researchers begin with a cross-sectional study. Then, after a period of years, they retest these subjects, which provides longitudinal data for several cohorts (a longitudinal sequence). At the same time, they test a new group of subjects which, together with the first cross-sectional study, forms a cross-sectional sequence. This whole process can be repeated over and over (every 5 or 10 years, for example) with retesting of old subjects (adding to the longitudinal data) and first testing of new subjects (adding to the cross-sectional data).

**Figure 5-3** Longitudinal and Cross-Sectional Sequences. Longitudinal sequences use two or more cohorts; cross-sectional sequences test the same age groups at two or more times, using independent samples.
As an illustration, suppose we were to give IQ tests to four groups of people, ranging in age from 30 to 60 in 1980 (see Figure 5-4). This is a straightforward cross-sectional study. In 1990, we retest as many of these same subjects as can be found. These data will show, for each of 4 cohorts, what happened to average IQ scores as the subjects grew 10 years older. At the same time, we recruit new subjects in the same age groups as the original subjects and test their IQ for the first time. We would probably add a new cohort of people who, at the time of second testing, are 30 years old to make the second cross-sectional study comparable to the first. The second cross-sectional study should yield results similar to the first; if it does not, we will have interesting clues to the nature of intellectual development. In 2000, we retest our first sample (the one that started in 1980) for the third time, adding more data to our longitudinal sequence. We also retest the subjects who were new in 1990, adding an entirely new longitudinal sequence. Finally, we recruit new subjects to form a third replication of our cross-sectional study. As you can see, we will have generated a wealth of data.

Figure 5-4  Schaie’s Most Efficient Design. In 1980 four groups are tested (a cross-sectional study). They are retested in 1990 and 2000 (a longitudinal sequence with repeated measures). New groups from the same cohorts are first-tested in 1990 and in 2000 (cross-sectional sequences, independent samples). These new groups are later retested to form new longitudinal sequences.
Analyses

Data collected by using Schaie’s most efficient design or comparable designs can be analyzed in several ways. The approach of greatest interest to developmental psychologists is to contrast age changes against cohort effects (Schaie, 1996b; Schaie and Baltes, 1975). To do this, we need at least two cohorts, and we must observe each cohort for at least two different ages. To simplify our illustration, consider only the changes in IQ scores between the ages of 60 and 70. This analysis allows us to test for the presence of irreversible decrement as contrasted to changes in performance levels across successive cohorts. The people in our hypothetical study (see Figure 5-4) who were 60 in 1980 (the 1920 cohort) and 70 when retested in 1990 are compared to the people who were 60 in 1990 (the 1930 cohort) and 70 in 2000. Do their IQ scores increase, decrease, or remain stable over the 10-year period? If there is irreversible decrement with increasing age, we should find similar decline patterns for both cohorts. In a traditional longitudinal study, we would have data only for a single cohort and would not know, therefore, whether the observed change can be generalized beyond the specific cohort that was studied. For example, one cohort may show an increase while the other shows a decrease, or one cohort may increase at a slower rate than the other. One cohort may have a higher average IQ than the other at both 60 and 70, though the increase or decrease may be similar for the two cohorts. Obviously, a lot of interesting comparisons can be made from this type of analysis, which is called a cohort-sequential analysis (Figure 5-5).

Another type of analysis is called cross-sequential. As the cohort-sequential analysis contrasts cohort effects against age effects, the cross-sequential analysis contrasts cohort effects against time of measurement. At least two cohorts are compared at two or more times of measurement. This strategy is particularly appropriate for data sets that fit the adult stability model discussed earlier. No age changes are expected, and the primary interest turns to identifying the presence and magnitude of cohort and time-of-measurement effects. The cross-sequential analysis is helpful when the researcher is interested in, say, the effects of some event or sociocultural change that occurs between the two times of measurement and, in addition, suspects that different cohorts might react differently. For example, the effects of the sexual revolution culminating by the 1970s might be compared for a cohort whose members were in their twenties in 1970 and a second cohort whose members were in their early forties in 1970. In addition, if there is reason to suppose that time-of-measurement effects are slight or nonexistent, cross-sequential analysis can be used to estimate age changes because subjects are obviously older at the second time measurement.

If the cohort-sequential analysis contrasts cohort against age, and the cross-sequential analysis contrasts cohort differences against time of measurement, we have one logical possibility left: the time-sequential strategy that contrasts age against time of measurement. People of at least two different ages are compared at two or more times of measurement. Consider a study of changes over time in the generation gap: the attitudes of 50-year-olds are compared with those of 20-year-olds, both in 1970 and in 1990. We might find that the difference between the age groups narrows between 1970 and 1990, or perhaps both age groups become more liberal.
in their attitudes, but the gap between them remains sizable. The time-sequential method is also appropriate for a test of the decrement with compensation model. When a new compensatory method is introduced (for example, a computerized memory prosthesis or a drug affecting declining memory), the time-sequential method would show that age differences over the same age range would be smaller at Time 1 than at Time 2.
Repeated Measures Versus Independent Samples

In a typical longitudinal study, repeated measures are taken of the same persons at different times. Another possibility, however, is to use the same research design but with independent samples at each point on the longitudinal time scale. If we are interested in intellectual development, for example, we might begin a longitudinal study by testing the IQ of 30-year-olds, with plans to retest these same individuals every 10 years; these would then be repeated measures. The alternative would be to draw a new (independent) sample from the same cohort every 10 years. Thus, in 10 years, when our 30-year-olds are 40, we find a new representative sample of 40-year-olds, instead of retesting the old batch. The independent sampling approach works well when a large sample is drawn from a large population, and is commonly used in large-scale demographic or sociological surveys. If small samples are used it is, of course, necessary to make sure that successive samples are matched on factors such as gender, income, and education to avoid possible differences due to selection biases.

What do we gain from the independent-samples procedure? First, we gain a replication of sorts of the repeated-measures study, a second look at the same trends. If, for example, our typical longitudinal study shows decreases in average IQ from age 30 to 60, the independent-sample study should show the same thing. Of course, we cannot use the independent-sample data to follow a particular individual. We cannot say that Joan Doe increased her IQ scores, nor can we say that 87 percent of our subjects decreased and 13 percent increased. But we can say that the averages for the cohort decreased or increased.

In addition, independent samples allow us to estimate the effects in the repeated-measures study of such problems as losing subjects due to their inability or unwillingness to be retested; practice effects can also be examined. The independent samples are new each time and thus reflect what the composition of the single sample of the repeated-measures study would have been if no subjects had been lost between testing. As a new sample, the subjects have not taken our tests before, so practice effects can be estimated by comparing this “no practice” group with the repeated-measures sample, who practice each time the measures are repeated.

Finally, if the longitudinal study is a longitudinal sequence (that is, if at least two cohorts are studied), then the independent-samples replication will form a cross-sectional sequence, providing valuable information for cross-sectional analyses at each testing period. Suppose we begin our longitudinal study with four cohorts, ranging in age from 30 to 60, retesting each cohort every 10 years. In 10 years, in addition to the retestings, we also test new, independent samples for each cohort. The retestings are longitudinal but the new testings form a neat, cross-sectional study of the age range 40 to 70. One could add a new cohort, a new group of 30-year-olds, and begin a second longitudinal sequence, to be retested for the second time. Thus the independent samples comprise (1) a replication of the original longitudinal study, (2) a new cross-sectional study, and (3) the start of a new longitudinal sequence. It is not surprising, then, that independent samples are included in Schaie’s “most efficient design.”
It is not possible to conduct true experiments to assess the aging processes because we cannot assign research subjects randomly to different age levels. Nevertheless, a number of experimental designs (more properly called quasi-experiments [Campbell and Stanley, 1963; Cook and Campbell, 1979]; also see Mertens, 1998) can give us important insights and make it possible to examine alternative explanations for a variety of aging phenomena, provided that we keep in mind what such studies can and cannot tell us. As always, we must also be alert to the precautions that are necessary in these and in all experimental studies. In this section, we will briefly discuss three types of research designs that are commonly found in the aging literature and that the reader will encounter in many of the studies discussed in this book: age-comparative experiments, single age group intervention designs, and molar equivalence-molecular decomposition experiments.

**Age-Comparative Experiments**

One way of identifying the processes thought to be involved in age changes in behavior is to conduct a study in two groups of subjects of different ages that are matched on as many demographic variables as possible. We then assess the performance of the members of a younger and an older group on some behavior, for example, sorting a variety of objects into classes of objects. Suppose we find that the younger persons score higher on average. We now introduce a manipulation designed to improve performance. Perhaps we teach all of our subjects strategies for identifying various aspects of the objects that would make sorting the objects easier. We then retest all of our subjects to see whether there is differential improvement by age. This is what is called a nonequivalent control group, pretest-posttest design (Cook and Campbell, 1979).

Figure 5-6 shows the design of a study in which we compare the effects of an intervention on the older and the younger groups. Note that in the age-comparative experiment it is not particularly interesting to demonstrate that there is an age difference; we take that for granted. Neither are we concerned with demonstrating the effect of our intervention; this is also taken for granted, or we have chosen the wrong kind of manipulation. The critical effect that is of interest to us is the age by treatment interaction. If we can show that the older group gains significantly more than the younger group, we then argue that the age difference in the pretest was due to the fact that the older group performed less well than the younger because it lacked the sorting strategies needed to do as well as the younger group. In other words, if we can show age-differential effects of an intervention, we can then argue that we have determined a factor that explains the age difference in behavior.

**Single Age Group Interventions**

There are a number of instances in which intervention experiments might be conducted in a single age group because it does not make sense to have a younger comparison group (Krauss, 1980; Willis, 1987; Willis and Schaie, 1994b). For
example, suppose that we are interested in determining whether some educational procedures might be useful to reverse cognitive decline in older persons. In this instance, it would not make sense to have a younger control group, because the nature of the training effects would not be comparable. The training effect for the younger persons would involve enhancement of the initial level of functioning, while for the older persons it would involve the remediation of losses from a previous level of functioning. The most effective design in this case would be to study older people on whom we have longitudinal data over a reasonable period of time. We could then compare the effects of our intervention on individuals who had remained stable over time with the effects on individuals with known cognitive decline (see Chapter 12 for further discussion).

Single age group intervention studies, of course, must have suitable control groups. We usually require two types of control groups. The first receives the pretest and posttest but does not receive any intervention. The second receives the tests and some alternate or neutral procedure to measure the effect of an irrelevant intervention; the latter group is sometimes called a contact control group. A good intervention design, moreover, will assess not only the behavior that we wish
to modify, but also some other behaviors, which should not be affected by our intervention. This type of assessment is needed to prove that our intervention is targeted to a specific behavior, rather than being a general, nonspecific intervention.

**Molar Equivalence–Molecular Decomposition Experiments**

Some older people do every bit as well as younger persons on some behaviors, even though they seem to have declined as much as their age peers on other psychological functions that seem superficially to be related to the behavior on which they are successful. The experimenter attempts to discover what it is that these older persons might do to compensate for declines in the basic functions that may have occurred. The experimental strategy is to find groups of younger and older people that can be matched on a complex (molar) behavior, so that the correlation between that behavior and age is zero. The experimenter then breaks down the components of the molar behavior (molecular decomposition) to show that there is indeed age-related decline in some one component, but that this decline is compensated for by enhanced skill in another component of the molar behavior (Charness, 1983; Salthouse, 1979, 1987).

An example of the *molar equivalence–molecular decomposition* experiment is a study by Salthouse (1984) in which he examined the behavior of transcription...
typists. He investigated typing, because it is a skill that is widely represented over
great ranges of ability and age. Moreover, although typing is an integrated activity,
some of its aspects can be reliably “decomposed” into distinct factors such as
choice reaction time, tapping speed, acquisition of information in one form (read-
ing the text to be transcribed), and transcription of information in another form
(the actual typing). Salthouse found that both choice reaction and tapping time
were correlated with age; the older typists were slower on these components. How-
ever, the older typists were more proficient in acquiring the information to be
typed. They tended to pick up (read) larger chunks of information to be tran-
scribed, reducing the time that they needed to acquire the information and thus
compensated for their slower motor speed in hitting the typewriter keys.

MEASUREMENT ISSUES IN STUDIES OF ADULT DEVELOPMENT

The concepts about measurement that you might encounter in any tests and mea-
surements course in psychology all apply equally to the field of adult development.
We are concerned with the reliability and validity of formal test instruments and we
pay attention to the problems of making objective observations and transforming
the resultant qualitative data into quantifiable form (Dunn, 1989; Ikels, Keith, and
Fry, 1988; Keith, 1988). There are, however, some additional issues that bear on
the validity of our research findings when we study changes in behavior over time
and across age groups.

Direct Observations and Latent Constructs

Many developmental psychologists are concerned more with studying
changes and differences on latent (abstract) constructs, such as anxiety or intelli-
gence, than in the specific behaviors or tests that are thought to represent these
constructs. Nevertheless, the only way we can measure constructs is to observe the
performance of subjects on tests or to observe their behavior in specific situations,
from which we then infer the individual’s standing on the abstract construct that is
not directly measurable. One of the problems in studies of adulthood is that there
is no guarantee that the same observation or test will be an equally valid representa-
tion of the same construct over a wide age range. Studies with the Wechsler
Adult Intelligence Scale, for example, early on showed that the Digit Span subtest
of that scale is related to a general ability factor in young adulthood, but to a mem-
ory factor in old age (Cohen, 1957).

More recent work with the Primary Mental Abilities, comparing the relation
of observed tests to the underlying ability factors, has shown that there are shifts
over the adult life span in the relative efficiency with which specific tests measure
the latent ability constructs (Schaie, Maitland, Willis, and Intrieri, 1998; Schae,
Willis, Jay, and Chipuer, 1989). Similar issues arise when we study gender differ-
ences across the adult life span (Maitland, Intrieri, Schae, and Willis, 2000). Un-
less we can show that relationships between the observables and their latent
constructs are stable across age and time, it may be that reported changes in per-
performance levels result from comparing apples and oranges.
These relationships are demonstrated typically by a technique known as confirmatory factor analysis. In this method, we specify a model that demands that the regression (correlation) of the observed variables or behaviors with the latent constructs are the same across different ages or different age groups. The fit of this model to a particular data set is then estimated. If the difference in regression coefficients is less than chance, we can accept the equivalence of the measures across age or time. If the equivalence model does not fit, we would then test a weaker model, in which we require that the same observed variables represent the factors across age or time, but do not demand that the values of the regression coefficients be identical (see Horn and McArdle, 1992; Maitland, Intrieri, Schaie, and Willis, 2000; Meredith, 1993; Schaie, Maitland, Willis, and Intrieri, 1998).

**Generalizability of Research Findings**

Whether findings from research can be broadly applied depends to a large extent upon the representativeness of the sample used. In work on adult development, however, representativeness goes beyond merely seeking a reasonable representation of various demographic dimensions. To generalize across age and time, we need to attend to the fact that populations shift over time and that whenever we compose a sample that differs in age, we are dealing with individuals who belong to different populations (Nesselroade, 1988b; Riley, Foner, and Riley, 1999). The dramatic shifts in levels of education, occupational status, income, and mobility characteristics across successive cohorts affect virtually every behavior studied by psychologists (cf. Atchley, 1989b; Schaie, 1996a, 1996b; Willis, 1989a). Thus, any study that matches these characteristics exactly across age groups will at the same time introduce distortions in the degree to which the subsamples are representative of their cohort.

There are also problems in generalizing from special samples of older people with known pathologies to healthy, community-dwelling individuals. While it is impossible to equate young and old groups in terms of the incidence of certain diseases, it is still necessary to worry about the impact of disease on behavior in age-comparative studies (Bosworth, Schaie, and Willis, 1999; Elias, Elias, and Elias, 1990; Elinson, 1988; Solomon, 1999). In any event, it is always necessary to note the demographic and health status characteristics of the participants in any study, so that we can understand the degree to which findings may be relevant to populations with different characteristics.

**DEVELOPMENTAL RESEARCH**

The primary goal of developmental psychologists is to describe and explain age changes. To do this within a practical time frame, they often try to estimate age changes by observing the age differences obtained in cross-sectional studies. This approach works reasonably well if cohort differences in the variable under study are slight. The assumption of trivial cohort differences, however, is unreasonable when young adults and the elderly are compared. But it makes sense when
comparison is made of relatively small age slices, for example, comparing the late middle-aged with young-old persons or comparing persons in their twenties with those in their thirties. Alternatively researchers use longitudinal studies that directly record age changes for a variety of individuals. But longitudinal studies are expensive and very time-consuming. In addition, the pure effects of age may be masked by social-historical changes between measurements. Large segments of the original sample may die, move away, or refuse to be retested, changing the sample in significant ways that hinder the researchers’ ability to generalize their results to a larger population. Tests that were used on 20-year-olds in 1940 may also be considered obsolete by the time the subjects were retested as 80-year-olds in 2000.

Longitudinal studies, nevertheless, have the great advantage of directly measuring changes in characteristics such as an individual’s intellect or personality or brain function as he or she grows older. Longitudinal studies produce data that provide at least partial answers to such questions as: Do some people increase in intelligence as they grow older, and some decrease? How much variability is there? What causes the variability? Does intelligence increase for people in “intellectual” jobs and decrease for people in “manual” jobs? Does an increase in blood pressure relate to a decrease in intelligence? As expensive and time-consuming as longitudinal studies are, they provide data that are among the most important for theories in developmental psychology.

Developmental research, which covers the life span, encounters somewhat different difficulties at different stages of life. In a child development textbook, for example, you will find a discussion of the difference between cross-sectional and longitudinal studies, but the sense of drawing dangerous conclusions about maturation from cross-sectional studies will be less. In child studies, cross-sectional and longitudinal studies often produce similar results. One reason is that these studies cover a smaller age range. For example, 2-year-olds might be compared to 4-year-olds, and it is not likely that cohort differences will be pronounced in a span of two years. Also, many of the age changes in childhood are based on biological changes; the maturation of the brain and nervous system or the sexual-reproductive system, for example, accounts for fairly uniform changes in intellect and sexual activity among children. Adulthood, on the other hand, covers 60 years of an average 80-year life span, and social events rather than biological events are often the primary determinants of change (Hagestad and Neugarten, 1985; Riley, Foner, and Riley, 1999). Thus, cross-sectional studies of adults, which confound cohort effects with age effects, are fraught with danger; the time span is longer and the variability of social causes is much greater than the variability of biological causes. Longitudinal studies are much more important, therefore, in the study of adult development than they are in child development.

SUMMARY

1. Cross-sectional designs compare several age groups (cohorts) at the same time. They yield data on age differences. Longitudinal designs compare the same cohort at different times. They yield data on age
changes. Longitudinal studies suffer from subject loss, practice effects, and historical changes that affect behavior, but cross-sectional studies are more prone to confusion of age effects with differences between generations (cohorts). Difference scores tend to be less reliable than scores on single occasions; alternative methods require three or more measurement points. A major research problem in adult development is finding representative samples; white, middle-class males are generally overrepresented in the research literature, but in advanced old age women are overrepresented. Longitudinal studies also have the problem of keeping the sample representative once it is recruited; lower-class subjects, for example, tend to drop out in disproportionate numbers.

2. Sequential designs are complex combinations of the simple cross-sectional and longitudinal designs. A cross-sectional sequence consists of two or more cross-sectional studies run at different times. A longitudinal sequence consists of concurrent longitudinal studies of two or more cohorts. Schaie’s most efficient design includes both cross-sectional and longitudinal sequences formed by retesting the subjects of an earlier cross-sectional study while testing new subjects in a new cross-sectional study. Analyses of the resulting data can be cohort-sequential (cohort versus age), cross-sequential (cohort versus time of measurement), or time-sequential (age versus time of measurement).

3. In repeated measures designs the same subjects are tested at different times. Independent samples designs test new subjects from the same cohort, instead of retesting the same subjects at different times.

4. Many experimental designs (quasi-experiments) may yield important insights and allow researchers to examine alternative explanations for a variety of aging phenomena. Age-comparative experiments compare groups of subjects of different ages to assess their performance of some behavior in order to determine if a particular factor explains the age differences in behavior. Single age group intervention designs study a group on whom longitudinal data is available and compare the effects of intervention on stable individuals with the effects on individuals experiencing decline. Molar equivalence-molecular decomposition experiments seek to discover how the older person might compensate for declines in performance in one area by enhanced performance in another area.

5. In addition to concerns about reliability, validity, and objectivity, several issues bear on the validity of developmental research. Relationships between observables, such as test performance and behavior, and latent constructs, such as anxiety and intelligence, must be stable across time to provide valid representations of performance. The method of confirmatory factor analysis is used to test the equivalence of these relationships across time or different age groups. The generalizability of research findings must also be taken into account. The extent to which research
findings can be broadly applied depends greatly on the representativeness of the sample. It is also necessary to be aware of the demographics and health characteristics of subjects in order to understand how findings may be relevant to other populations.

SUGGESTED READINGS


Mertens, D. M. (1998). Research methods in education and psychology: Integrating diversity with quantitative and qualitative approaches. Thousand Oaks, CA: Sage. A very accessible presentation of research designs and elementary statistics suitable for students who have little or no background in measurement and social science statistics. This text also emphasizes issues in research with minority populations.