Jacoby, W. G. (1991). *Data theory and dimensional analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–078). Newbury Park, CA: Sage.
Nunnally, J. C. (1978). *Psychometric theory.* New York: McGraw-Hill.
Weisberg, H. F. (1974). Dimensionland: An excursion into spaces. *American Journal of Political Science, 18,* 743–776.

MULTI-ITEM MEASURES

The item is the basic unit of a psychological scale. It is a statement or question in clear, unequivocal terms about the measured characteristics (Haladyna, 1994). In social science, SCALES are used to assess people's social characteristics, such as attitudes, personalities, opinions, emotional states, personal needs, and description of their living environment.

An item is a "mini" measure that has a molecular score (Thorndike, 1967). When used in social science, multi-item measures can be superior to a single, straightforward question. There are two reasons. First, the RELIABILITY of a multi-item measure is higher than a single question. With a single question, people are less likely to give consistent answers over time. Many things can influence people's response (e.g., mood, specific thing they encountered that day). They may choose yes to a question one day and say no the other day. It is also possible that people give a wrong answer or interpret the question differently over time. On the other hand, a multi-item measure has several questions targeting the same social issue, and the final composite score is based on all questions. People are less likely to make the above mistakes to multiple items, and the composite score is more consistent over time. Thus, the multi-item measure is more reliable than a single question. Second, the VALIDITY of a multi-item measure can be higher than a single question. Many measured social characteristics are broad in scope and simply cannot be assessed with a single question. Multi-item measures will be necessary to cover more content of the measured characteristic and to fully and completely reflect the construct domain.

These issues are best illustrated with an example. To assess people's job satisfaction, a single-item measure could be as follows:

> I'm not satisfied with my work. (1 = *disagree*, 2 = *slightly disagree*, 3 = *uncertain*, 4 = *slightly agree*, 5 = *agree*)

To this single question, people's responses can be inconsistent over time. Depending on their mood or specific things they encountered at work that day, they might respond very differently to this single question. Also, people may make mistakes when reading or responding. For example, they might not notice the word *not* and agree when they really disagree. Thus, this single-item measure about job satisfaction can be notoriously unreliable. Another problem is that people's feelings toward their jobs may not be simple. Job satisfaction is a very broad issue, and it includes many aspects (e.g., satisfaction with the supervisor, satisfaction with coworkers, satisfaction with work content, satisfaction with pay, etc.). Subjects may like certain aspects of their jobs but not others. The single-item measure will oversimplify people's feelings toward their jobs.

A multi-item measure can reduce the above problems. The results from a multi-item measure should be more consistent over time. With multiple items, random errors could average out (Spector, 1992). That is, with 20 items, if a respondent makes an error on 1 item, the impact on the overall score is quite minimal. More important, a multi-item measure will allow subjects to describe their feelings about different aspects of their jobs. This will greatly improve the precision and validity of the measure. Therefore, multi-item measures are one of the most important and frequently used tools in social science.

—Cong Liu

See also SCALING

REFERENCES

Haladyna, T. M. (1994). *Developing and validating multiple-choice test items.* Hillsdale, NJ: Lawrence Erlbaum.
Spector, P. (1992). *Summated rating scale construction: An introduction.* Newbury Park, CA: Sage.
Thorndike, R. L. (1967). The analysis and selection of test items. In D. Jackson & S. Messick (Eds.), *Problems in human assessment* (pp. 201–216). New York: McGraw-Hill.

MULTILEVEL ANALYSIS

Most of statistical inference is based on replicated observations of UNITS OF ANALYSIS of one type (e.g., a sample of individuals, countries, or schools). The analysis of such observations usually is based on the assumption that either the sampled units themselves or the corresponding RESIDUALS in some statistical model are independent and identically distributed. However, the complexity of social reality and social science theories often calls for more COMPLEX DATA SETS, which include units of analysis of more than one type. Examples are studies on educational achievement, in which pupils, teachers, classrooms, and schools might all be important units of analysis; organizational studies, with employees, departments, and firms as units of analysis; cross-national comparative research, with individuals and countries (perhaps also regions) as units of analysis; studies in GENERALIZABILITY THEORY, in which each factor defines a type of unit of analysis; and META-ANALYSIS, in which the collected research studies, the research groups that produced them, and the subjects or respondents in these studies are units of analysis and sources of unexplained variation. Frequently, but by no means always, units of analysis of different types are hierarchically nested (e.g., pupils are nested in classrooms, which, in turn, are nested in schools). *Multilevel analysis* is a general term referring to statistical methods appropriate for the analysis of data sets comprising several types of unit of analysis. The *levels* in the multilevel analysis are another name for the different types of unit of analysis. Each level of analysis will correspond to a POPULATION, so that multilevel studies will refer to several populations—in the first example, there are four populations: of pupils, teachers, classrooms, and schools. In a strictly nested data structure, the most detailed level is called the first, or the lowest, level. For example, in a data set with pupils nested in classrooms nested in schools, the pupils constitute Level 1, the classrooms Level 2, and the schools Level 3.

HIERARCHICAL LINEAR MODEL

The most important methods of multilevel analysis are variants of REGRESSION analysis designed for hierarchically nested data sets. The main model is the HIERARCHICAL LINEAR MODEL (HLM), an extension of the GENERAL LINEAR MODEL in which the probability model for the errors, or residuals, has a structure reflecting the hierarchical structure of the data. For this reason, multilevel analysis often is called hierarchical linear modeling. As an example, suppose that a researcher is studying how annual earnings of college graduates well after graduation depend on academic achievement in college. Let us assume that the researcher collected data for a reasonable number of colleges—say, more than 30 colleges that can be regarded as a sample from a specific population of colleges, with this population being further specified to one or a few college programs—and, for each of these colleges, a random sample of the students who graduated 15 years ago. For each student, information was collected on the current income (variable Y) and the grade point average in college (denoted by the variable X in a metric where $X = 0$ is the minimum passing grade). Graduates are denoted by the letter i and colleges by j. Because graduates are nested in colleges, the numbering of graduates i may start from 1 for each college separately, and the variables are denoted by Y_{ij} and X_{ij}. The analysis could be based for college j on the model

$$Y_{ij} = a_j + b_j X_{ij} + E_{ij}.$$

This is just a linear regression model, in which the INTERCEPTS a_j and the regression coefficients b_j depend on the college and therefore are indicated with the subscript j. The fact that colleges are regarded as a random sample from a population is reflected by the assumption of random variation for the intercepts a_j and regression coefficients b_j. Denote the population mean (in the population of all colleges) of the intercepts by a and the college-specific deviations by U_{0j}, so that $a_j = a + U_{0j}$. Similarly, split the regression coefficients into the population mean and the college-specific deviations $b_j = b + U_{1j}$. Substitution of these equations then yields

$$Y_{ij} = a + bX_{ij} + U_{0j} + U_{1j}X_{ij} + E_{ij}.$$

This model has three different types of residuals: the so-called Level-1 residual E_{ij} and the Level-2 residuals U_{0j} and U_{1j}. The Level-1 residual varies over the population of graduates; the Level-2 residuals vary over the population of colleges. The residuals can be interpreted as follows. For colleges with a high value of U_{0j}, their graduates with the minimum passing grade $X = 0$ have a relatively high expected income—namely, $a + U_{0j}$. For colleges with a high value of U_{1j}, the effect of one unit GPA extra on the expected income of their graduates is relatively high—namely, $b + U_{1j}$. Graduates with a high value of E_{ij} have an income that is relatively high, given their college j and their GPA X_{ij}.

This equation is an example of the HLM; in its general form, this model can have more than one independent variable. The first part of the equation, $a + bX_{ij}$, is called the fixed part of the model; this is a linear function of the independent variables, just like in linear regression analysis. The second part, $U_{0j} + U_{1j}X_{ij} + E_{ij}$, is called the random part and is more complicated than the random residual in linear regression analysis, as it reflects the unexplained variation E_{ij} between the graduates as well as the unexplained variation $U_{0j} + U_{1j}X_{ij}$ between the colleges. The random part of the model is what distinguishes the hierarchical linear model from the general linear model. The simplest nontrivial specification for the random part of a two-level model is a model in which only the intercept varies between Level-2 units, but the regression coefficients are the same across Level-2 units. This is called the random intercept model, and for our example, it reads

$$Y_{ij} = a + bX_{ij} + U_{0j} + E_{ij}.$$

Models in which also the regression coefficients vary randomly between Level –2 units are called random slope models (referring to graphs of the regression lines, in which the regression coefficients are the slopes of the regression lines).

The dependent variable Y in the HLM always is a variable defined at the lowest (i.e., most detailed) level of the hierarchy. An important feature of the HLM is that the independent, or explanatory, variables can be defined at any of the levels of analysis. In the example of the study of income of college graduates, suppose that the researcher is interested in the effect on earnings of alumni of college quality, as measured by college rankings, and that some meaningful college ranking score Z_j is available. In the earlier model, the college-level residuals U_{0j} and U_{1j} reflect unexplained variability between colleges. This variability could be explained partially by the college-level variable Z_j, according to the equations

$$a_j = a + c_0Z_j + U_{0j}, \quad b_j = b + c_1Z_j + U_{1j},$$

which can be regarded as linear regression equations at Level 2 for the quantities a_j and b_j, which are themselves not directly observable. Substitution of these equations into the Level-1 equation $Y_{ij} = a_j + b_jX_{ij} + E_{ij}$ yields the new model,

$$Y_{ij} = a + bX_{ij} + c_0Z_j + c_1X_{ij}Z_j + U_{0j} + U_{1j}X_{ij} + E_{ij},$$

where the parameters a and b and the residuals E_{ij}, U_{0j}, and U_{1j} now have different meanings than in the earlier model. The fixed part of this model is extended compared to the earlier model, but the random part has retained the same structure. The term $c_1X_{ij}Z_j$ in the fixed part is the INTERACTION EFFECT between the Level-1 variable X and the Level-2 variable Z. The regression coefficient c_1 expresses how much the college context (Z) modifies the effect of the individual achievement (X) on later income (Y); such an effect is called a cross-level interaction effect. The possibility of expressing how context (the "macro level") affects relations between individual-level variables (the "micro level") is an important reason for the popularity of multilevel modeling (see DiPrete & Forristal, 1994).

A parameter that describes the relative importance of the two levels in such a data set is the INTRACLASS CORRELATION coefficient, described in the entry with this name and also in the entry on VARIANCE COMPONENT MODELS. The similar variance ratio, when applied to residual (i.e., unexplained) variances, is called the residual intraclass correlation coefficient.

ASSUMPTIONS, ESTIMATION, AND TESTING

The standard assumptions for the HLM are the linear model expressed by the model equation, normal distributions for all residuals, and independence of the residuals for different levels and for different units in the same level. However, different residuals for the same unit, such as the random intercept U_{0j} and the random slope U_{1j} in the model above, are allowed to be correlated; they are assumed to have a multivariate normal distribution. With these assumptions, the HLM for the example above implies that outcomes for graduates of the same college are correlated due to the influences from the college—technically, due to the fact that their equations for Y_{ij} contain the same college-level residuals U_{0j} and U_{1j}. This dependence between different cases is an important departure from the assumptions of the more traditional general linear model used in regression analysis.

The parameters of the HLM can be estimated by the MAXIMUM LIKELIHOOD method. Various ALGORITHMS have been developed mainly in the 1980s (cf. Goldstein, 2003; Longford, 1993); one important algorithm is an iterative reweighted least squares algorithm (see the entry on GENERALIZED LEAST SQUARES), which alternates between estimating the regression coefficients in the fixed part and the parameters of the random part. The regression coefficients can be tested by T-TESTS or Wald tests. The parameters defining the structure of the random part can be tested by LIKELIHOOD RATIO TESTS (also called deviance tests) or by chi-squared tests. These methods have been made available since the 1980s in dedicated multilevel software, such as HLM and MLwiN, and later also in packages that include multilevel analysis among a more general array of methods, such as M-Plus, and in some general statistical packages, such as SAS and SPSS. An overview of software capabilities is given in Goldstein (2003).

MULTIPLE LEVELS

As was illustrated already in the examples, it is not uncommon that a practical investigation involves more than two levels of analysis. In educational research, the largest contributions to achievement outcomes usually are determined by the pupil and the teacher, but the social context provided by the group of pupils in the classroom and the organizational context provided by the school, as well as the social context defined by the neighborhood, may also have important influences. In a study of academic achievement of pupils, variables defined at each of these levels of analysis could be included as explanatory variables. If there is an influence of some level of analysis, then it is to be expected that this influence will not be completely captured by the variables measured for this level of analysis, but there will be some amount of unexplained variation between the units of analysis for this level. This should then be reflected by including this unexplained variation as random residual variability in the model. The first type of residual variability is the random main effect of the units at this level, exemplified by the random intercepts U_{0j} in the two-level model above. In addition, it is possible that the effects of numerical variables (such as pupil-level variables) differ across the units of the level under consideration, which can again be modeled by random slopes such as the U_{1j} above. An important type of conclusion of analyses with multiple levels of analysis is the partitioning of unexplained variability over the various levels. This is discussed for models without random slopes in the entry on VARIANCE COMPONENT MODELS. How much unexplained variability is associated with each of the levels can provide the researcher with important directions about where to look for further explanation.

Levels of analysis can be nested or crossed. One level of analysis—the lower level—is said to be nested in another, higher level if the units of the higher level correspond to a partition into subsets of the units of the lower level (i.e., each unit of the lower level is contained in exactly one unit of the higher level). Otherwise, the levels are said to be crossed. Crossed levels of analysis often are more liable to lead to difficulties in the analysis than nested levels: Estimation algorithms may have more convergence problems, the empirical conclusions about partitioning variability over the various levels may be less clear-cut, and there may be more ambiguity in conceptual and theoretical modeling.

The use of models with multiple levels of analysis requires a sufficiently rich data set on which to base the statistical analysis. Note that for each level of analysis, the units in the data set constitute a sample from the corresponding population. Although any rule of thumb should be taken with a grain of sand, a sample size less than 20 (i.e., a level of analysis represented by less than 20 units) usually will give only quite restricted information about this population (i.e., this level of analysis), and sample sizes less than 10 should be regarded with suspicion.

LONGITUDINAL DATA

In LONGITUDINAL RESEARCH, the HLM also can be used fruitfully. In the most simple longitudinal data structure, with repeated measures on individuals, the repeated measures constitute the lower (first) and the individuals the higher (second) levels. Mostly, there will be a meaningful numerical time variable: For example, in an experimental study, this may be the time since onset of the experimental situation, and in a developmental study, this may be age. Especially for nonbalanced longitudinal data structures, in which the numbers and times of observations differ between individuals, multilevel modeling may be a natural and very convenient method. The dependence of the outcome variable on the time dimension is a crucial aspect of the model. Often, a linear dependence is a useful first approximation. This amounts to including the time of measurement as an explanatory variable; a random slope for this variable represents differential change (or growth) rates for different individuals. Often, however, dependence on time is nonlinear. In some cases, it will be possible to model this while remaining within the HLM by using several nonlinear transformations (e.g., polynomials or splines) of time and postulating a model that is linear in these transformed time variables (see Snijders & Bosker, 1999, chap. 12). In other cases, it is better to forgo the relative simplicity of linear modeling and construct models that are not linear in the original or transformed variables or for which the Level-1 residuals are autocorrelated (cf. Verbeke & Molenberghs, 2000).

NONLINEAR MODELS

The assumption of normal distributions for the residuals is not always appropriate, although sometimes this assumption can be made more realistic by transformations of the dependent variable. In particular, for dichotomous or discrete dependent variables, other models are required. Just as the GENERALIZED LINEAR MODEL is an extension of the general linear model of regression analysis, nonlinear versions of the HLM also provide the basis of, for example, multilevel versions of LOGISTIC REGRESSION and LOGIT MODELS. These are called hierarchical generalized linear models or generalized linear mixed models (see the entry on HIERARCHICAL NONLINEAR MODELS).

—Tom A. B. Snijders

REFERENCES

de Leeuw, J., & Kreft, I. G. G. (Eds.). (in press). *Handbook of quantitative multilevel analysis.* Dordrecht, The Netherlands: Kluwer Academic.

DiPrete, T. A., & Forristal, J. D. (1994). Multilevel models: Methods and substance. *Annual Review of Sociology, 20,* 331–357.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Hodder Arnold.

Hox, J. J. (2002). *Multilevel analysis, techniques and applications.* Mahwah, NJ: Lawrence Erlbaum.

Longford, N. T. (1993). *Random coefficient models.* New York: Oxford University Press.

Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer.

MULTIMETHOD-MULTITRAIT RESEARCH. *See* MULTIMETHOD RESEARCH

MULTIMETHOD RESEARCH

Multimethod research entails the application of two or more sources of data or research methods to the investigation of a research question or to different but highly linked research questions. Such research is also frequently referred to as *mixed methodology.* The rationale for mixed-method research is that most social research is based on findings deriving from a single research method and, as such, is vulnerable to the accusation that any findings deriving from such a study may lead to incorrect inferences and conclusions if MEASUREMENT ERROR is affecting those findings. It is rarely possible to estimate how much measurement error is having an impact on a set of findings, so that monomethod research is always suspect in this regard.

MIXED-METHOD RESEARCH AND MEASUREMENT

The rationale of mixed-method research is underpinned by the principle of TRIANGULATION, which implies that researchers should seek to ensure that they are not overreliant on a single research method and should instead employ more than one measurement procedure when investigating a research problem. Thus, the argument for mixed-method research, which in large part accounts for its growth in popularity, is that it enhances confidence in findings.

In the context of measurement considerations, mixed-method research might be envisioned in relation to different kinds of situations. One form might be that when one or more constructs that are the focus of an investigation have attracted different measurement efforts (such as different ways of measuring levels of job satisfaction), two or more approaches to measurement might be employed in combination. A second form might entail employing two or more methods of data collection. For example, in developing an approach to the examination of the nature of jobs in a firm, we might employ structured observation and structured interviews concerning apparently identical