MEASURES OF FAMILY CHARACTERISTICS

Murray A. Straus
Susan M. Ross

A 1964 review of tests and scales used in family research found serious deficiencies (Straus 1964), and subsequent reviews showed very little improvement (Straus 1992; Straus and Brown 1978). However, changes in the nature of the field have contributed to an increase in the use of standardized tests to measure characteristics of the family. This is an important development because standardized tests are vital tools for both clinical assessment and research. New tests tend to produce a flowering of research focused on the newly measurable concept. Examples of tests that have fostered much research include measures of marital satisfaction (Spanier 1976), adequacy of family functioning (Olson, Russell, and Sprenkle 1989), and family violence (Straus 1990a). Hundreds of family measures are abstracted or reproduced in compendiums such as Family Assessment (Grotevant and Carlson 1989), Handbook of Measurements for Marriage and Family Therapy (Fredman and Sherman 1987), and Handbook of Family Measurement Techniques (Touliatos, Perlmuter, and Straus 1990). There is also a growing methodological literature on techniques for constructing measures of family characteristics, such as those by Karen S. Wampler and Charles F. Halverson, Jr. (1993) and Thomas W. Draper and Anastascios C. Marcos (1990). The state of testing in family research, however, is not as healthy as these publications might suggest. In fact, the data indicate that the validity of tests used in family research is rarely known.

For purposes of this entry, the term “measure” includes test, scale (such as Likert, Thurstone, Guttman, and Semantic Differential scales), index, factor score, scoring system (when referring to methods of scoring social interaction such as Gottman 1994 or Patterson 1982), and latent variables constructed by use of a structural equation modeling program. The defining feature is that they “combine the values of several items [also called indicators, questions, observations, events] into a composite measure . . . used to predict or gauge some underlying continuum which can only be partially measured by any single item or variable” (Nie et al. 1978, p. 529).

Advantages of Multiple-Item Measures

Multiple-item measures are emphasized in this entry because they are more likely to be valid than single-item measures. Although one good question or observation may be enough and thirty bad ones are useless, there are reasons why multiple-item measures are more likely to be valid. One reason is that most phenomena of interest to family researchers have multiple facets that can be adequately represented only by use of multiple items. A single question, for example, is unlikely to represent the multiple facets of marital satisfaction adequately.

A second reason for greater confidence in multiple-item measures occurs because of the inevitable risk of error in selecting items. If a single item is used and there is a conceptual error in formulating or scoring it, hypotheses that are tested by using that measure will not be supported even if they are true. However, when a multiple-item test is used, the adverse effect of a single invalid item is limited to a relatively small reduction in validity (Sträus and Baron 1980). In a fifteen-item scale, for example, a defective item is only 6.6 percent of the total, so the findings would parallel those obtained if all fifteen items were correct.

Multiple items are also desirable because measures of internal consistency reliability are based on the number of items in the measure and the correlation between them. Given a certain average correlation between items, the more items, the higher the reliability. If only three items are used, it is rarely possible to achieve a high level of reliability. Reliability needs to be high because it sets an upper limit on validity.

Status and Trends in Family Measurement

To investigate the quality of measurement in family research, all empirical studies published in two major U.S. family journals (Journal of Marriage and the Family and Journal of Family Psychology) were examined. To determine trends in the Journal of Marriage and the Family, issues from 1982 and 1992 were compared. For the Journal of Family Psychology, issues from 1987 (the year the journal was founded) and 1992 were compared. Of the 161 empirical research articles reviewed, slightly fewer than two-thirds used a multiple item measurement. This increased from 46.9 percent initially to 68.1 percent in 1992. A typical article used more than one such instrument, so that a total of 219 multiple item measures were used in these 161 articles. Reliability was reported in 79.4 percent of these articles. Reliability reporting increased from 53.3 percent initially to 90.6 percent in 1992. Six percent of the articles had as their main purpose describing a new measurement instrument or presenting data concerning an existing instrument.
How one interprets these statistics depends on the standard of comparison. Articles in sociology journals and child psychology and clinical psychology journals are appropriate comparisons because these are the disciplines closest to family studies and in which many family researchers were trained. For sociology, the findings listed above can be compared to those reported in a study by Murray A. Straus and Barbara Wauchople (1992), in which they examined empirical articles from the 1979 and 1989 issues of American Sociological Review, American Journal of Sociology, and Sociological Methods and Research. This comparison shows that articles in family journals pay considerably more attention to measurement than articles in leading sociological journals. None of the 185 articles in sociology journals was on a specific measure, whereas 8 percent of the articles in the family journals were devoted to describing or evaluating an instrument. This portends well for family research because it is an investment in tools for future research. Only one-third of the articles in the sociology journals used a multiple-item measure, compared to more than two-thirds (68%) of articles in the family journals. The record of family researchers also exceeds that of sociologists in respect to reporting reliability. Only about 10 percent of the articles in sociology journals, compared to 80 percent of the articles in family journals, reported the reliability of the instruments. The main problem area is validity; only 12.4 percent of the articles in family journals described or referenced evidence of validity. The fact that this is three times more than in sociology is not much consolation because 12 percent is still a small percentage. Moreover, reporting or citing information on validity did not increase from the base period. Since validity is probably the most crucial quality of an instrument, the low percentage and the lack of growth indicate that more attention needs to be paid to measurement in family research.

There is no comparable study of measures in child or clinical psychological journals.

Reasons for Underdevelopment of Measures

The limited production of standard and validated measures of family characteristics is probably the result of a number of causes. Conventional wisdom attributes it to a lack of time and other resources for instrument development and validation. This is not an adequate explanation because it is true of all the social sciences. Why do psychologists devote the most resources to developing and validating tests, sociologists the least, and family researchers fall in between?

One likely reason is a difference in rewards for measurement research. A related reason is a difference in the opportunities and constraints. In psychology, there are journals devoted to psychological measures in whole or in part, such as Educational and Psychological Measurement and Journal of Clinical and Consulting Psychology. There are no such journals in sociology or family studies. Moreover, there is a large market for psychological tests, and several major firms specialize in publishing tests. It is a multimillion-dollar industry, and authors of tests can earn substantial royalties. By contrast, sociology lacks the symbolic and economic reward system that underlies the institutionalization of test development as a major specialization in psychology. The field of family studies lies in between. In principle there should be a demand for tests because of the large number of family therapists, but few family therapists actually use tests.

A second explanation for the differences among psychology, family studies, and sociology in attention to measurement is a situational constraint inherent in the type of research done. A considerable amount of family research is done by survey methods—for example, the National Survey of Families and Households. Surveys of this type usually include measures of many variables in a single thirty- to sixty-minute interview. Clinical psychologists, on the other hand, often can use longer and therefore more reliable tests, because their clients have a greater stake in providing adequate data and will tolerate undergoing two or more hours of testing.

Third, most tests are developed for a specific study and there is rarely a place in the project budget for adequate measure development—test/retest reliability, concurrent and construct validity, and construction of normative tables. Even when the author of a measure does the psychometric research needed to enable others to evaluate whether the measure might be suitable for their research, family journals rarely allow enough space to present that material.

Fourth, the optimum procedure is for the author to write a paper describing the test, the theory underlying the test, the empirical procedures used to develop the test, reliability and validity evidence, and norms. This rarely occurs because of the lack of resources indicated above. In addition, most investigators are more interested in the substantive issues for which the project was funded.

Another reason why standardized tests are less frequently used in family research is that many studies are based on cases from agencies. A researcher studying child abuse who draws the cases from child protective services might not need a method of measuring child abuse. However, standardized tests are still needed because an adequate understanding of child abuse cannot depend solely on officially identified cases. It is important also to do research on cases that are not known to agencies, because such cases are much more numerous than cases known to agencies (Straus 1990b).
The Future of Family Research Measures

There are grounds for optimism and grounds for concern about the future of family tests. The grounds for concern are, first, that in survey research on the family, concepts are often measured by a single interview question. Second, even when a multiple-item test is used, it is rarely on the basis of empirical evidence of reliability and validity. Third, the typical measure developed for use in a family study is never used in another study. One can speculate that this hiatus in the cumulative nature of research occurs because of the lack of evidence of reliability and validity and because authors rarely provide sufficient information to facilitate use of the instrument by others.

The grounds for optimism are to be found in the sizable and slowly growing number of standardized instruments, as listed in compendiums (e.g., Grotevant and Carlson 1989; Fredman and Sherman 1987; Touliatas, Perlmuttern, and Strauss 1990). A second ground for optimism is the rapid growth in the number of psychologists doing family research, because psychologists bring to family research an established tradition of test development. Similarly, the explosive growth of family therapy is grounds for optimism, because it is likely that more tests will gradually begin to be used for intake diagnosis.

There is a certain irony in the second source of optimism, because basic researchers usually believe that they, not clinicians, represent quality in science. In respect to measurement, clinicians tend to demand instruments of higher quality than do basic researchers because the consequences of using an inadequate measure are more serious. When a basic researcher uses an instrument with low reliability or validity, it can lead to a Type II error—that is, failing to accept a true hypothesis. This may result in theoretical confusion or a paper not being published. But when a practitioner uses an invalid or unreliable instrument, the worst-case scenario can involve injury to a client. Consequently, clinicians need to demand more evidence of reliability and validity than do researchers. As a result, clinically oriented family researchers tend to produce and make available more adequate measures. Hubert M. Blalock (1982) argued that inconsistent findings and failure to find empirical support for sound theories may be due to lack of reliable and valid measures of operationalizing concepts in the theories being tested. It follows that research will be on a sounder footing if researchers devote more attention to developing reliable and valid measures of family characteristics.

(See also: Marital Quality; Marital Typologies; Research Methods)

BIBLIOGRAPHY


Murray A. Straus

Saban M. Ross