1995

1. Measurement Beyond the Individual

Charles F. Halverson
University of Georgia, chalvers@fcs.uga.edu

Follow this and additional works at: http://digitalcommons.unl.edu/burosfamily
Part of the Educational Assessment, Evaluation, and Research Commons, Family, Life Course, and Society Commons, and the Quantitative, Qualitative, Comparative, and Historical Methodologies Commons

http://digitalcommons.unl.edu/burosfamily/4

This Article is brought to you for free and open access by the Buros-Nebraska Series on Measurement and Testing at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Family Assessment by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
This chapter has several goals. First, I will briefly review the history of measurement as it applies to family assessment. This history has been recounted by many and is available in many recent publications, so I shall be fairly brief. Second, I will discuss family measurement in terms of important issues still facing the family measurement field—issues that are not, in my opinion, being well addressed at this time. And finally, I will attempt to weave these various threads into some speculations about the future directions that family measurement might (or maybe needs) to take.

I will confine this discussion to quantitative measurement because the available space does not permit any extensive discussion of the growing area of qualitative research. I will confess that although my biases permit qualitative research to be complementary to quantitative research, I believe we will always find it necessary to use quantitative measures when we entertain questions involving comparisons across

---

Author Notes: Portions of this chapter are based on material contained in a chapter by Karen Wampler and Charles Halverson that appeared in The Sourcebook of Family Theories and Methods: A Contextual Approach (1993). Edited by P. Boss, W. Doherty, R. LaRossa, W. Schuum & S. Steinmetz (Chapter 8).

Thanks go to several anonymous reviewers who considerably strengthened the chapter. The work was supported in part by Grant MH39899 awarded to C. F. Halverson.
families or when we desire generalizations to populations with certain defined characteristics.

Let me be clear about what is meant by quantitative research methods. Quantitative measurement is simply the assigning of numerical values to abstract, theoretical constructs that constitute the core of family theory. Further, my emphasis is not on measures pertaining to individuals but rather on family measures—when a family relationship is measured or a set of relationships in the family are measured (See Draper & Marcos, 1990; Huston & Robins, 1982; and Thompson & Walker, 1982, among others for discussions on individual- vs. family-level measurement). Note that the distinctions made about family versus individual measurement are independent from types of measurement (e.g., observations, self-report, diaries) as well as data analytic techniques (e.g., combining scores from different family members, using dyadic codes, etc.).

When discussing the measurement of relationships we can for our purposes summarize a definition of relationship as clearly involving more than one individual over time (Wampler & Halverson, 1993). When considering the conceptualization of family relationships, we must also take into account the idea that families have a specifiable past and an expected future—that is, the relationships are intimate and extended as opposed to casual and brief.

Before I discuss history and issues, I will mention just one brief aside about family theory and its relation to measurement of family constructs. For most family researchers in the past, there has been a conscious attempt to link broad, theoretical positions to certain types of measurement. For example, symbolic interactionists have demonstrated a strong commitment to qualitative, grounded methods, whereas behaviorists have verified their commitment to observational data by focusing on behaviors, right down to microcoding small behavioral sequences in family interactions.

My position is that measurement cannot be formulated without theory—such a theory is not a global theory, however, but rather a set of theories about constructs that will dictate what measures we should collect to identify each abstract construct. This theory about measurement stems from a "multiplist position" (Houts, Cook, & Shadish, 1986) that advances a measurement pluralism where every abstract social science construct is best measured from multiple perspectives—no one measurement system (self-report, observational, short-term, long-term, etc.) is adequate to measure any complex construct. Every construct's meaning is more than that indicated by any one measure, source, setting, etc. I will return to this notion of
"construct building" when we discuss some of the shortcomings of current-day family assessment. (See also Bank, Dishion, Skinner, & Patterson, 1990; Patterson & Bank, 1989.)

HISTORY

In our recent chapter on quantitative measurement we proposed that the history of family measurement coincides with the history of the scientific study of the family (Wampler & Halverson, 1993). It has culminated with the multiagent and multimethod strategies that are increasingly being used today.

MEASUREMENT OF MARRIAGE VARIABLES

The earliest attempts at measurement of marriage were almost entirely based on self-reports and focused on indicators of either satisfaction or marital adjustment (Burgess & Cottrell, 1939; Locke, 1951; Locke & Williamson, 1958). Revisions of early measures by Spanier (1976), along with Gottman’s (1979) research indicating the centrality of marital adjustment and satisfaction for marital functioning have led to many measures of the core evaluative constructs of marriage. For the most part, they are self-report and individually-based measures that are widely used today in nearly every study of marriage.

In the area of marriage assessment, there has been what could be termed a growing methodological dualism with the rise of observational studies of marital interaction in the late 60s and 70s, continuing to the present. Beginning in the 60s at the Old Child Research Branch (where I had a new post-doctoral assignment), Harold Rausch and his colleagues began the study of filmed marital improvisations (Rausch, Barry, Hertel, & Swain, 1974). Later, his colleagues Bob Ryder and Dave Olson began to use adaptations of the old Revealed Difference technique along with such innovative assessments of interaction as the color-matching test (e.g., Olson & Ryder, 1970).

Latter-day clinical psychologists have continued to refine the observational armamentarium over recent years (e.g., the Couples Interaction Scoring System [CISS], Gottman, 1979; the Spouse Observation Checklist [SOC], Weiss & Perry, 1983; and the Marital Interaction Coding System [MICS], Weiss, Hops, & Patterson, 1973). Ironically, these two traditions in the marriage assessment area are still relatively non-cross-fertilizing, existing in parallel tracks; both viewed as valid and sufficient in themselves. With the rise of video and high-tech coding procedures based on video records, we seem to have a proliferation of custom-coding systems, designed for special
uses. As we shall see, this parallel system and the proliferation of measures has not clarified the marriage measurement field.

MEASUREMENT OF FAMILY VARIABLES

Measurement of family variables has had a somewhat different history, with observational coding systems arising early from the small-group work of Bales (1950) and others. These early studies focused for the most part on the verbal interactions among parents and their children. Many of these studies, like some of the marriage observational studies, were laboratory based and used various techniques to elicit interaction (e.g., Revealed Differences, card sorts; Reiss & Klein, 1987; SIMFAM, Straus & Tallman, 1971; Building Houses, Halverson & Wampler, 1993).

Although self-report measures of family functioning have been around for many years, self-report measures of the family really did not become widely available until David Olson and his colleagues began to create and make available a wide variety of questionnaire scales, most notably the Family Adaptability and Cohesion Evaluation Scales (FACES) with its associated circumplex model of family functioning (e.g., Olson, McCubbin, Barnes, Larsen, Muxen, & Wilson, 1985). This instrument, more than any other, has become the benchmark for family assessment. Olson has reported over 600 studies using one of the versions of FACES; the researchers continue developing its norms based upon different family forms, ethnic groups, and studies on cross-national differences in adaptability and cohesion. Clearly, FACES is a popular instrument. The only other self-report instrument that approaches it in popularity is the Moos Family Environment Scale (FES; Moos & Moos, 1984).

MEASUREMENT OF PARENT-CHILD VARIABLES

In the area of parenting and parent-child relationships, self-report instruments of parenting practices and attitudes like Block’s Child-Rearing Practices Report (CRPR, Block, Block, & Morrison, 1981) have been developed along with observational measures of dyadic interaction in the family—observation of parent-child and sibling interaction supplementing the observations of husbands and wives, (see, for example, the Family Interaction Coding System, Patterson, 1982; the Family Interaction Q-Sort, Gjerde, 1986). These observational measures allowed investigators to conceptualize parent-child relationships as bi-directional (Bell, 1968; Sears, Maccoby, & Levin, 1957). The use of both self-report parenting measures and observations
of parent-child interaction allowed investigators to realize that parental report is often inconsistent with self-reported attitudes. (See Bradbury & Fincham, 1990, for an extended discussion of this issue.)

Sibling and intergenerational relationships have been increasingly measured in recent years (see Bengtson, 1989; Brody & Stoneman, 1990). Interestingly, the sibling measures are self-report, interview, and observational whereas nearly all measures of intergenerational relationships are one-respondent self-reports. Below, I have quoted a passage from our recent chapter on quantitative family measurement that directs the interested reader to one or more of the many reviews of quantitative family measurement (Wampler & Halverson, 1993).

The reader is referred to the following publications starting with the most comprehensive: Touliatos, Perlmutter, and Straus (1990), Jacob and Tennenbaum (1988); Fredman and Sherman (1987); Grotevant and Carlson (1989; family and parent-child); Filsinger (1983b; family and marital); O'Leary (1987; marital); Skinner (1987; family self-report); Forman and Hagan (1983; 1984; family self-report); Sabatelli (1988; marital self-report); Filsinger and Lewis (1981; marital observation); Gilbert and Christensen (1985; marital observation); Markman and Notarius (1987; marital and family observation); Margolin (1987; behavioral self-report); Beere (1990; gender roles); and Mangen, Bengtson, and Landry (1988; intergenerational self-report). Schumm (1990) provides a summary of the major reviews and compendia of marriage and family measures. (pp. 184-185)

The history of quantitative family measurement reflects the influences of many social science subdisciplines with their varying conceptual and methodological preferences influencing how the family measurement enterprise has been conducted over the years. (Bradbury & Fincham, 1990; Gottman, 1979; Grotevant & Carlson, 1989; Jacob, 1987). Some of the subdisciplines relied on survey and interview methods whereas the more hard-nosed behaviorists developed rigorous observational protocols to study relationships. Only in the most recent years have we seen the use of multiple measurement strategies to identify family-level constructs and to assess to some extent the biases of mono-method approaches. More on this below.

The previous sections have been a quick tour of the complex and burgeoning family assessment area. In one sense, there certainly appears to be much vitality to the enterprise—many studies, many measures, lots of publications—but how well are we doing? Not as well as the mini-history might indicate. Let us look at some of the problems in this area as I see them (and as seen by others as well!).
PROBLEMS IN FAMILY ASSESSMENT

When you examine the literature the first thing that strikes you is that much of family measurement research is still rather small-scale with investigators working in relative isolation from each other. Part of this is no surprise—the "engine" of family research is the graduate thesis or dissertation, done by people with limited means. This research is often never replicated nor the measure used again by other people. Generations of this kind of research has led to the greatest weakness in family assessment: There are too many measures measuring too many constructs. Any review of family measurement (e.g., Touliatos, Perlmutter, & Straus, 1990) will quickly reveal there are hundreds of family measures, most with limited reliability and barely adequate psychometric properties. Even the few measures that might possess decent psychometrics have been used in just a study or two. Indeed, Schumm (1990) cites the research of Straus (1969) that 80% (!) of surveyed measures had never been used more than once. Schumm also cites Bonjean, Hill, and McLemore (1967) who report equally dismal findings: 28% of the measures had been used more than once and only 2.2% had been used as many as five times! Coupled with the above is the fact that many investigators seemed prone to develop new measures when they needed one to measure their favorite construct, or worse, adopt ones with unknown psychometric properties.

Closely allied to the problem of too many measures (and really a result of it) is the problem of too many constructs being assessed by all these family instruments. This problem of too many constructs reflects the fact that there is really no consensus on what are the most important constructs in assessing family relationships. If you look carefully at most constructs defined by the various assessment devices, it becomes apparent that many constructs with the same name may not be measuring the same underlying variable and there is always the possibility that constructs with very different labels may be capturing the same underlying variance (what my colleague Jack Block [personal communication, June, 1991] refers to as the "jingle-jangle" problem).

With the multitude of measures partially identifying many, many constructs and very little in the way of replicated findings, it is really quite impossible for most family researchers to identify potentially useful measures of family functioning. All these measures of unknown validity and reliability leads to a serious dilution of research efforts. Instead of systematic research on a small number of constructs identified by a manageable number of measures, we have instead
example after example of one or two studies that identify a construct with only one measure and then little or no follow-up, or replication.

In an earlier chapter the utility of the theory of critical multiplism was noted (Wampler & Halverson, 1993):

Recall the perspective of critical multiplism (Houts et al., 1986) cited earlier. A multiplist perspective asserts that no one measurement system is adequate to exhaust the meaning of any complex social science construct. In the family area we mostly deal with highly complex, abstract, “nonvisible” constructs that must be estimated from fallible and biased measurement systems. The bias is maximized when our constructs are estimated by one measure from one source with one method at one point in time (e.g., self-report questionnaire from wife on family cohesion). This typical case must be remedied by “building constructs” across methods, sources, different times, and contexts if we intend to have constructs general across such domains. The point is to “average out” the limitations and biases from any one single source and method and to aggregate the underlying construct variance across sources, methods, contexts, and time for a stable, well-defined construct that is not tied to any one source setting or method. We must devote both theory and empirical work to aggregation and construct building (cf., Patterson & Bank, 1989) that can include all sorts of measurement at all levels of quantitative sophistication from nominal data to ratio scales. (p. 189)

Let me give you a brief example from our own recent work. In her dissertation, Nancy Hollett (1992) discovered the value of aggregation of measures over time and source in predicting some peer-acceptance outcomes some 3 years after we had stopped collecting family data. Originally in our modeling of predictors of peer acceptance (measured in the classroom), we used data from the 4th-year observation of the parent-child interaction in our lab to predict peer acceptance. The predictions derived from these observational Q-sort ratings (of about 20 minutes of interaction) showed little convergence with other measures of parenting (self-report and interview ratings) and no predictability to our criterion measure. We could have at that point concluded that there was no predictability from our family data to peer data. We decided to instead aggregate the three sources of data over 4 years of observations, self-reports, and interview ratings to see if we could construct a more robust and reliable measure of competent parent-child interaction. Consistent with the lessons we have learned from our personality-researcher colleagues (see Epstein & O'Brien, 1985), such aggregate measures proved much more adequate than any single measure. Each measure contained theoretically relevant
components that were not in the other measures such that aggregated 4-year observation Q-sorts became more reliable and converged with the other measures to form a latent construct with path coefficients of .40 and .47 to peer acceptance (for mother and father parenting style respectively). Obviously, the aggregate measures must show convergent validity in order to be useful in a prediction equation. In this case we built a construct with both relatively molar and molecular variables that came from three different sources that converged on a construct of competent parenting having real predictive potential.

This example leads me to yet another weakness in family assessment—the lack of studies where one can compare the usefulness and distinctiveness of various constructs included in the same study. In terms of the multiplist agenda, we have almost no work done on the nomological net of our constructs (Campbell & Fiske, 1959). In Campbell and Fiske’s article on the multitrait-multimethod matrix, construct validity could be demonstrated when two or more methods were used to measure two or more traits in a nomological net. Multitrait-multimethod studies allow us to examine construct validity as well as to distinguish truly different traits from those with overlapping variance. Studies where this is possible are mostly missing from family research. It is difficult to find a study that includes three or four operationalizations (even if all self-report!) of some key family construct. This problem is especially serious in family data where most comparisons are within method (e.g., method variance is almost always confounded with construct variance). Indeed, the lack of convergence of measures from different or same sources is always ambiguous. We seem to be swimming in a sea of measures of unknown meaning most of the time!

One might think the solution to this problem could be solved by getting large Ns and using a potful of family measures to see “what is related to what.” I believe this strategy is a mistake. Along with Jacob Cohen (1990) I think less is more. Cohen convincingly demonstrates the folly of studies with “prodigious numbers of dependent variables...[and] far too many independent variables, or (heaven help us) both” (p. 1304). There is considerable muddle in relating, for example, 10 predictors to 6 outcome measures. A little thought reveals that the Type I error rate is very high—there are going to be many “significant” chance correlations (more with larger Ns) and we really cannot tell which are the real associations.

Related to the weakness above is the unfortunate fact that many or most of the measures we have of the family are self-report. These instruments elicit information from individuals who report on their family’s functioning. Almost without exception, instruments are
developed, normed, and used as *individual* measures rather than as family measures—even when the content refers to families. Obviously, the reliance on single-source, single-method data as a proxy for family functioning produces many difficulties that cannot be solved by factoring a large number of instruments together (assuming we could get a huge sample to fill out 35 or 40 of these measures). Such studies might find a few, broad replicable "factors" (probably evaluation—good family—bad family). Reports of internal consistency, reliability, etc., could not erase the problem of method and source variance in the measures: These new clusters will always refer to single individual’s perceptions of family, not to descriptions of the family based on multiple *sources, settings, and times*!

Another way to phrase this issue is by asking the question: Are family measures capturing unique variance about family relationships or are they just individual characteristics disguised as family measures? (Wampler & Halverson, 1993). The issue is most salient for self-report measures because they measure individual perceptions of relationships rather than actual quality of relationships (Christensen & Arrington, 1987). In our data, correlations are consistently high between self-report measures of family constructs such as cohesion and individual constructs such as depression, clearly an individual measure. Further, for many analyses, family-level constructs do not add significant variance in a step-wise multiple regression after we have first entered individual measures. In an earlier publication (Wampler & Halverson, 1993) we wrote:

The individual difference issue is closely tied to the treatment of two or more different sources of information about a relationship (Fisher, Kokes, Ransom, Phillips, & Rudd, 1985; Schumm, Barnes, Bollman, Jurich, & Milliken, 1985). These discussions are often, however, in the context of how to combine scores rather than conceptualizing how different perceptions may be central to family process, a possibly far-reaching conceptual issue. If measures are simply individual scores, most information could be gained by leaving them separate. In contrast, if they are biased indicators of a construct, they should be combined. The point is that although we may combine individual perceptions of the family (and we have several proposed ways of doing it—see Schumm et al., 1985; Walters, Pittman, & Norrell, 1984), these combined scores then must be thought of as biased indicators of constructs to be combined with other methods and sources—and that is almost never done! (p. 187)

**SAMPLING AND FAMILY MEASUREMENT**

Finally, no discussion about the family measurement field could be complete without some discussion of the issues revolving around
sampling. First, let me note that for the most part, family assessment measures have been developed on relatively small and restricted samples. One lesson we have not heeded from our psychometric mentors is that without large sample sizes, much of our data are unstable and therefore mostly uninterpretable. I have seen many, many instruments “developed” with nearly as many items as subjects; factor analyses done on small samples and results interpreted as stable and meaningful, etc. Obviously, this is part of the general validity problem. The constructs we measure are subtle and complex. The indicators of those constructs need to be very carefully crafted and tested, and that requires large samples and replication of factor structures before we can be sure of our indices. Samples of 100-200 with as many variables are unfortunately all too common!

A second issue related to sample size is the lack of normative data on most of the extant instruments in the family area. Because most measures have been developed on small or restricted samples, the interpretation of mean level scores remains moot. What is a high score on cohesion? On conflict? What is the norm? This issue often escapes us because we tend to deal in correlations, but the usefulness of family measures would be greatly enhanced if we could interpret mean level scores against a normative base. Let me give you an example from the child psychopathology area—namely the Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1988); Achenbach’s group has normed his behavior problem checklist on large, national (and now international) samples by age, gender, social class, clinical status, etc. (Achenbach, McConaughy, & Howell, 1987). If I use the measure on my sample I can describe the sample relative to those norms (e.g., we have 21% of our children above the 90% percentile on behavior problems, Mavis Hetherington had 80% of her boys of divorced parents scoring above the clinical cutoff, etc.).

What do we know with most of our family measures? Not much. What is a high score? A “clinical” score? It is clear that demographic variables do affect family functioning and that family form (divorced, step, single parent, reconstituted, etc.) will make our assessment job more difficult. Do we study normal or distressed families with the same or different measures? More basically, how do we define family? Can we sample by living arrangements, setting, etc.? Large, diverse samples need to be used during construct development to allow us to begin to develop preliminary answers to these important questions. We simply do not know whether the same instruments will work for all family forms in most settings (allowing us to compare mean levels) or whether we need different kinds of measures.
for different groups. "For example, since wives generally indicate lower levels of marital satisfaction than husbands, does it make sense to use separate norms based on gender or is it preferable to use raw scores?" (Wampler & Halverson, 1993, p. 189). As long as we muddle along on small, convenience samples, we can never begin to address these questions of when, where, and for whom our measure applies, not to mention how we can interpret mean scores.

SUMMARY

So, whither the field of family measurement? Clearly, there are many things left undone in my opinion. My assessment of the maturity of the field is not positive at this time. We still have many measures of many theoretical constructs. Many of those measures have not only poor psychometric qualities, but they are also saturated with method variance and of unknown discriminant validity. I am sure many measures with different names tap mostly the same variance whereas others with the same name (e.g., cohesion) measure quite different things.

We also seem to lag far behind theory in places. I agree with Grotevant and Carlson (1989) that the "theoretically powerful transactional view of socialization processes has not yet been matched in terms of measurement technology" (p. 149). Further, we are still without tests and measures with known normative data and cross-replicated findings from different studies employing "benchmark" measures.

What is to be done? I believe that there most likely will be no nation-wide "rigorous, and programmatic efforts" to improve measurement technology (Jacob, Tennenbaum, & Krahn, 1987, p. 322). It is difficult to fund large-scale measurement studies where there is still much disagreement about the key constructs to be measured. More likely are the cross-laboratory replications of measures derived from programmatic research programs like Patterson’s (1982) and Gottman’s (1979).

For example, Patterson’s (1982) ongoing research has always stressed the need to build constructs from multiple methods and sources. When those multiple-measured constructs are used by multiple investigators across the country, their replicated usefulness as well as the ever-growing nomological net (when these constructs are compared to new ones for predictive efficacy) will help us to know which are the best multiple-source/method constructs to use in our own work. These techniques, many based on new multivariate procedures like LISREL and confirmatory factor analysis, are just now
beginning to have an impact on family measurement. In a sense, we have a very short history of solid, sophisticated measurement that is psychometrically sound and theoretically useful. I remain hopeful as we increasingly emphasize data collection from multiple sources and recognize the importance of replication, more agreement will emerge about the basic dimensions related to family functioning and the best ways to measure them.

REFERENCES


a mild congenital risk factor. In R. Cole, & D. Reiss (Eds.), How do families cope with chronic illness? (pp. 71-94) Hillsdale, NJ: Lawrence Erlbaum Associates.


