ARLAND THORNTON   *The University of Michigan*

GEORGINA BINSTOCK*   *The University of Michigan and Centro de Estudios de Población (CENEP), Buenos Aires, Argentina*

The Reliability of Measurement and the Cross-Time Stability of Individual and Family Variables

Using data from a 31-year panel study, we evaluate both the reliability of measurement of family attitudes, relationships, and self-concepts and the stability of these variables across time. We also compare the reliability of measurement and the stability of variables across time in these domains of family life with the reliability and stability of behavioral dimensions. Our results provide considerable support for the hypothesis that family relationships, attitudes, and self-concepts can be measured reliably. We also demonstrate that self-concepts and family relationships and attitudes have high levels of stability across significant periods of time, and that these are comparable to those for the behavioral indicators we examined.

Research concerning family and child well-being and the causal forces and processes influencing them requires that the research community have an extensive package of clear and reliable concepts and empirical indicators. There is widespread recognition in the research community that procuring valid and reliable information about even the most salient and important dimensions of life is difficult. There are important impediments to the ability and willingness of people to recall, formulate, and report some of the most salient and important facts of their lives, including such things as age, education, marriage, number of children, abortion, employment, health, and income (Alwin, 1989; Alwin & Thornton, 1984; Castro Martin & Bumpass, 1989; Cherlin, Griffith, & McCarthy, 1983; Duncan, Mathiowetz, Cannell, Hill, & Ponza, 1985; Ewbank, 1981; Greenberg & Halsey, 1983; Jones & Forrest, 1992; Madow, 1967; Myers, 1993; Sweet, 1990). Extensive work has gone into enhancing and measuring the reliability of the indicators in these important areas of human life.

Although the difficulty of ascertaining reliable data across all domains of life is widely understood, obtaining information about such cognitive and ideational matters as values, attitudes, goals, preferences, expectations, beliefs, constraints, and religious involvement is frequently seen as additionally problematic (Alwin, 1989; Kalton & Schuman, 1982). Measuring family relationships, interpersonal interactions, and processes may also be especially difficult. Although study participants are usually willing to provide responses to our questions, there is concern about how much their answers reflect the truth about the individual and his or her family relationships and processes.

These concerns about the measurement of attitudes, values, and family relationships and processes are supported by empirical evidence demonstrating that measures of virtually all of these phenomena are subject to measurement error (Alwin, 1989, 1995; Thornton, Orbuch, & Axinn, 1995). However, despite the evidence about the difficulties of measuring attitudes, values, and family processes, we also have considerable evidence that we are able to ascertain reliable differences among individuals.

Another concern of students of family and child well-being is the across-time stability of values, attitudes, and family relationships. Even if these phenomena can be reliably measured, these variables may have little importance for researchers if they are ephemeral and subject to passing whims and fancies. Such ephemeral matters may not be of particular interest as dependent variables and may have little power as explanatory factors (for discussions of the stability of values, attitudes, and family relationships, see Alwin, 1994; Converse, 1964, 1970; Moss & Susman, 1980; Newcomb, 1943; Thornton, Alwin, & Camburn, 1983; Thornton et al., 1995).

In this article we examine both the reliability of measurement and the stability of attributes across time. These issues of reliability and stability are investigated using data and techniques that permit us to separate measurement reliability from trait persistence across time. Although these issues could be investigated across a wide range of domains, we focus our attention most explicitly on the domains of family attitudes and values, interpersonal relationships, and self-esteem.

Following Alwin (1989), we also place our work in a comparative framework and ask how the reliabilities and stabilities of values, self-esteem, and family relationships compare with similar indicators of such phenomena as education and employment. Although Alwin's work makes us expect that such facts as education and employment are reported more reliably than attitudes and values, both his work and that of others demonstrate that there is also measurement error in these phenomena. This article will compare reliabilities and stabilities of family attitudes and values, interpersonal relationships, and self-esteem with similar measures for education and work.

METHOD

Data

The data used in this analysis come from a 31-year eight-wave panel study of mothers and their children. Mothers were originally selected using a probability sample of 1961 birth records from the Detroit Metropolitan Area (Wayne, Oakland, and Macomb counties). Approximately equal numbers of these White married women who had given birth to a first, second, or fourth child were chosen for a personal interview in the winter of 1962. Subsequent telephone interviews were conducted in the fall of 1962, 1963, 1966, 1977, 1980, 1985, and 1993. The children born in 1961 were interviewed in 1980, 1985, and 1993 (at ages 18, 23, and 31 years, respectively).

In 1980, most of the children's interviews were conducted in person (80%), as they were for approximately 50% and 40% of the subsequent interviews conducted in 1985 and in 1993. The rest of the children's interviews were conducted by telephone, and only a small minority was conducted by mail (1%–3% in each of the 3 years). Additionally, in 1980 and 1985, data about children's attitudes toward cohabitation, premarital sex, divorce, marriage, children's self-esteem, and children's relation with mother were obtained through a self-administered questionnaire in those interviews that were conducted in person.

The original 1962 sample had a response rate of 92%, and the study has retained a large percentage of the families over its full 31 years. In 1993, 884 of the mothers and 906 of the children were interviewed, representing 84% of the mothers originally interviewed in 1962 and 81% of their children. Because attrition has been kept extremely low, it has had almost no effect on the composition of the sample.

This analysis relies primarily on data for both mothers and children for 1980, 1985, and 1993. In addition, data concerning the mothers in 1962, 1977, and 1980 are also used.

Measures of Family Attitudes and Values, Interpersonal Relationships, and Self-Esteem

The data set contains multiple measures of a variety of concepts for both mothers and children in 1980, 1985, and 1993. There are two indicators each for cohabitation, divorce, and premarital sex attitudes, four indicators for marriage attitudes, and eight indicators of sex role attitudes. Each of these indicators is identical in wording for mothers and children. For mothers, the data contain five indicators each for her relationship with her husband and for her relationship with her child. For the children, the data contain seven indicators

each for self-esteem and for his or her relationship with the mother. The exact wording of each of the questions is presented in Table 1, with the name of the variable given in parentheses.

For each cohabitation, premarital sex, divorce, sex role, and marriage attitude measure, the respondents were asked to indicate the extent to which they agreed or disagreed with each of the statements listed in Table 1. The response categories were: *strongly agree, agree, disagree,* and *strongly disagree.* For analysis, each of the items was ordered on a 5-point scale with ''uncertain'' or ''depends'' responses coded at the midpoint. Cohabitation, premarital sex, and divorce variables were coded so that high values reflect approval and low values indicate disapproval. Sex role items were ordered so that a high score reflects an egalitarian response and a low score represents a nonegalitarian orientation. Finally, marriage attitude variables were coded so that high values reflect positive attitudes toward singleness and low values indicate disapproval.

For the mother–child relationship items, both mothers and children indicated the extent to which each of correspondent statements were *always true, usually true, sometimes true,* or *never true.* This 4-point scale was maintained in the analysis, and each of the items was coded so that high values reflect in each case a good relationship. The analysis within this domain is restricted to the respondents who referred to the biological mother during each of the 3 years to ensure the comparability of the questions during the 3 years.

For self-esteem, the children were asked to indicate the extent to which each of the seven statements listed in Table 1 was *always true, almost always true, often true, sometimes true, seldom true,* or *never true.* This 6-point scale was maintained in the analysis, and each of the items was coded so that high values reflect high self-esteem.

Finally, each mother was asked four questions in order to assess the quality of her relationship with her husband. Unlike in the other domains, each of the questions that refers to the mother's relationship with her husband possesses a different set of four response categories. Table 1 lists the exact wording and the response alternatives for each question. In order to ensure the comparability of the questions and respondents across the three points in time, the analysis within this domain is restricted to those mothers who were married to the same person from 1980 to 1993.

Measures of Mother's Education, Family Income, Work Hours, and Religious Attendance

We also use data on mother's education, family income, work hours, and religious attendance. Education comes from the single question asked in 1962, 1977, and 1980: ''What is the highest grade of school or year of college you have completed?'' Response categories range from 0 to 17 and more. Mother's family income was measured in all eight surveys through a question that inquires about the total income received by the mother and her family during the previous year. The income figures include total earnings from jobs and money received from Social Security, unemployment compensation, and income from investments. Work hours comes from the single question, ''About how many hours do you work on your main job in an average week?'' The codes reflect the actual hours, with zero hours representing not working. Finally, religious attendance comes from the question: ''How often do you usually attend religious services—would you say several times a week, once a week, a few times a month, once a month, or less than once a month?'' Respondents who volunteered that they never attended were so coded in the analysis.

Estimating Reliability and Stability

Our methodology for estimating reliability and stability follows the single-variable three-wave panel procedures outlined by Alwin (1989) and summarized in Figure 1. As Figure 1 indicates, we posit a model with a single underlying true score (τ), which has a single indicator (y) that is observed at three points in time: 1980, 1985, and 1993.

The observed indicators in 1980, 1985, and 1993 are assumed to be linked to the underlying constructs in the same year through Equations 1, 2, and 3, where the observed indicator (y) is conceptualized as being composed of the true score (τ) and error of measurement (ϵ), which is assumed to be random. Reliability is a concept that refers to how much an indicator represents the underlying construct rather than the error component. More specifically, it indicates the extent to which the variance of a particular variable can be said to be ''true'' variance rather than ''random error'' variance. Reliability is operationalized as the ratio of the variance of the true score (τ) to the variance of the observed indicator (y). This is also the square of the correlation between τ and

TABLE 1. MEASURES OF FAMILY ATTITUDES AND VALUES, INTERPERSONAL RELATIONSHIPS, AND SELF-ESTEEM, MOTHERS AND CHILDREN

A. Cohabitation

(1) It's all right for a couple to live together without planning to get married. (live together all right)
(2) A young couple should not live together unless they are married. (should not live together)

B. Premarital sex

(1) Young people should *not* have sex before marriage. (before marriage)
(2) Premarital sex is all right for a young couple planning to get married. (planning marriage)

C. Divorce

(1) When there are children in the family, parents should stay together even if they don't get along. (not stay together)
(2) Divorce is usually the best solution when a couple can't seem to work out their marriage problems. (divorce best)

D. Sex role

(1) Most of the important decisions in the life of the family should be made by the man of the house. (decisions)
(2) It's perfectly all right for women to be active in clubs, politics, and other outside activities before the children are grown up. (women active)
(3) There is some work that is men's and some that is women's, and they shouldn't be doing each other's. (men's/ women's work)
(4) A wife shouldn't expect her husband to help around the house after he's come home from a hard day's work. (housework)
(5) A working mother can establish as warm and secure a relationship with her children as a mother who does not work. (relations/working mother)
(6) It is much better for everyone if the man earns the main living and the woman takes care of the home and family (men work/women home)
(7) Women are much happier if they stay at home and take care of their children. (women happier home)
(8) It is more important for a wife to help her husband's career than to have one herself. (husband/wife career)

E. Marriage

(1) Married people are usually happier than those who go through life without getting married. (married happier)
(2) There are few good or happy marriages these days. (few good marriages)
(3) It's better for a person to get married than to go through life being single (better married)
(4) All in all, there are more advantages to being single than to being married. (better single)

F. Mother's relationship with child

(1) NAME's ideas and opinions about the important things in life are ones you can respect. (respects child)
(2) He or she respects your ideas and opinions about the important things in life (child respects mother)
(3) You find it easy to understand him or her. (understands child)
(4) You enjoy doing things together with NAME. (enjoy doing with child)
(5) You enjoy talking to NAME. (enjoys talking child)

G. Mother's relationship with husband

(1) How well do you think your husband understands you—your feelings, your likes and dislikes, and any problems you may have; do you think that he understands you very well, fairly well, not very well, or not well at all? (husband understands mother)
(2) How well do you think you understand your husband—very well, fairly well, not very well, or not well at all? (mother understands husband)
(3) Generally speaking, would you say that the time you spend together with your husband is extremely enjoyable, very enjoyable, enjoyable, not too enjoyable? (time spent together)
(4) Taking things all together, how would you describe your marriage—would you say your marriage is very happy, a little happier than average, just about average, or not too happy? (happy marriage)

H. Child's relationship with mother

(1) My mother's ideas and opinions about the important things in life are ones I can respect. (child respects mother)
(2) My mother respects my ideas and opinions about the important things in life. (mother respects child)
(3) My mother accepts and understands me as a person. (mother understands child)
(4) I enjoy doing things together with my mother. (enjoy doing with mother)
(5) My mother makes it easy for me to confide in her. (confides mother)
(6) My mother gives me the right amount of affection. (affection mother)
(7) When something is bothering me, I am able to talk it over with my mother. (talks over mother)

I. Child's self-esteem

(1) I take a positive attitude toward myself. (positive attitude)
(2) I feel I do not have much to be proud of. (not proud)
(3) I am able to do things as well as most other people. (able to do)
(4) I feel that I can't do anything right. (anything right)

TABLE 1. CONTINUED.

(5) As a person I do a good job these days. (do good job)
(6) I feel that I have a number of good qualities. (have good qualities)
(7) I feel that I'm a person of worth, at least on an equal level with others. (person of worth)

y and ranges from zero to 1. Thus a reliability of 1 indicates that all of the variance in the indicator reflects the variance of the underlying construct, whereas a reliability of zero indicates that none of the variance in the indicator comes from the underlying construct.

$$y_1 = \tau_1 + \epsilon_1 \quad (1)$$
$$y_2 = \tau_2 + \epsilon_2 \quad (2)$$
$$y_3 = \tau_3 + \epsilon_3 \quad (3)$$

Simplifying assumptions are required to estimate the model. Following Heise (1969) and Alwin (1989), we assume that the reliabilities of the measures are constant across time. We implement this assumption by standardizing the observed variables with means of zero and unit variance and fixing the variance of the errors of measurement to be constant across the three observation waves.

Equations 4 and 5 are posited to link the latent variables or true scores together across time (see Figure 1). This model posits a Markovian or lag-1 process in which each occurrence of the underlying construct is assumed to be the product only of the immediately preceding occurrence of that construct and an error of prediction (ζ) (Alwin, 1989; Jöreskog, 1970, 1974). That is, the 1980 occurrence of an underlying construct is assumed

FIGURE 1. SINGLE-ITEM RELIABILITY AND STABILITY MODEL

to influence the 1993 construct only through its influence on the 1985 construct. The beta coefficients represent the stability of the underlying constructs across time. They indicate the extent to which individual scores on an underlying construct—purged of measurement error—correlate across time. Given that each of the equations has only one predictor variable, the beta coefficients are also equal to the correlations between two observations of the same underlying construct at two points in time.

$$\tau_2 = \beta_{21}\tau_1 + \zeta_2 \quad (4)$$
$$\tau_3 = \beta_{32}\tau_2 + \zeta_3 \quad (5)$$

As indicated earlier, many of the measures in the data set were designed to be multiple indicators of the same underlying construct. The psychometric literature indicates that scales composed of items designed to measure the same underlying construct have greater reliability than the individual items themselves (Kim & Mueller, 1978; Nunnally, 1967). We evaluated the effectiveness of this approach with our measures by forming domain-specific scales and estimating their reliabilities and stabilities. We did this by adding together and averaging individual items within each domain to form a domain-specific scale and then utilizing the model in Figure 1 to estimate reliabilities and stabilities for the scales, exactly as just described for individual items. We estimated these models using LISREL (Jöreskog & Sörbom, 1993).

Measurement error, which we assume is random, contains error from both the population and the measurement instrument. As described in the Data section, we analyze the same population over time. However, the mothers were interviewed in 1962 in face-to-face interviews and thereafter by telephone. In addition, three different modes of administering the questionnaire were applied among children: personal, telephone, and self-administered interviews. We are aware that this procedure violates the assumption of constancy of measurement conditions. Although our procedures must assume that error variance remains invariant over time, this assumption may be violated to some degree for the children because of different
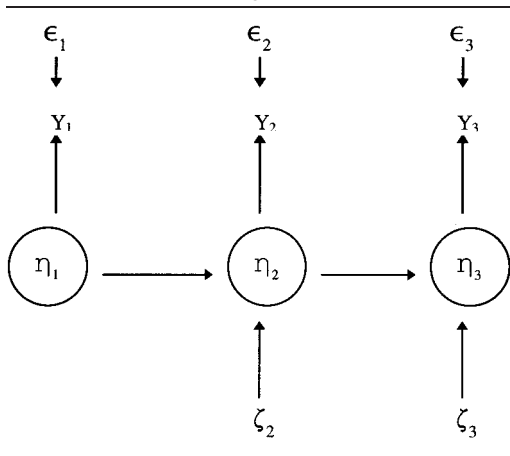
modes of administration both within and across interview waves. Mode-of-interview differences between mothers and children also affect our ability to compare reliability and stability across generations. Although the mode of interview is constant for mothers after 1962, any analyses including the 1962 data involve the same violation of the constancy of measurement assumption.

<center>Results</center>

<center>*Reliabilities of Attitudes, Relationships,*
and Self-Esteem</center>

Table 2 provides a summary of the reliability estimates obtained from the model of Figure 1. Those reliabilities are the ratios of the variance of the true scores (τ) to the variance of variables (y) and indicate the proportion of the variance in the observed variables accounted for by their underlying true scores or, in other words, how well the concept or variable is being measured. Included in Table 2 are reliabilities for both individual measures and for scales formed by summing together and averaging the individual measures. The reliabilities for individual measures are in the first column of coefficients, whereas the reliabilities for the scales are listed in the second column of coefficients. Also listed for each group of variables is the average reliability of the individual items within that group—measured by the simple arithmetic mean.

Examination of Table 2 reveals a substantial range of reliabilities across the different individual measures included in the study, ranging from .25 to .71, with the bulk of the reliabilities between .4 and .7. These data clearly indicate that the bulk of the items measured are capturing substantial amounts of true variance; between 40% and 70% of the observed variance in most of the measures represents true variance in the underlying variables. This, of course, is contrary to the position that measures of attitudes, relationships, and personal identities reflect only noise and no reliable variance. Of course, at the same time, a significant part of the observed variance of most items—between 30% and 60%—can be attributed to random error.

Looking in more detail at the reliabilities, we see that many of the family domains include items measured with quite high reliabilities—.64 or higher. These include individual items from the domains of cohabitation attitudes, premarital sex attitudes, sex role attitudes, the mother's relation-

ship with her husband, and the child's relationship with his or her mother. Although there are also items in some of these domains with lower reliabilities, these results indicate that it is possible to obtain substantial reliabilities across a number of different domains of family attitudes and relationships.

Although the attitudinal, relationship, and identity items with relatively high reliability are scattered across several domains, the data suggest that reliabilities may be generally higher in some domains than in others. For example, the estimated reliabilities for both of the premarital sex and cohabitation items for both mothers and children are clustered toward the high end of the reliabilities observed (averaging from .63 to .69). This suggests that cohabitation and premarital sex may have been more salient, crystallized, and central to young people and their parents during the 1980s and 1990s than some of the other domains examined.

Among the attitudinal domains, the reliabilities of the marriage and divorce attitudinal items cluster toward the lower end of the range. This is particularly true for some of the marriage items—with the item about remaining single being better than marriage (better single) having reliabilities around .3 for both mothers and children. The reliabilities for the divorce items for the children are similarly low but are more substantial for the mothers. Although the highest reliabilities in the marriage and divorce domains (.59) are comparable with those of numerous other items, the overall low levels of reliability suggest that people may be more ambivalent and less clear about marriage and divorce than about some of the other domains.

The higher reliabilities on divorce for mothers than for children suggest that there may be less ambivalence in this area for mothers than for children. This may be related to the fact that divorce is seen very differently by children than by their parents (Thornton, 1985, 1989).

Although there is a significant range of reliabilities among the eight sex role attitude items, they generally cluster in the middle range of reliabilities (means = .56 for mothers and .49 for children). For mothers, the bulk of the reliabilities are between .5 and .6. The sex role attitude reliabilities, however, are somewhat lower among the children than among the mothers. For both mothers and children, the item about decision making seemed to be measured particularly reliably (.62 for mothers and .59 for children). The highest re-

liability for mothers was for the question about the husband's career (husband/wife career) (.7), but this item was not particularly reliably measured among the children (.46). Similarly, the highest reliability item among the children, husband helping at home (housework) (.7), was not measured especially reliably among the mothers (.47). These inconsistencies across generations make it difficult to identify exactly which of the items might generally have the highest reliabilities.

Turning now to the relationship items, we see that the mother's relationship with her husband is measured with the greatest reliability, with most of these reliabilities clustered around .6 (mean of .58). Thus, husband–wife relationships are measured nearly as reliably as the premarital sex and cohabitation attitudes items. The mother's relationship with her child is measured substantially less reliably; these reliabilities seem to center around .4 (mean of .41). The reliabilities of the child's relationship (mean of .51) with his or her mother generally seem to lie between the two sets of mother items. With one notable exception (having a positive attitude toward self), the self-esteem items also have relatively low reliabilities. Even with the one outlier, the average reliability in the self-esteem domain is only .38.

Note that as predicted in the psychometric literature, the reliabilities of the scales are substantially higher than the reliabilities of the individual items. As expected, this is particularly true for the scales with large numbers of items. Whereas the least reliable scale has a reliability of only .45 (divorce attitudes for children), all of the remaining scales have reliabilities of .58 or higher. Furthermore, many of the scale reliabilities exceed .7, with the highest being .81 (sex role attitudes for mothers). Thus, by measuring multiple items in the same domain and combining them into a scale, reliabilities can be increased substantially.

We have mentioned some differences in reliabilities between mothers and children. For example, at least on the items measured in this study, mothers tend to have higher reliabilities than children concerning attitudes toward divorce (.63 vs. .45 on the scale) and sex role attitudes (.81 vs. .73 on the scale), whereas intergenerational relationships may be more reliably measured for children than for mothers (difference between .72 and .58 on the scale, but with different numbers of items). In the other domains there do not appear to be any substantial or consistent differences between mothers and children. These findings suggest that although certain issues may be more salient or central to one generation than to the other, there do not seem to be any overwhelming age or generation differences in overall measurement reliability—at least across the ages observed in this study. This finding is consistent with Alwin's (1989) conclusions about age differences across a wide range of general measures.

Reliabilities of Income, Education, Work, and Religious Attendance

In order to obtain comparative information about the reliabilities and stabilities of behavioral phenomena we estimated single-item models for family income, mother's education, mother's work hours, and frequency of religious services using the same procedures used earlier. The reliabilities of these measures are reported in Table 3. Note that some of these models are estimated using data from the 1962, 1977, and 1980 waves of interviews, whereas others come from the 1980–1993 interviews.

As expected, the reliabilities of the behavioral indicators are generally higher than the attitudinal and relational items examined earlier. The lowest estimated reliabilities among the behavioral measures were for the mother's hours of employment and family income, which were around .7. The highest reliability observed was for mother's education, which exceeded .9. Most of the reliabilities for these measures were between .7 and .8.

As a group, the reliabilities of the behavioral indicators are clearly higher than those for individual indicators of family attitudes, relationships, and self-evaluation. However, with the exception of education, the reliabilities of the behavioral measures are clearly in the same general area of the most reliably measured attitudinal and relationship indicators—around .7. Thus, the reliabilities of the more behavioral indicators are, with the exception of education, not all that different from the most reliably measured attitudinal and relational indicators. Furthermore, when we consider the reliabilities of the attitudinal and relational indicators combined as scales with the reliabilities of the behavioral indicators, the scales compare very favorably. As discussed earlier, their reliabilities range from .45 to .81. These data, therefore, provide little support to the position that attitudinal, relational, and self-concept matters cannot be measured reliably but that behavioral information can.

TABLE 2. ESTIMATED MEASUREMENT RELIABILITIES FOR ATTITUDES, RELATIONSHIPS, AND SELF-ESTEEM, MOTHERS AND CHILDREN

	Mothers		Children	
	Individual Measures	Scales	Individual Measures	Scales
Cohabitation		.74		.74
Live together all right	.67		.62	
Should not live together	.63		.71	
Average	.65		.67	
Premarital sex		.71		.79
Before marriage	.57		.70	
Planning marriage	.68		.68	
Average	.63		.69	
Divorce		.63		.45
Not stay together	.60		.42	
Divorce best	.53		.32	
Average	.57		.37	
Sex role		.81		.73
Decisions	.62		.59	
Women active	.57		.40	
Men's/women's work	.50		.44	
Housework	.47		.70	
Relations/working mother	.57		.42	
Man work/women home	.58		.52	
Women happier home	.46		.39	
Husband/wife career	.70		.46	
Average	.56		.49	
Marriage		.59		.60
Married happier	.53		.46	
Few good marriages	.55		.37	
Better married	.39		.52	
Better single	.31		.29	
Average	.45		.41	
Average of Cohabitation through Marriage items	.55		.50	
Relation with child		.58		
Respects child	.47			
Child respects mother	.31			
Understands child	.45			
Enjoy doing with child	.44			
Enjoys talking child	.38			
Average	.41			
Relation with husband		.73		
Husband understands mother	.59			
Mother understands husband	.46			
Time spent together	.60			
Happy marriage	.65			
Average	.58			
Relation with biological mother				.72
Child respects mother			.53	
Mother respects child			.48	
Mother understands child			.50	
Enjoy doing with mother			.49	
Confides mother			.66	
Affection mother			.47	
Talks over mother			.50	
Average			.51	

TABLE 2. CONTINUED.

Self-esteem		
Positive attitude	.56	
Not proud	.25	
Able to do	.38	
Anything right	.33	
Do good job	.34	
Have good qualities	.42	
Person of worth	.41	
Average	.38	.60

Note: The number of cases ranges from 835 to 870, except for mother's relation with husband (602) and child's relation with biological mother (780).

TABLE 3. MEASUREMENT RELIABILITIES FOR INCOME, EDUCATION, WORK, AND RELIGIOUS ATTENDANCE, MOTHERS AND CHILDREN

Variable	Reliability
Mothers	
Family income[a]	
1962, 1977, and 1980	.79
1977, 1980, and 1985	.77
1980, 1985, and 1993	.78
1980, 1985, and 1993 (for mothers married to the same husband during the period	.69
Number of years of schooling, 1962, 1977, and 1980[b]	.95
Number of hours worked per week, 1980, 1985, and 1993 (for all mothers; nonworking = 0 hours)	.71
Number of hours worked per week, 1980, 1985, and 1993 (for all mothers working in all 3 years)	.72
Frequency of religious attendance, 1980, 1985, and 1993	.77
Children	
Frequency of religious attendance; 1980, 1985, and 1993	.74

Note: The number of cases range from 705 to 966, except for number of hours worked among mothers who worked in all 3 years (378) and family income among mothers married to the same husband (550).

[a] In 1962, family income is coded in $1,000 income groups, and it was coded at the midpoint except the last category ($15,000 or more), which was coded at $20,000. The wording of the 1962 question is slightly different than in the rest of the years. Additionally, there are slight differences in terms of specifying ''before taxes'' or not, ''mother and husband'' versus ''mother and family'' in the rest of the years (1977 to 1993).

[b] In 1962, categories were: 0–4, 5–8, 9–11, 12, 13–15, and 16 or more, and they were coded at the midpoint. Also, the wording of the questions was slightly different in 1962 than in 1977 and 1980.

Stabilities of Attitudes, Relationships, and Self-Esteem

We now shift our attention from the reliability of measurement to the stability of underlying constructs across time. Using the model of Figure 1, we estimated how each of the attitudinal domains remained stable across the 5-year period from 1980 through 1985 and the 8-year period from 1985 through 1993 among mothers and children. In order to take into account the differential periods of time in the two intervals, we converted the 5-year stabilities from 1980 through 1985 into 8-year stabilities by taking the estimated stabilities to the 8/5 power. The estimated 8-year stabilities for the various measures are provided in Table 4. The 8-year stabilities from the single-item models are shown in the first and second columns for the 1980–1985 and 1985–1993 periods, respectively. The third and fourth columns provide comparable stabilities for the multiple items summed and averaged into scales.

One of the most noteworthy patterns in Table 4 is the consistent and strong difference in stabilities for the children between the first and second intervals. With only two exceptions, all of the stabilities between 1980 and 1985 (ages 18 and 23) are lower than the comparable stabilities between 1985 and 1993 (ages 23 and 31) (controlling for length of period). Furthermore, almost all of these differences in stability across time are of substantial magnitude, with a substantial number of differences exceeding .2. This finding of a substantially increasing level of stability across time is consistent with the hypothesis that attitudes and relationships solidify and crystallize significantly between the years centered around age 20 and the period of the late 20s. This finding is consistent with Alwin's (1994) earlier findings that attitudes and values are particularly unstable during the years of early adulthood but quickly become more

TABLE 4. EIGHT-YEAR STABILITIES OF ATTITUDES, RELATIONSHIPS, AND SELF-ESTEEM, MOTHERS AND CHILDREN

	Mothers				Children			
	Individual Measures		Scales		Individual Measures		Scales	
	1980–1985	1985–1993	1980–1985	1985–1993	1980–1985	1985–1993	1980–1985	1985–1993
Cohabitation			0.84	0.89			0.62	0.83
Live together all right	0.86	0.89			0.69	0.80		
Should not live together	0.84	0.88			0.53	0.84		
Premarital sex			0.94	0.90			0.49	0.76
Before marriage	0.92	0.90			0.49	0.76		
Planning marriage	0.92	0.89			0.45	0.78		
Divorce			0.70	0.83			0.74	0.88
Not stay together	0.76	0.88			0.83	0.92		
Divorce best	0.54	0.77			0.77	0.92		
Sex role			0.86	0.92			0.74	0.79
Decisions	0.83	0.88			0.66	0.82		
Women active	0.73	0.84			0.58	0.85		
Men's/women's work	0.86	0.92			0.70	1.00		
Housework	0.84	0.91			0.41	0.45		
Relations/working mother	0.64	0.76			0.57	0.72		
Men work/women home	0.82	0.90			0.49	0.75		
Women happier home	0.95	0.88			0.64	0.66		
Husband/wife career	0.64	0.72			0.66	0.76		
Marriage			0.70	0.81			0.54	0.72
Married happier	0.60	0.85			0.60	0.75		
Few good marriages	0.64	0.81			0.79	1.00[a]		
Better married	1.00[a]	0.99			0.54	0.64		
Better single	0.64	1.00[a]			0.70	0.86		
Relation with child			0.97	0.87				
Respects child	1.00[a]	0.84						
Child respects mother	1.00[a]	1.00[a]						
Understands child	0.74	0.90						
Enjoy doing with child	0.95	0.86						
Enjoys talking child	1.00	0.77						
Relation with husband			0.84	0.90				
Husband understands mother	0.86	0.89						
Mother understands husband	0.98	0.88						
Time spent together	0.84	0.90						
Happy marriage	0.82	0.96						
Relation with biological mother							0.51	0.77
Child respects mother					0.35	0.73		
Mother respects child					0.44	0.85		
Mother understands child					0.54	0.75		
Enjoy doing with mother					0.63	0.80		
Confides mother					0.44	0.78		
Affection mother					0.76	0.93		
Talks over mother					0.62	0.85		
Self-esteem							0.73	0.86
Positive attitude					0.69	0.80		
Not proud					1.00[a]	0.95		
Able to do					0.67	0.79		
Anything right					0.73	0.92		
Do good job					0.79	0.72		
Have good qualities					0.70	0.91		
Person of worth					0.66	0.90		

Note: These models sometimes estimate beta coefficients greater than 1. The reason is that the correlation between observations at Times 1 and 3 is infrequently bigger than the correlation at Times 1 and 2 or at Times 2 and 3. The actual beta coefficients from Equations 3 and 5 in these cases range from 1.02 to 1.10 but are listed as 1.0.

[a]Parameter listed as 1.0 as noted.

stable in later years. Note that whereas the stabilities of children's attitudes and relationships increase dramatically between the 1980–1985 and 1985–1993 periods, the children's stabilities during the latter period are, with a small number of exceptions, still lower than the mothers' stabilities during the same years. It is clear from these data that the maximum level of stability in adulthood was probably not reached for this cohort of young adults by age 23, because their later stabilities were still generally below those of their mothers. It should be noted that although we are interpreting the differences in children's stabilities across time as representing aging effects, it is also possible that the differences were due to historical period, which in this sample of young adults is perfectly conflated with age. Note, however, that the possibility of the age differences among the children being the result of differences in historical period is reduced by the observation that the mothers did not show the same substantial and consistent differences in stabilities between 1980–1985 and 1985–1993.

Although there were no general changes in the stabilities of maternal attitudes and relationships, there appear to be some noticeable shifts in certain domains. Perhaps most noticeable among these are the increases in stability of attitudes concerning marriage and divorce. However, this upward movement in these two domains was generally balanced by a decline in stabilities of relationships with the study child. Explanations concerning these differential trends are not readily apparent.

Shifting now from generational, age, and time differences to the effects of substantive domains, we find few large and consistent differences. Particularly in the 1985–1993 period, the stabilities are virtually all clustered within a relatively limited band. Furthermore, it appears that there is relatively little variability in stabilities within or between substantive domains. We find these similarities in stabilities across various attitudes, interpersonal relationships, and self-identity to be remarkable.

We are also impressed with the relatively high magnitudes of the stabilities of attitudes, relationships, and identities. Discounting the fluidity of the late teenage and early 20s years, we note that the children's stabilities generally cluster around the .8 mark. Maternal stabilities for the 8-year period between 1985 and 1993 are, as noted earlier, somewhat higher than the children's—being generally above .8, with many ranging up to .9 and higher. This means that during the mature adult years, these subjective factors at Time 1 generally share from about two thirds to about four fifths of their variance with the same subjective factor 8 years later.

We believe that stabilities in this range are clearly contrary to the view that attitudes, relationships, and self-identities are ephemeral qualities with little continuity. To the contrary, these results indicate rather substantial stability of these important dimensions of people's lives. Of course, we know that if we were to look at the continuity of attitudes, relationships, and identities over a longer period of time, the stabilities would be lower. For example, over a 16-year period, the stabilities would be approximately the square of the 8-year stabilities reported in Table 4. Nevertheless, we believe that such high stabilities over 8 years reflect substantial continuity rather than ephemerality.

Also note that the estimated stabilities in the model utilizing scales of individual items are remarkably consistent with the stabilities estimated from single-variable models. That is, the stabilities from the models using scales tend to fall into the middle of the range of stabilities from the single-item models.

Stabilities of Income, Education, Work, and Religious Attendance

Table 5 provides estimates of the stability of income, education, work, and religious attendance, estimated comparably to the earlier procedures. General assurance about the stabilities of attitudinal and relational phenomena relative to behavioral indicators is provided in Table 5. Education, which tends to become quite fixed later in life, had an 8-year stability for the mothers of .94 between 1962 and 1977 and .97 between 1977 and 1980. However, the stabilities of mother's family income measured in 8-year units range from a low of only .63 to a high of .83, depending on period and sample definition. The stabilities of the mother's work hours were also generally at a middle level—ranging from .55 to .70, again depending on the period and sample definition. Comparisons of the stabilities of these behavioral indicators with the stabilities of the attitudinal and relational items ascertained from the mothers certainly do not support the proposition that the attitudinal and relational factors are ephemeral whereas behavioral factors are consistently more stable.

Variable	Individual Measures	
	1980–1985	1985–1993
Mothers		
Family income[a]		
1962 to 1977 to 1980	.68	.78
1977 to 1980 to 1985	.78	.67
1980 to 1985 to 1993	.63	.69
1980 to 1985 to 1993 (for mothers married to the same husband during the period)	.83	.83
Number of years of schooling, 1962 to 1977 to 1980[b]	.94	.97
Number of hours worked per week, 1980 to 1985 to 1993 (for all mothers; nonworking = 0 hours)	.70	.69
Number of hours worked per week, 1980 to 1985 to 1993 (for all mothers working in all 3 years)	.55	.65
Frequency of religious attendance, 1980 to 1985 to 1993	.89	.91
Children		
Frequency of religious attendance; 1980 to 1985 to 1993	.67	.82

[a]In 1962, family income is coded in $1,000 income groups, and it was coded at the midpoint except the last category ($15,000 or more), which was coded at $20,000. The wording of the 1962 question is slightly different than in the rest of the years. Additionally, there are slight differences in terms of specifying "before taxes" or not, "mother and husband" versus "mother and family" in the rest of the years (1977 to 1993).

[b]In 1962, categories were: 0–4, 5–8, 9–11, 12, 13–15, and 16 or more, and they were coded at the midpoint. Also, the wording of the questions was slightly different in 1962 than in 1977 and 1980.

DISCUSSION

This article began with the observation that empirical studies of family and child well-being require reliable measures of numerous dimensions of well-being and the factors that influence them. Although many of the important dependent and independent variables of interest are behavioral, many other dimensions important to the study of family and child well-being come from the domains of values, preferences, attitudes, relationships, and self-concepts. At the same time that researchers recognize the difficulties of obtaining reliable measures of behavior and experience, there are additional concerns about the difficulty of reliably measuring the dimensions of attitudes, relationships, and self-concepts. In addition, attitudes, relationships, and self-concepts are sometimes seen as ephemeral, and therefore unimportant, in studies of the long-term well-being of children and families. The purpose of this article is to evaluate both the reliability of measurement of family attitudes, relationships, and self-concepts and the stability of these variables across time. We also compare the reliability of measurement and the stability of variables across time in these domains of family life with the reliability and stability of behavioral dimensions.

The data analyzed in this article provide considerable support for the hypothesis that family relationships, attitudes, and self-concepts can be measured reliably. Our empirical estimates indicate that substantial fractions of the variance in observed indicators are shared with their underlying constructs. In addition, if multiple indicators of the same underlying constructs are obtained and summed together into a scale, reliability levels reach even higher levels for most of the domains we studied. In fact, with one exception, the amount of variance shared between the scales we constructed and their underlying construct ranged from .58 to .81.

As expected, the reliabilities of the individual family relationships, attitudes, and self-concept measures were generally lower than the reliabilities for behavioral measures, although the differences were not as great as might be expected. In addition, the reliabilities of the scales compare quite favorably to the reliabilities of the behavioral indicators.

Our results also suggest that if researchers do not use techniques that adjust explicitly for measurement error, the use of multi-item scales is preferable to using single items. The reason is that, as we have seen, the scales have higher measurement reliabilities.

However, recent advances in psychometric techniques make it possible for researchers to adjust for measurement error explicitly within their analytical models. By combining measurement and analytical models in the same analysis, researchers are now able to obtain estimates of structural or substantive parameters while at the same time taking into account the measurement errors in the data (Bollen, 1989; Jöreskog & Sörbom, 1979). Although these structural equation models require assumptions and complex estimation techniques, they are now widely used in the social sciences (e.g. Hauser & Sin-Kwok Wong, 1989; Hauser & Warren, 1997; Teachman, 1995;

Thornton et al., 1995). These techniques can be implemented with only a modest number of indicators of the same underlying construct.

We have also demonstrated that self-concept and family relationships and attitudes are not ephemeral variables with little continuity across time. In fact, these dimensions have high levels of stability across significant periods of time. Furthermore, their levels of stability are comparable to those for the behavioral indicators we examined.

At the same time that our data suggest considerable reliability in the measurement of family attitudes, relationships, and self-concept, they also indicate that our measures include substantial measurement error. Furthermore, the amount of measurement error varies from dimension to dimension. This finding suggests that we have little room for complacency about the current status of our measurement tools. There is room for improvement in all of our measures, and for some dimensions we have examined, the improvement needed is substantial. However, although further improvements are needed, our results suggest that our current measurement capabilities are sufficient to support high-quality empirical research using indicators of these central concepts in the study of family and child well-being.

REFERENCES

Alwin, D. F. (1989). Problems in the estimation and interpretation of the reliability of survey data. *Quality & Quantity, 23,* 277–331.

Alwin, D. F. (1994). Aging, personality, and social change: The stability of individual differences over the adult life span. In D. L. Featherman, R. M. Lerner, & M. Perlmutter (Eds.), *Life-span development and behavior* (Vol. 12, pp. 135–185). Hillsdale, NJ: Lawrence Erlbaum.

Alwin, D. F. (1995, April). *The reliability of survey data.* Paper presented at the International Conference on Measurement Error, Bristol, England.

Alwin, D. F., & Thornton, A. (1984). Family origins and the schooling process: Early vs. late influence of parental characteristics. *American Sociological Review, 49,* 784–802.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley.

Castro Martin, T., & Bumpass, L. L. (1989). Recent trends in marital disruption. *Demography, 26,* 37–51.

Cherlin, A., Griffith, J., & McCarthy, J. (1983). A note on maritally-disrupted men's reports of child support in the June 1980 Current Population Survey. *Demography, 20,* 385–389.

Converse, P. E. (1964). The nature of belief systems in the mass public. In D. E. Apter (Ed.), *Ideology and discontent* (pp. 168–189). London: Free Press of Glencoe.

Converse, P. E. (1970). Attitudes and non-attitudes: Continuation of a dialogue. In E. R. Tufte (Ed.), *The quantitative analysis of social problems* (pp. 206–261). Reading, MA: Addison-Wesley.

Duncan, G. J., Mathiowetz, N. A., Cannell, C. F., Hill, D. H., & Ponza, M. (1985). *A validation study of economic survey data* (Research Report Series). Ann Arbor: Survey Research Center, Institute for Social Research, The University of Michigan.

Ewbank, D. C. (1981). *Age misreporting and age-selective underenumeration: Sources, patterns and consequences for demographic analysis.* Washington, DC: National Academy Press.

Greenberg, D., & Halsey, H. (1983). Systematic misreporting and effects of income maintenance experiments on work effort: Evidence from the Seattle-Denver experiment. *Journal of Labor Economics, 1,* 380–407.

Hauser, R. M., & Sin-Kwok Wong, R. (1989). Sibling resemblance and intersibling effects in educational attainment. *Sociology of Education, 62,* 149–171.

Hauser, R. M., & Warren, R. J. (1997). Socioeconomic indexes for occupations: A review, update, and critique. *Sociological Methodology, 27,* 177–298.

Heise, D. R. (1969). Separating reliability and stability in test–retest correlations. *American Sociological Review, 34,* 93–101.

Jones, E. F., & Forrest, J. (1992). Underreporting of abortion in surveys of U.S. women: 1976 to 1988. *Demography, 29,* 113–126.

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology, 23,* 121–145.

Jöreskog, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In D. H. Kranz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Measurement, psychophysics, and neural information processing* (pp. 1–56). San Francisco: Freeman.

Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models* (J. Magidson, Ed.). Cambridge, MA: Abt Books.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Hillsdale, NJ: Lawrence Erlbaum.

Kalton, G., & Schuman, S. (1982). The effect of the question on survey responses: A review. *Journal of the Royal Statistical Association, 145,* 42–73.

Kim, J., & Mueller, C. W. (1978). *Factor analysis. Statistical methods and practical issues* (Sage University Paper Series on Quantitative Applications in the Social Sciences No. 14). Beverly Hills, CA: Sage.

Madow, W. G. (1967). *Interview data on chronic conditions compared with information derived from medical records* (PHS Vital and Health Statistics, Series 2, No. 23). Washington, DC: Public Health Service.

Moss, H. A., & Susman, E. J. (1980). Longitudinal study of personality development. In O. G. Brim, Jr., & J. Kagan (Eds.), *Constancy and change in human development* (pp. 530–595). Cambridge, MA: Harvard University Press.

Myers, R. J. (1993). Errors and bias in the reporting of ages in census data. In D. J. Bogue, E. E. Arriaga, & D. L. Anderton (Eds.), *Readings in population research methodology* (Vol. 1, pp. 23–29). Chicago:

United Nations Population Fund by Social Development Center.

Newcomb, T. M. (1943). *Personality and social change: Attitude formation in a student community.* New York: Dryden.

Nunnally, J. C. (1967). *Psychometric theory.* New York: McGraw-Hill.

Sweet, J. A. (1990). *Differentials in precision of reporting of dates of marital events in the National Survey of Families and Households* (NSFH Working Paper No. 20). Madison: University of Wisconsin, Center for Demography and Ecology.

Teachman, J. D. (1995). A procedure for estimating the potentially biasing influence of families on relation-ships between variables. *Journal of Family Issues, 16,* 519–537.

Thornton, A. (1985). Changing attitudes toward separation and divorce: Causes and consequences. *American Journal of Sociology, 90,* 856–872.

Thornton, A. (1989). Changing attitudes toward family issues in the United States. *Journal of Marriage and the Family, 51,* 873–898.

Thornton, A., Alwin, D. F., & Camburn, D. (1983). Causes and consequences of sex-role attitudes and attitude change. *American Sociological Review, 48,* 211–227.

Thornton, A., Orbuch, T. L., & Axinn, W. G. (1995). Parent–child relationships during the transition to adulthood. *Journal of Family Issues, 16,* 538–564.