

University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Faculty Publications, Department of Psychology

Psychology, Department of

1992

In Stories We Trust: Studies of the Validity of Autobiographies

George S. Howard
University of Notre Dame

Arthur C. Maerlender
University of Notre Dame, amaerlender2@unl.edu

Paul R. Myers
University of Notre Dame

Tom D. Curtin
University of Notre Dame

Follow this and additional works at: <http://digitalcommons.unl.edu/psychfacpub>

 Part of the [Psychology Commons](#)

Howard, George S.; Maerlender, Arthur C.; Myers, Paul R.; and Curtin, Tom D., "In Stories We Trust: Studies of the Validity of Autobiographies" (1992). *Faculty Publications, Department of Psychology*. 721.
<http://digitalcommons.unl.edu/psychfacpub/721>

This Article is brought to you for free and open access by the Psychology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, Department of Psychology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



Published in *Journal of Counseling Psychology* 39:3 (1992), pp. 398–405.

Copyright © 1992 American Psychological Association, Inc. Used by permission.

Submitted May 28, 1991; revised October 25, 1991; accepted October 31, 1991.

In Stories We Trust: Studies of the Validity of Autobiographies

George S. Howard, Arthur C. Maerlender, Paul R. Myers,
and Tom D. Curtin

University of Notre Dame

Corresponding author – George S. Howard, Department of Psychology, University of Notre Dame, Notre Dame, Indiana 46556

Abstract

There is a dearth of research addressing the validity of life stories, or autobiographies. In part, this is because it is unclear how such data sources might be validated. This article recommends two different perspectives for obtaining evidence relevant to the validity of autobiographical data. A study is conducted from each of these perspectives, and multiple validity estimates are obtained in each investigation. The construct validity of standard psychological constructs (e.g., assertiveness, trustworthiness) obtained in Study 1 from autobiographies was equal to that of standard instruments designed to assess these constructs. Evidence for the validity of life themes, extracted from autobiographies in Study 2, was also apparent. Because life histories, autobiographies, and case studies are relied on heavily in the practice of counseling psychology and are rapidly regaining popularity in scientific studies, these positive findings suggest the wisdom in this increasing use of life history data.

The use of autobiographical materials in psychological research has a long history (Allport, 1942; Murray, 1938; White, 1952). Unfortunately after a golden age of life history research (spanning the 1930s and early 1940s), studies of life stories generally declined for several decades (from the end of World War II into the early 1970s). However, a vigorous renaissance of research on life stories, autobiographies, and psychobiographies has occurred in the past decade and a half (Cocks & Crosby, 1987; Howard, 1989, 1991; McAdams, 1985, 1990; McAdams & Ochberg, 1988; Rabin, Aronoff, Barclay, & Zucker, 1981; Runyan, 1982;

Spence, 1982; Tyler, 1978). (See McAdams, 1985, pp. 19–24, for a more detailed history of research in personology.)

On the other hand, life stories have always been a staple of psychological practice. For example, therapy usually begins with an invitation to a client to tell his or her story. Counselors have favored ways of phrasing their readiness to hear the client's tale, such as "Can you tell me what brings you here?" "How can I be of help to you?" or "What seems to be the problem?" Clients generally know that these invitations do not request the telling of one's complete life story. Rather, they understand that their task is to tell the part of their life story that appears most relevant to their presenting problem. Whereas the life stories of clients often represent the staple data of counseling practice (whether of the personal-emotional or career-vocational sort), life stories have been underutilized in counseling psychology research. If one could demonstrate the validity, viability, or trustworthiness of autobiographical data in some way, it might help to integrate the science and practice functions in the discipline of counseling psychology (see Heppner et al., in press; Howard, 1985, 1986).

With the continuing importance of life stories in counseling practice, and with the resurgent importance of life stories in research, this might be an appropriate point at which to consider the evidence that tests the validity of life stories. Are there ways to transform current validation strategies that will enable researchers to appreciate the validity of life themes extracted from autobiographies? That is, How can validation strategies be modified to obtain empirical evidence for testing the validity of stories (such as life histories, case studies, etc.) in which counselors often place great trust?

There has always been concern about the validity of life histories (e.g., Gottschalk, Kluckhohn, & Angell, 1945; Murray, 1938; Runyan, 1982, 1990), but the record of empirical validation efforts is sparse. Allport, Bruner, and Jandorf (1953) studied life histories of 90 persons who lived through the Nazi experience in Germany. A subset of 10 subjects were interviewed by one of the investigators, and these interview summaries were compared by independent judges to summaries extracted from the life histories. The investigators reported a 95% agreement rate for summaries of the same person from the two independent rating sources. Unfortunately, the report of this research is presented in one brief paragraph, in no greater detail than is described here. Dailey (1971, chapter 6) showed that judges could use fragments of life stories to make predictions about subjects' later life events with greater accuracy than would be expected by chance. An extensive literature search revealed a dearth of literature purporting to assess the validity of life histories (see Allport, 1942, and Rubin, 1986, for exceptions). However, the absence of such a literature is understandable, given the histories of the fields of personality, personology, and measurement theory.

Personology (i.e., research focused on the study of whole lives; cf. Murray, 1938) developed as an outlier, in contrast to the mainstream research in personality, which focuses on personality traits.

The scientific study of the whole person, however, was never well integrated into the mainstream of American psychology in general, nor personality psychology in particular. . . . Personality psychologists became experts in "extraversion" or

the “need for achievement,” but few deemed it worthwhile or profitable to become experts on “persons” (see the reviews provided by Adelson, 1969; Block, 1981; Carlson, 1971, 1975; Helson & Mitchell, 1978; Maddi, 1982; Singer & Singer, 1972). (McAdams, 1985, p. 21)

The split of personology from mainstream personality research was also important because greatly differing notions developed for testing the validity of traits (in personality theory) and the validity of life themes (in personology theory). Of equal importance for the purposes of this article, the two intellectual traditions developed differing notions of the kinds of evidence that might be marshaled in support of the validity of various measurement approaches (self-report questionnaires, behavioral measures, autobiographies, etc.). Finally, because of this split, applied psychologists often have been actively dissuaded from conducting research on intensive case studies. For example, the participants at the Boulder Conference, *Clinical Training in Psychology*, Boulder, Colorado, 1948, doubted that a single case study, no matter how brilliantly executed, could satisfy the requirements of a doctoral dissertation but noted, “However, this question deserves further study before a categorical denial of the suitability of this form of research is made” (Raimy, 1950, p. 89).

Mainstream personality theory developed in conjunction with standard measurement theory (Cook & Campbell, 1979; Cronbach, 1971). The concepts of face validity, content validity, criterion validity, and construct validity became central to both fields and are in fact familiar to most students of psychology. The ways in which such concepts relate to the validation of psychological traits is well documented (e.g., Cook & Campbell, 1979; Cronbach & Meehl, 1955). Yet, how can these concepts be applied to biographical data? One can well imagine the ways in which more rudimentary forms of measurement validity (i.e., face and content validity) might be determined for life stories. What is less clear is how the more compelling forms of measurement adequacy (i.e., criterion and construct validity) relate to life history data.

Study 1

Our ambition in the first study of this article is to evaluate the validity of traditional constructs (i.e., assertiveness, trustworthiness), assessed through autobiographies, and to compare these data with evidence from a set of traditional measurement methodologies developed specifically to assess subjects’ levels of functioning on these traits. If autobiographies yield criterion and construct validity estimates of assertiveness and trustworthiness that are close to the levels obtained with instruments specifically designed to achieve these ends, then one’s confidence in the portrayal of persons in autobiographies would be greatly bolstered. However, advocates of the use of life histories (e.g., Allport, 1942, Howard, 1989, 1991; Mair, 1989; McAdams, 1985; Murray, 1938; Polkinghorne, 1988; Runyan, 1982; White, 1952) might consider this an unfair competition, as autobiographies are not tailored to the assessment of traditional personality constructs.

Criterion and Construct Validity

The literature on test theory and validation draws a clear distinction between criterion validity and construct validity. Cronbach (1971) noted that criterion validation “consists of checking the test score against some other observation that serves as *criterion*. The aim of testing is to predict this criterion, and the merit of the test is judged simply by the accuracy of prediction” (p. 443). Cronbach (1971) also noted that “whenever one classifies situations, persons, or responses, he [or she] uses *constructs*” (p. 462). However, for the vast majority of the constructs dealt with by psychologists, there is no single criterion that measures the construct perfectly. Lyman (1978) underscored and expanded on this fact: “Sometimes, in fact, the test may be a *better* measure of the characteristic than any criterion is” (p. 30). Therefore, a potentially serious problem exists. When an investigator uses a single index as a criterion measure, he or she is involved in criterion validation, which demands that the criterion measure be a perfectly accurate index. However, because the assumption that one index can assess the construct perfectly is almost always untenable, most validation efforts in clinical and personality research involve problems of construct validity.

Campbell and his colleagues (Campbell, 1969; Campbell & Fiske, 1959; Cook & Campbell, 1979) recognized the problems of construct validation and formalized their recommendations in the concept of multiple operationalism. Cook and Campbell (1979) summarized as follows:

We cannot in reality achieve widely accepted definitions of most constructs. This is because propositions about constructs are more reliable if they have been successfully tested not only across many overlapping operational representations of a single definition of a construct, but also across representations of many overlapping definitions of the *same* construct. (pp. 62–63)

The implication is that in most instances multiple operationalizations (methods) are required to obtain adequate construct estimates for construct validation purposes.

A recent series of investigations (Cole, Howard, & Maxwell, 1981; Howard, Conway, & Maxwell, 1985; Howard, Maxwell, Weiner, Boynton, & Rooney, 1980) has demonstrated that the relative validity of various measurement methods (e.g., self-reports, role playings, in vivo behavioral measures, significant-other reports) can be assessed when data is collected according to a multitrait-multimethod matrix format (Campbell & Fiske, 1959). In such studies, the relative criterion validity (mono-operationalized validity coefficients) and the relative construct validity (multiply operationalized validity coefficients) of various alternative measurement strategies can be estimated. In the first study, we seek to estimate the relative validity, in measuring subjects’ levels of assertiveness and trustworthiness, of judges’ ratings of autobiographies, self-report questionnaires, responses to standard vignettes, and significant-other reports. We include a third construct (subjects’ writing ability) as a discriminant variable in this 3 (trait) × 4 (method) model, to estimate both convergent and discriminant validity (cf. Campbell & Fiske, 1959). Of course, we are not assessing “the facts” of a person’s life (e.g., whether a particular event actually occurred) but rather whether a data source (i.e., an autobiography) is an effective instrument in assessing participants’ level of functioning on hypothetical latent traits (i.e., assertiveness, trustworthiness).

Method

Participants

Participants in this study were 22 senior undergraduate psychology majors from a mid-western university who wrote their autobiographies as a class requirement. The consent for use of the material was obtained after the autobiography had been written so that participants had no knowledge that the autobiography was to become part of an investigation. Participants were informed that a study was being conducted in which their autobiography could serve as one of four methods of data collection. All but 1 student agreed to participate, and the participants received course credit for their part in the research. Participants understood that the study would probe the relative validity of several different approaches to assessing psychological traits.

Each participant identified a significant other, who then rated the participant on the three traits of interest. Roommates, close friends, and siblings were chosen as significant others for this study. Six advanced graduate students in counseling psychology served as raters of autobiographies (three raters) and vignettes (three raters).

Procedure

The autobiographies from the undergraduate psychology classes were collected before the data collection session, as per class requirement. The autobiographies were distributed to three raters, after the consent forms were signed and collected at the data collection session. The raters assessed the participants' assertiveness, trustworthiness, and writing ability on the basis of a reading of the participants' autobiographies. The mean rating of the three judges constituted the participants' score on the autobiography measure.

Participants and significant others were instructed to come together to a data collection session, at which time they were given number-coded packets of materials. Participants and significant others were then separated and asked to complete a number of forms. In a few cases, participants asked to take their significant-other packets to the significant other. In addition to autobiographies, participants gave responses to vignettes (described below) designed to elicit both trustworthy and assertive behaviors (and provided a sample of their writing to be judged later for participants' writing ability). When all of these data had been received, responses to the vignettes were distributed to the other three raters, who then completed the rating tasks. Each judge read the participants' responses to the vignettes and scored them for assertiveness, trustworthiness, and writing ability. The average across judges on each trait constituted the vignette score.

Measures

In this study, each of three traits (assertiveness, trustworthiness, and writing ability) was measured by four methods (self-report, significant-other report, autobiography, and vignette).

Autobiographical rating of trustworthiness. Participants were given instructions to write a detailed story of their lives, highlighting what seemed to be significant events and themes at the time of writing. The autobiographies ranged in length from 6 to 19 typed

pages ($M = 13$ pages). Three raters read and rated the autobiographies for trustworthiness on a scale ranging from *extremely untrustworthy* (1) to *extremely trustworthy* (5).

Pencil-and-paper measures of trustworthiness: Self-report and significant-other report. We selected Rotter's (1967) Interpersonal Trust Scale (ITS) for use in this study as a self-report measure of trustworthiness. The ITS is a 99-item questionnaire in which participants respond to brief statements on a scale ranging from *strongly agree* (1) to *strongly disagree* (5). We developed a modified version of this measure by altering the wording of the scale items to facilitate rating of the participant by a significant other (the significant-other report measure). Rotter (1971) reported a split-half reliability estimate of .76 for the original scale.

Rated behavior samples of trustworthiness: Vignettes. Written contrived situation vignettes were rated on the basis of Barak and LaCrosse's (1976) criteria developed for assessing trustworthiness in counselor behavior, as judged by college undergraduates. Although the situations in the vignettes used in this article are different from those researched by Barak and LaCrosse, the criteria Barak and LaCrosse used provided clear parameters for assessing this construct. The situations used in the present study involved noncontroversial situations that typically might be encountered by college or postcollege students. Three situations designed to place the participant in a situation calling on his or her trustworthiness were written for this study. Participants responded to these vignettes by writing 1 page or less about how they would respond to the situation. Three raters evaluated these essay responses to the vignettes for trustworthiness.

Autobiographical rating of assertiveness. The same raters who read and rated the autobiographies for trustworthiness rated the autobiographies for participants' assertiveness on a scale ranging from *extremely unassertive* (1) to *extremely assertive* (5).

Pencil-and-paper measures of assertiveness: Self-report and significant-other report. For assertiveness, we used the College Self-Expression Scale (CSE) as the self-report and significant-other report instruments. The CSE is a 96-item questionnaire in which participants respond to brief statements on a scale ranging from *strongly agree* (1) to *strongly disagree* (5). The significant-other report reflected appropriate grammatical changes to the CSE questions. A rate-rerate reliability value of .89 for the CSE was reported by Galassi, Delo, Galassi, and Bastien (1974).

Rated behavior samples of assertiveness: Vignettes. For the contrived-situation vignette ratings, we used criteria for rating assertiveness situations from studies by Eisler, Hersen, and Miller (1973). In these studies, observers rated subjects' responses to several videotaped scenarios. Raters were trained to use these criteria as guidelines for their rating of assertiveness in the written vignettes. The situations used in the present study involved noncontroversial situations that typically might be encountered by college or postcollege students. Three situations placing the participant in a situation calling on his or her assertiveness were written. As with the measure of trustworthiness, we constructed three situations for this study that college-age students might actually encounter, each of which called for an assertive response.

Autobiographical rating of writing ability. The set of raters who read and rated the autobiographies for trustworthiness and assertiveness also rated the autobiographies for

participants' writing ability on a scale ranging from *extremely poor writer* (1) to *extremely good writer* (5).

Pencil-and-paper measures of writing ability: Self-report and significant-other report. Because writing ability was used as a discriminant variable, which likely had little or no correlation with the other constructs, a simple question was added to the self-report and other-report instruments: "In general, considering sentence structure, grammar, punctuation, and ability to express thoughts clearly, I am [he is/she is] a good writer." To report their perceived writing ability, participants and significant others used a 5-point scale, ranging from *poor writer* (1) to *superior writer* (5).

Rated behavior samples of writing ability: Vignettes. After reading each participant's essay response to the vignettes, each judge rated the participant's writing ability on the 5-point scale described in the previous paragraph. Judges who read autobiographies also rated participants' writing abilities on the same 5-point scale.

Results

Table 1 presents the raw data (means and standard deviations) for the three traits of interest, as assessed by the four measurement methods.

| Measure | Assertiveness | | Trustworthiness | | Writing ability | |
|--------------------------|---------------|-----------|-----------------|-----------|-----------------|-----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Self-report | 3.50 | 0.37 | 3.08 | 0.33 | 4.14 | 0.99 |
| Significant-other report | 3.40 | 0.52 | 3.01 | 0.38 | 4.40 | 0.50 |
| Autobiography | 3.20 | 0.83 | 3.80 | 0.43 | 3.76 | 0.56 |
| Vignettes | 3.05 | 0.51 | 3.29 | 0.55 | 3.20 | 0.69 |

Note: All ratings were made on a 5-point scale, with a score of 5 indicating the greatest levels of assertiveness, trustworthiness, and writing ability.

Convergent Validity

Convergent validity is demonstrated by the extent to which the different methods of measuring a trait are correlated with each other. We used two methods of demonstrating convergent validity: mono-operationalization and multiple-operationalization. The data used in the analyses of convergent and discriminant validity largely derive from the intercorrelation matrix presented in table 2.

Table 2. Intercorrelations among Three Traits Measured by Four Methods

| Trait/ measure | Assertiveness | | | | Trustworthiness | | | | Writing ability | | | |
|-------------------|---------------|-------|------|------|-----------------|------|------|------|-----------------|-----|------|-----|
| | SR | SOR | Auto | Vig | SR | SOR | Auto | Vig | SR | SOR | Auto | Vig |
| Assertiveness | | | | | | | | | | | | |
| SR | — | | | | | | | | | | | |
| SOR | .69** | — | | | | | | | | | | |
| Auto | .52** | .60** | — | | | | | | | | | |
| Vig | .04 | .00 | -.11 | — | | | | | | | | |
| Trustworthiness | | | | | | | | | | | | |
| SR | -.11 | -.31 | .05 | .03 | — | | | | | | | |
| SOR | .38 | .39* | .03 | .19 | .04 | — | | | | | | |
| Auto | .22 | .26 | .34 | -.04 | .17 | .44* | — | | | | | |
| Vig | -.04 | .02 | .26 | .47* | .19 | .34 | .34 | — | | | | |
| Writing ability | | | | | | | | | | | | |
| SR | .60* | .27 | .24 | -.04 | -.20 | .00 | -.06 | -.05 | — | | | |
| SOR | .51 | .23 | .48* | .15 | .31 | .14 | .25 | .19 | .36 | — | | |
| Auto | .29 | .27 | .62 | -.10 | .18 | .01 | .18 | .34 | .32 | .31 | — | |
| Vig | .06 | -.04 | .22 | .25 | .40 | -.23 | .14 | .45 | .14 | .43 | .24 | — |

Note: SR = self-report; SOR = significant-other report; Auto = autobiography; Vig = vignettes. $N = 22$. Higher correlations between different methods of measuring the same trait suggest evidence of convergent validity. * $p < .05$. ** $p < .01$.

Mono-operationalization validation. Each method of measuring the traits of assertiveness, trustworthiness, and writing ability was used to validate every other method within that trait construct. The average validity coefficient for each measurement method was calculated from the correlation matrix by averaging the correlations of that method with the three other methods of measuring that same trait. This gives a single estimate of the convergence of each method with every other method. The results of this analysis are presented in the mono-operationalization columns of table 3. The results of autobiography ratings are encouraging, in comparison with all other methods of measuring the three traits. The mean of the validity coefficients across all traits shows autobiography to be at least as effective a method (with an average validity coefficient of .32) as self-report (.27) and other report (.36). The vignettes seemed to be inferior to the other three measurement methods.

Table 3. Estimates of Criterion (Mono-Operationalization) and Construct (Multiple-Operationalization) Validity of Four Methods for Three Traits

| Measure | Mono-operationalization | | | | Multiple-operationalization | | | |
|---------|-------------------------|-------|-------|-----|-----------------------------|-------|-------|-----|
| | Assert | Trust | Write | M | Assert | Trust | Write | M |
| Auto | .34 | .32 | .29 | .32 | .62 | .44 | .40 | .49 |
| Vig | -.02 | .29 | .27 | .18 | .02 | .38 | .32 | .24 |
| SR | .42 | .13 | .27 | .27 | .64 | .06 | .80 | .50 |
| SOR | .43 | .27 | .37 | .36 | .70 | .09 | .52 | .44 |

Note: Assert = assertiveness; Trust = trustworthiness; Write = writing ability; Auto = autobiography; Vig = vignette; SR = self-report; SOR = significant-other report. Higher correlations suggest evidence of convergent validity.

Multiple-operationalization validation. In the multiple-operationalization approach, every method is correlated with a different composite criterion. This composite criterion was computed by forming an unweighted linear combination of the standardized scores for the remaining three measures of the construct. In this manner every method's contribution to the criterion is equal, and the effects of method variance and error variance are minimized, in contrast with the mono-operationalized approach. Thus, no method is presumed to have more or less value as a criterion than any other method. In carrying out these analyses, we standardized scores to prevent scaling differences from biasing the mean calculations. The results are reported in the multiple-operationalization columns of table 3.

One can see that the average validity coefficients obtained with this method are generally improved over those obtained with the mono-operationalization approach (with the exception of the self-report measure of trustworthiness). Because the multiple-operationalization approach averages over method variance and random error variance, the random error variance is canceled out, leaving only the effects of true (trait) variance and method variance (as opposed to the mono-operationalization approach, in which the validity coefficients are distorted by both random error variance and method variance). These results confirmed that autobiography ratings were comparable to self-report and other report measures, whereas the validity estimates for the vignettes again lagged behind the other three measurement methods.

Discriminant Validity

Discriminant validity assesses the extent to which a method differentiates between distinct traits. The goal of discriminant validity is to find evidence of divergence between different constructs (e.g., assertiveness and writing ability) when measured by the same method (e.g., self-report). Discriminant validity can be understood as a comparison of methods, through removal of the method variance that may have been present in the data on convergent validity. According to Campbell and Fiske (1959), there are three criteria for testing discriminant validity: (a) monotrait, heteromethod correlations should be higher than different heterotrait, monomethod correlations; (b) monotrait, heteromethod correlations should be greater than heterotrait, heteromethod correlations; and (c) the same pattern of trait interrelationships should be demonstrated in all of the heterotrait correlations. These

rules of thumb are cumbersome to assess, and only a summary is reported here. (See table 2 for the full Trait \times Method intercorrelation matrix). Divergence, as shown by relatively small correlations between similar measures of separate constructs (e.g., the construct of interest and a discriminant variable), may be construed as evidence that the construct of interest was uniquely measured (Cook & Campbell, 1979).

The median value for heterotrait, monomethod correlations across all traits was .295. Only the median correlations among the various measures of writing ability was higher (.315). However, similar comparisons for the trustworthiness (.265) and assertiveness (.280) constructs were sufficiently close to the .295 critical value to warrant examination by the second criterion.

The more stringent second criterion was easily met, as the median heterotrait, heteromethod correlation was .04, which was clearly smaller than the monotrait, heteromethod medians reported above (.280, .265, and .315).

Although it is less clear exactly how to test the third criterion, there do appear to be similar patterns of relationships for the four measurement methods. Taken together, the convergent and discriminant validity data support the conclusion that rated autobiography is a valid method of measuring these particular traits.

Study 2

The strong evidence of criterion and construct validity, for estimates of various traits obtained from autobiographies, suggests that accurate pictures of persons might be obtained from such intensive case-study materials. Although personologists and practitioners might applaud this research support for the validity of autobiographical life histories, they would hasten to add that such efforts fail to capitalize on one of the most important aspects (or insights) that might be gleaned from life stories. Namely, one can gain an appreciation of a person's life theme from an autobiography. Yet, how is one to assess the construct validity of a life theme? It seems unlikely that one can treat life themes as if they were constructs; thus, the multitrait-multimethod approach used in Study 1 would be of little help. The reason for this is that construct validation strategies demand that differentiations be made among participants in the *quantity* of the trait possessed (e.g., Subject A's IQ is 90, whereas Subject B's IQ is 110; Subject A is more assertive than Subject B, who, in turn, is more assertive than Subject C). Yet, one is hard-pressed to justify a single overarching dimension along which to rank order or quantitatively rate life themes. Life themes exhibit differences in focus or orientation, which make them noncomparable in any way other than an arbitrary or trivial manner. That is, life themes exhibit differences in *kind* rather than differences in *quantity*. Therefore, the most basic premise of construct validation strategies appears to be unsatisfied in our efforts to assess the validity of life themes. However, one can obtain a sense for the convergent and discriminant validity of life themes in these autobiographies, by developing alternative methodologies that test important notions of measurement accuracy in a slightly different manner.

In this second study, evidence for convergent validity of life themes extracted from the autobiographies is obtained by considering agreement among various types of judges possessing vastly different bodies of information about the research participant on which to

judge the life theme's adequacy. Evidence for discriminant validity can be seen in each judge's ability to endorse the life theme construed from each participant's autobiography, while at the same time rejecting life themes gleaned from the life stories written by the other participants. Thus, we develop slightly modified procedures for obtaining evidence of convergent and discriminant validity to test the measurement adequacy of life themes obtained from autobiographies.

Method

Participants and Procedure

Participants were 20 undergraduate students from a midwestern university who received course credit for participating in this study. Sixteen participants participated in both Study 1 and Study 2. Participants were first asked to write their autobiography. Each autobiography ranged from 6 to 19 pages in length ($M = 14$ pages).

Next, one of a panel of four summarizers (the authors) read each of the autobiographies (which were identified by participant number only) and wrote summaries of the original autobiographies in the third person. These summaries were either one or two paragraphs in length. Summaries were written so as to represent the character of the participant and to edit out specific, factual pieces of information that might give away the identity of the participant (i.e., gender, place or residence, death of family member, divorce, etc., represent typical information that was deleted from the summaries). Thus, the summaries contained descriptions of personality characteristics, general issues about which the participant expressed concern (e.g., "this person struggles with religious issues, and he [she] is very competitive"), and the summarizer's sense of the subject's life theme.

When all of the summaries were complete, we constructed the packets. Each packet contained a designated target summary and three randomly selected foil summaries (one foil from each of the three summarizers who did not summarize that target participant's autobiography). The participants were then asked to rate each summary in his or her packet on a 10-point scale, on which a value of 1 represented no belief that the summary was of his or her autobiography, as it did not seem to grasp the participant's life theme, and a value of 10 represented total confidence that the summary was of his or her autobiography because it grasped his or her life theme perfectly. This constituted the "participant" measure of the summaries, as it was made by each research participant himself or herself. Again, every participant rated four summaries: one (the target) was a summary of his or her own autobiography and three others (the foils) were summaries of the other participants' autobiographies.

In designing this study, two threats to the validity of this method came to mind, and these challenges need to be addressed before one can have confidence in this validation method. First, one needs to rule out the possibility that the participant was somehow matching specific information from the text of the autobiography to information in the summary (e.g., key words or expressions). To address this issue, a second rating procedure was conducted in which another rater (a significant other), who stated that he or she had not read the autobiography, rated four summaries on the basis of what he or she knew about the target participant and his or her life. The same packet that had been rated by the

participant (the target summary and the same three foils) was also rated by the significant other on a scale ranging from *not like my friend* (1) to *exactly like my friend* (10). The significant-other rating took place within 1 week after the participant furnished the name of the significant other to the researchers.

The second threat to the validity of the autobiography/summary method was the possibility that information about the participant's life that was not contained in the autobiography but was known by other means could be used in rating the summaries. To address this objection, we conducted a third rating procedure in which a new set of raters (three advanced graduate students who were not involved in generating the summaries and who knew nothing about the target subject) read the target's autobiography and rated the same packet of target and foil summaries (described above) as had the participant and the significant other.

To summarize, in the second rating procedure, the 20 participants each identified a significant other (typically a classmate or a roommate), who was then recruited to participate in the study. The significant other was shown the same four summaries that their friend (the target participant) rated. The significant other was also asked to rate his or her confidence that each summary represented his or her friend. The significant others asserted that they had not read the original autobiographies, they simply compared the summaries to their experiences and knowledge of their friend.

Finally, in the third rating procedure, three advanced graduate students who did not know any of the research participants were asked to read about one third of the 20 autobiographies. Attached to each autobiography were the same three randomly selected foil summaries, along with one target summary. Thus, judges rated the summaries with no knowledge of the subject, except that gained from reading the autobiography.

Results

Differences between mean endorsement of the target summary versus the foil summaries for each rater group (participant, significant other, and judge) are shown in table 4. Judges evidenced the greatest ability to separate the target summaries ($M = 8.57$) from the foil summaries ($M = 2.27$). The difference between targets and foils was statistically reliable for judges, $t(19) = 20.50, p = .001$; also, for none of the 20 subjects was the judges' mean rating of the target summary less than or equal to their mean rating of the foils. The participants also demonstrated a striking ability to discriminate target summaries ($M = 8.90$) from foil summaries ($M = 3.58$). This difference was also statistically reliable, $t(19) = 7.74, p = .001$, and in only one instance (of 20 participants) was the mean foil rating greater than or equal to the target rating. Finally, significant others were also able to discriminate target summaries ($M = 7.60$) from foils ($M = 3.93$) reliably, $t(19) = 4.62, p = .001$. However, in 5 cases (of 20), significant others rated the foils equal to or higher than the target summaries.

Table 4. Mean Endorsement of Target and Foil Summaries by Three Types of Raters

| Rater | Target | | Foil | |
|-------------------|----------|-----------|----------|-----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Participant | 8.90 | 2.04 | 3.58 | 1.81 |
| Significant other | 7.60 | 2.83 | 3.93 | 1.46 |
| Judge rating | 8.57 | 1.07 | 2.27 | 0.76 |

Note: All ratings were made on a 10-point scale, with a score of 10 indicating a summary that accurately depicts the participant's characteristics and life theme.

We performed a two-way split-plot multivariate analysis of variance to compare the relative adequacy of the rating techniques (participants, judges, significant others) in discriminating target from foil summaries. The dependent measures in this analysis were the four ratings made by the participants, significant others, and judges of the four summaries. There was a significant difference among the three methods in ability to discriminate targets from foils, $F(2, 18) = 5.92, p = .011$. Significant others showed reliably less ability to discriminate between targets and foils than did the judges, $F(1, 19) = 11.98, p = .003$ and significantly less discriminating ability than the participants themselves, $F(1, 19) = 6.69, p = .018$. However, judges and participants were not reliably different from one another on this dimension, $F(1, 19) = 2.29, p = .147$.

General Discussion

Perhaps the most important information people convey when they write their life history, or autobiography, is their life theme. Each person attempts to depict the core issues and values that make his or her life an integrated whole. If identity is formed by weaving the "matter" of one's experience into a life story (see Bruner, 1986; Gergen & Gergen, 1983, 1987; Howard, 1989; Mair, 1989; Mancuso & Sarbin, 1983; McAdams, 1985; Polkinghorne, 1988; Runyan, 1982; Sarbin, 1986), then are researchers not obligated to view autobiography as a rich subject matter for psychological investigations? What constitutes the empirical data for counseling and psychotherapy, if not clients' life stories?

What is the status of autobiography as scientific evidence? Dilthey (as cited in Rickman, 1961) stated that, "Autobiography is the highest and most instructive form in which the understanding of life confronts us" (p. 85). However, as indicated in the introduction of this article, empirical support of the construct validity of autobiographies is meager. In Study 1, we found that judges' ratings of participants' psychological constructs from autobiographies (a task for which autobiographies are not designed) was as good as (and sometimes better than) standard assessment techniques (e.g., self-report questionnaires, ratings of vignettes, significant-other reports) that were specifically developed to assess participants' levels of functioning on the target constructs. This finding should be very encouraging to advocates of autobiographies.

However, autobiographies are tailored toward revealing an understanding of a construct that questionnaires, role playings, behavioral measures, projective measures, and so

forth are hard-pressed to accurately assess, namely, a person's life theme. In his psychobiographic research, Sartre (as cited in Charne, 1984) found that a singular, holistic structure of meaning emerged. Sartre termed this unifying theme a person's "fundamental project." Similarly, McAdams (1985) found evidence of coherent life themes (organized around issues of power and intimacy) in his study of autobiographical materials. Finally, Csikszentmihalyi and Beattie (1979) found life themes to be an affective and cognitive system composed of a central existential problem, with possible causes and solutions. In Study 2 we sought to extract such life themes from autobiographies and to assess the participant's, judges', and significant other's ability to recognize the subject's life theme on the basis of firsthand knowledge of the person only (significant other), the autobiography only (the judges), or both (the participant).

Participants were perfectly able (with 100% accuracy) to discriminate their life theme from the life themes of other participants. Furthermore, after reading only the autobiography, raters could identify the proper life theme from foil summaries with an amazing 95% accuracy. Although significant others were less accurate (75% accurate) in identifying target themes than the participants and the raters, this might have resulted from some of the particulars of this study. Because the study was conducted at a university that is almost completely residential, participants were unable (in most cases) to involve lifelong significant others. For pragmatic reasons, most subjects nominated classmates to serve as significant others. Thus, the significant others' experiences with the autobiographer tended to be time limited and focused heavily on academic experiences. Had subjects been able to involve significant others who had shared longer and more varied experiences with them, the significant-other accuracy rate might have approached the remarkable rates of the participants and the judges.

An ever-growing number of psychologists are using life history data as the foundation of their scientific and applied efforts. The current dearth of empirical evidence to support our trust in autobiographical data motivated the present studies. Using two very different measurement validation strategies, we found good support for the criterion and construct validity of traditional personality constructs obtained from autobiographies (Study 1) and we obtained strong support for the validity of life theme summaries from autobiographies (Study 2). Successful replication of these findings will enable a wide range of psychologists to have greater confidence in the construct validity of autobiographical materials in their scientific and practice efforts. Finally, with continued research on life stories, psychologists will come to a richer understanding of the special characteristics of autobiographical data that render such evidence more useful than data collected through standard instrumentation. For example, autobiographical data locate the place and importance of a trait in a person's life history and life project. Knowledge of the specifics of how a person became, for example, anxious, trustworthy, or assertive will aid the counselor in enlisting that trait (and the past experiences) in forming a plan of action for the counselee.

A preliminary effort was made to extend current construct validation methodologies and perspectives to a consideration of entities (such as life themes) that possess differences in kind (or quality) but that cannot be quantitatively assessed or rank ordered in nonarbitrary ways. By capitalizing on the ability of various groups of raters (participants, significant others, and judges) to recognize qualitative differences in life themes, we extended

the notions of convergent and discriminant validity and applied these to entities for which they were formerly ill suited. With these methods, the measurement adequacy of explanatory concepts (such as life themes) and data sources (such as autobiographies), which are becoming increasingly popular of late, can now be more sensitively and appropriately assessed.

References

- Allport, G. W. (1942). *The use of personal documents in psychological science*. New York: Social Science Research Council.
- Allport, G. W., Bruner, J. S., & Jandorf, E. M. (1953). Personality under social catastrophe: Ninety life-histories of the Nazi revolution. In C. Kluckhohn and H. A. Murray (Eds.), *Personality in nature, society, and culture* (2nd ed., pp. 436–455). New York: Knopf.
- Barak, A., & LaCrosse, M. B. (1976). Multidimensional perception of counselor behavior. *Journal of Counseling Psychology, 22*, 471–476.
- Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.
- Campbell, D. T. (1969). Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), *Psychology: A study of a science*. (Vol. 6, pp. 94–172). New York: McGraw-Hill.
- Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.
- Charne, S. L. (1984). *Meaning and myth in the study of lives: A Sartrean perspective*. Philadelphia: University of Pennsylvania Press.
- Cocks, G., & Crosby, T. L. (Eds.). (1987). *Psychohistory: Readings in the method of psychology, psychoanalysis, and history*. New Haven, CT: Yale University Press.
- Cole, D. A., Howard, G. S., & Maxwell, S. E. (1981). Effects of mono- versus multiple-operationalization in construct validation efforts. *Journal of Consulting and Clinical Psychology, 49*, 395–405.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 443–507). Washington, DC: American Council on Education.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 55*, 281–302.
- Csikszentmihalyi, M., & Beattie, O. (1979). Life themes. *Journal of Humanistic Psychology, 19*, 45–63.
- Dailey, C. A. (1971). *Assessment of lives*. San Francisco: Jossey-Bass.
- Eisler, R. M., Hersen, M., & Miller, P. M. (1973). Components of assertive behavior. *Journal of Clinical Psychology, 29*, 295–299.
- Galassi, J., Delo, J., Galassi, M., & Bastien, S. (1974). The college self-expression scale: A measure of assertiveness. *Behavior Therapist, 5*, 165–175.
- Gergen, K. J., & Gergen, M. M. (1983). Narratives of the self. In T. R. Sarbin & K. E. Scheibe (Eds.), *Studies in social identity* (pp. 254–273). New York: Praeger.
- Gergen, K. J., & Gergen, M. M. (1987). The self in temporal perspective. In R. P. Abeles (Ed.), *Lifespan perspectives in social psychology* (pp. 121–137). Hillsdale, NJ: Erlbaum.
- Gottschalk, L., Kluckhohn, C., & Angell, R. (1945). *The use of personal documents in history, anthropology, and sociology*. New York: Social Science Research Council.

- Heppner, P. P., Carter, J., Claiborn, C. D., Brooks, L., Gelso, C. J., Fassinger, R. E., Holloway, E. L., Stone, G. C., Wampold, B. E., & Galassi, J. P. (in press). A proposal to integrate science and practice in counseling psychology. *The Counseling Psychologist*.
- Howard, G. S. (1985). Can research in the human sciences become more relevant to practice? *Journal of Counseling and Development*, 63, 539–544.
- Howard, G. S. (1986). The scientist-practitioner model in counseling psychology: Toward a deeper integration of theory, research, and practice. *The Counseling Psychologist*, 14, 61–105.
- Howard, G. S. (1989). *A tale of two stories: Excursions into a narrative approach to psychology*. Notre Dame, IN: Academic Publications.
- Howard, G. S. (1991). Culture tales: A narrative approach to thinking, cross-cultural psychology and psychotherapy. *American Psychologist*, 46, 188–197.
- Howard, G. S., Conway, C. G., & Maxwell, S. E. (1985). Construct validity of measures of college teaching effectiveness. *Journal of Educational Psychology*, 77, 187–196.
- Howard, G. S., Maxwell, S. E., Weiner, R. L., Boynton, K. S., & Rooney, W. M. (1980). Is a behavioral measure the best measure of behavioral parameters? Perhaps not. *Applied Psychological Measurement*, 4, 293–311.
- Lyman, H. B. (1978). *Test scores and what they mean* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Mair, M. (1989). *Between psychology and psychotherapy*. New York: Routledge, Chapman & Hall.
- Mancuso, J. C., & Sarbin, T. R. (1983). The self-narrative in the enactment of roles. In T. R. Sarbin & K. E. Scheibe (Eds.), *Studies in social identity* (pp. 233–253). New York: Praeger.
- McAdams, D. P. (1985). *Power, intimacy, and the life-story: Personological inquiries into identity*. Homewood, IL: Dorsey Press.
- McAdams, D. P. (1990). Unity and purpose in human lives: The emergence of identity as a life story. In A. I. Rabin, R. A. Zucher, R. A. Emmons, & S. Frank (Eds.), *Studying persons and lives* (pp. 148–200). New York: Springer.
- McAdams, D. P., & Ochberg, R. (Eds.). (1988). Psychobiography and life narratives [Special issue]. *Journal of Personality*, 56(1).
- Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.
- Polkinghorne, D. (1988). *Narrative knowing and the human sciences*. Albany, NY: State University of New York Press.
- Rabin, A. I., Aronoff, J., Barclay, A. M., & Zucker, R. A. (Eds.). (1981). *Further explorations in personality*. New York: Wiley.
- Raimy, V. C. (Ed.). (1950). *Training in clinical psychology (Boulder Conference)*. Englewood Cliffs, NJ: Prentice-Hall.
- Rickman, H. P. (Ed.). (1961). *Meaning in history: W. Dilthey's thoughts on history and psychology*. Winchester, MA: Allen & Unwin.
- Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35, 651–665.
- Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, 26, 443–452.
- Rubin, D. C. (1986). *Autobiographical memory*. Cambridge, England: Cambridge University Press.
- Runyan, W. M. (1982). *Life histories and psychobiography: Explorations in theory and method*. New York: Oxford University Press.

- Runyan, W. M. (1990). Individual lives and the structure of personality psychology. In A. I. Rabin, R. A. Zucher, R. A. Emmons, & S. Frank (Eds.), *Studying persons and lives* (pp. 10–40). New York: Springer.
- Sarbin, T. R. (Ed.). (1986). *Narrative psychology: The storied nature of human conduct*. New York: Praeger.
- Spence, D. P. (1982). *Narrative truth and historical truth*. New York: Norton.
- Tyler, L. B. (1978). *Individuality: Human possibilities and personal choice in the psychological development of men and women*. San Francisco: Jossey-Bass.
- White, R. W. (1952). *Lives in progress*. New York: Holt.