

- Testa, M. A., & Simonson, D. C. (1996). Assessment of quality-of-life outcomes. *New England Journal of Medicine*, *334*, 835-840.
- UNAIDS. (1998). *Epidemiological fact sheet on HIV/AIDS and sexually transmitted diseases*. Geneva: UNAIDS/WHO Working Group on Global HIV/AIDS and STD Surveillance.
- Wade, S. H. (1989). *The development of a scale to measure forgiveness*. Unpublished doctoral dissertation, Fuller Graduate School of Psychology, Pasadena, CA.
- WHOQOL Group. (1993). Study protocol for the World Health Organization project to develop a Quality of Life Assessment Instrument (WHOQOL). *Quality of Life Research*, *2*, 153-159.
- WHOQOL Group. (1994). The development of the World Health Organization Quality of Life Assessment Instrument (the WHOQOL). In J. Orley & W. Kuyken (Eds.), *Quality of life assessment: International perspectives* (pp. 41-57). Berlin: Springer-Verlag.
- WHOQOL Group. (1995). The World Health Organization Quality of Life Assessment (WHOQOL): Position paper from the World Health Organization. *Social Science and Medicine*, *41*, 1403-1409.
- WHO/Division of Mental Health/Temoshok, L. R. (1997). *Report on the consultation regarding quality of life and HIV/AIDS*. Geneva: WHO.
- WHO/UNAIDS. (1998). *Report on the global HIV/AIDS epidemic, June 1998*. Geneva: WHO Working Group on Global HIV/AIDS and STD Surveillance.
- Winiarski, M. G. (1991). Spiritual issues in AIDS-related psychotherapy. In M. G. Winiarski (Ed.), *AIDS-related psychotherapy* (pp. 183-187). New York: Pergamon Press.
- Zich, J., & Temoshok, L. (1987). Perceptions of social support in persons with AIDS and ARC. *Journal of Applied Social Psychology*, *17*, 193-215.

CHAPTER 4

What We Know (and Need to Know) about Assessing Forgiveness Constructs

Michael E. McCullough, William T. Hoyt,
and K. Chris Rachal

Improving the measures available for assessing forgiveness is one of the most important tasks necessary for creating a sustainable future for the psychology of forgiveness (McCullough & Worthington, 1995; McCullough, Worthington, & Rachal, 1997; McCullough, Sandage, Rachal, & Worthington, 1997). Currently, many aspects of forgiveness appear to be adequately assessed by existing measures. However, many aspects of forgiveness still cannot be studied properly because instruments for assessing them have not yet been developed. As well, many questions remain about how to use the existing measures in ways that optimize their performance.

In the present chapter, we set out to review what we know about assessing forgiveness-related constructs. As well, we address several of the most pressing issues in future psychometric work on forgiveness-related constructs.

A 3 × 2 × 4 FRAMEWORK

We have developed a taxonomy for categorizing the existent *prima facie* measures of forgiveness. All of the measurements included in this taxon-

omy consider the individual as the unit of analysis, although forgiveness might also be observed and assessed at the level of the dyad, the family, the neighborhood, or larger population aggregates. Such considerations, however, are beyond the scope of this chapter.

Our $3 \times 2 \times 4$ taxonomy categorizes the available instruments along three dimensions. The first dimension refers to the level of *specificity* with which forgiveness is assessed. McCullough and Worthington (1999) classified the existing measures of forgiveness into three levels of specificity; offense-specific, dyadic, and dispositional.

Offense-specific measures are assessments of the extent to which a person has forgiven a specific offender for a specific offense. Ostensibly, an offense-specific measure of forgiveness would represent the extent to which a person has forgiven (or sought forgiveness) in the context of a single, circumscribed interpersonal offense. *Dyadic* measures of forgiveness ostensibly represent an aggregate of the extent to which an individual forgives (or seeks forgiveness) in a single relationship across multiple offenses. Such dyadic measures are more general than offense-specific measures, as they are aggregating (at least in theory) people's forgiveness responses across many offenses occurring within a single relationship. In a sense, dyadic measures loosely might be considered as a sort of weighted mean of a person's offense-specific forgiveness responses summed across multiple offenses within a single relationship. *Dispositional* measures of forgiveness ostensibly represent a person's tendency to grant (or seek) forgiveness across a variety of interpersonal offenses occurring in a variety of relationships. Thus, dispositional measures of forgiveness represent (at least in theory) a sort of weighted mean of a person's offense-specific forgiveness responses summed across multiple offenses and multiple relationships.

The second dimension along which existing forgiveness measures may be categorized is their *direction* of measurement. Most measures assess forgiveness in the direction of granting forgiveness (i.e., from the perspective of the forgiver). A few others also measure forgiveness in the direction of seeking or accepting forgiveness (i.e., from the perspective of the transgressor). Because fairly little research has examined the contours of seeking or accepting forgiveness from others (Gassin, 1998; Meek, Allbright, & McMinn, 1995), the measurement of forgiveness from the perspective of the person who seeks or accepts forgiveness is similarly undeveloped.

A third dimension on which the existing measures of forgiveness may be categorized refers to the *method* with which forgiveness is assessed. Offense-specific forgiveness (i.e., both granting forgiveness and receiving forgiveness) might ostensibly be assessed through at least four methods. First, using self-report methods, an offended person can report the extent to which he or she has forgiven the offending partner (or the offender can

report the extent to which he or she has sought forgiveness from the offended partner). Second, using partner-report methods, the offending relationship partner can report the extent to which the offended relationship partner has granted forgiveness (or the offended relationship partner can report the extent to which the offending partner appears to feel forgiven). Third, an outside observer (e.g., a clinician or other third party) can assess the extent to which a partner has forgiven the offending relationship partner (or appears to have sought the offended partner's forgiveness). Fourth, measures of constructive or destructive behaviors toward an offending relationship partner, which do not rely on verbal or written reports, can be used to infer the extent to which an offended partner has forgiven an offending relationship partner, or to which the offender has sought the forgiveness of the offended partner.¹ In the section of the chapter that follows, we use the $3 \times 2 \times 4$ taxonomy to review the existing measures for assessing forgiveness constructs.

A REVIEW OF THE EXISTING MEASURES OF FORGIVENESS

Offense-Specific Measures of Forgiveness

At the offense-specific level, a measure of forgiveness assesses the extent to which a person has forgiven (or sought forgiveness) for a single interpersonal offense. Offense-specific measures are used to answer questions such as, "Did this person forgive her father for abandoning her and her mother?" or "Did this person seek forgiveness from his wife for his extramarital affair?" Most of the existing measures of forgiveness have assessed forgiveness at this level of specificity (McCullough & Worthington, 1999).

Offense-Specific Measures of Granting Forgiveness

Self-Report Measures. For nearly 20 years, social psychologists have assessed offense-specific granting of forgiveness with single-item self-report measures (e.g., Darby & Schlenker, 1982). Even today, correlation with such a face-valid, single-item measure of forgiveness remains something of a litmus test for making the case for the validity of a more complex measure (e.g., Subkoviak et al., 1995; McCullough et al., 1998).

As early as 1981, researchers also began to develop offense-specific, multi-item measures that assessed granting forgiveness by self-report. Trainer (1981) developed a nine-item self-report measure of offense-specific forgiveness that she called "general forgiveness" (i.e., absence of hostility, grudge-holding; presence of positive feelings and hopes for the offender's well-being). Along with this general measure of offense-specific forgiving, she also developed three self-report measures that ostensibly

measure motivations for forgiving: intrinsic motivation (14 items), expedient motivation (10 items), and role-expected motivation (10 items). The four scales manifested internal consistency reliabilities ranging from .77 to .89 (Trainer, 1981). Park and Enright (1997) used seven of the nine items from Trainer's (1981) general forgiveness scale, added three new items, and used the 10-item total score (with an internal consistency reliability of $\alpha = .87$) as a measure of general forgiveness.

In Trainer's (1981) work, the four forgiveness subscales were correlated with the extent to which offended persons blamed their offenders, experienced empathic attributions for their offenders' behavior, and maintained bitter/hostile feelings toward the offender. Moreover, there were important differences in the correlates of the four subscales, suggesting that they might be assessing different aspects of forgiveness.

However, factor analyses disputed the validity of a four-factor model for describing the interrelationships of the items on these four scales. Most of the general forgiveness and intrinsic items loaded on a first factor that explained approximately 18% of the total item variance. This factor seemed to represent forgiveness most adequately (indeed, we can be comforted that the item that loaded most strongly with this first factor was a single item assessing the extent to which the respondent had forgiven his or her offender). A second factor seemed to represent an angry, pragmatic, unempathic view of forgiveness that reflected respondents' use of forgiveness as a way of getting revenge or propping up a damaged self-image. A third factor appeared to represent a form of forgiveness that was motivated by a sense of hopelessness about the viability of other options, a sense of duty, and religious convictions. Nevertheless, given the encouraging initial work conducted with these scales, they might possibly merit greater attention in future research.

A variety of other self-report measures of granting forgiveness have been developed in the nearly 20 years since Trainer's (1981) work. More than 10 years ago, Wade (1987, 1989) developed an 81-item self-report pencil-and-paper measure that purports to assess nine dimensions of forgiveness. The measure has been used in several research efforts (Dreelin, 1993; Davidson & Jurkovic, 1993; McCullough & Worthington, 1995; Rhode, 1990). The subscales of Wade's Forgiveness Scale have demonstrated "known groups validity," successfully discriminating between people who report having forgiven an offender and those who report not having forgiven an offender (Wade, 1989). Subscales also appear to be correlated with spiritual well-being (Dreelin, 1993), and narcissism (Davidson & Jurkovic, 1994).

McCullough et al. (1998) reported the results of four studies in which they developed a 12-item measure based on items from Wade's (1989) Forgiveness Scale. Their 12-item scale—the Transgression-Related Interpersonal Motivations (TRIM) Inventory—is designed to assess the two neg-

ative motivational elements that are outlined in McCullough et al.'s (1997, 1998) theorizing about the nature of interpersonal forgiving: (1) the motivation to avoid the offender (Avoidance) and (2) the motivation to seek revenge (Revenge). The essence of forgiving, according to McCullough et al. (1998), is relative reductions in these two interpersonal motivations; thus, reductions in avoidance and revenge motivations are considered to be equivalent to "forgiving." The two TRIM subscales are correlated highly with many of the constructs that are central to McCullough et al.'s theorizing regarding the determinants of forgiving, including relational satisfaction, and commitment, and closeness, apology, empathy, and rumination. Confirmatory factor analyses also demonstrated that the TRIM subscale scores are distinct from these other constructs. As well, the two subscales predict approximately 50% of the variance (multiple $R = .70$) in a single-item measure of forgiveness and predict the reestablishment of relational closeness following an interpersonal offense.

Another important self-report measure has been developed by Enright and his colleagues at the University of Wisconsin-Madison (Subkoviak et al., 1995; Subkoviak, Enright, & Wu, 1992). The 60-item Enright Forgiveness Inventory (EFI) has been used in several studies and has a variety of desirable psychometric properties. It is intended to assess six aspects of forgiving another person: presence of positive affect, cognition, and behavior, and the absence of negative affect, cognition, and behavior. Internal consistency reliabilities for all six subscales are high, and internal consistency reliability for the total scale score is in the high .90s (Subkoviak et al., 1995; McCullough, 1995). In some subsamples, the EFI also was correlated negatively with measures of depression and anxiety (Subkoviak et al., 1995). Moreover, it is sensitive to experimental manipulations designed to facilitate forgiveness (Coyle & Enright, 1997).

Prior to the advent of the EFI Hebl and Enright (1993), Al-Mabuk, Enright, and Cardis (1995), and Freedman and Enright (1996) used a 30-item measure that they called the Psychological Profile of Forgiveness Scale. This scale was designed to assess the same six constructs as does the EFI. Internal consistency reliabilities for the total score on this 30-item measure tend to be above .90 (Al-Mabuk et al., 1995; Freedman & Enright, 1996; Hebl & Enright, 1993). Despite that fact that the 30-item Psychological Profile of Forgiveness Scale appears to be sensitive to experimental manipulations designed to facilitate forgiveness (e.g., Freedman & Enright, 1996), the 60-item EFI has been used in the stead of the 30-item Psychological Profile of Forgiveness Scale in more recent intervention studies by Enright and his colleagues (Coyle & Enright, 1997).²

Partner-Report Measures. We are aware of no measure that assesses the extent to which an offender perceives that an offended relationship partner has forgiven him or her for the offense. Given the essentially interper-

sonal nature of the concept of forgiveness, this seems like a tremendous oversight in the development of measures of forgiveness. Such measures would be relevant to (1) understanding the impact of forgiveness on the offender and (2) necessary for studying offense-specific forgiveness at the dyadic rather than simply the individual level (see Exline & Baumeister, Chapter 7, this volume).

Observer-Report Measures. Observer-report measures of forgiveness would be analogous to other interview-based measures for assessing psychological constructs, such as the Hamilton Depression Rating Scale for assessing depressive symptoms (Hamilton, 1960), and the Type A Structured Interview for assessing the Type A Behavior Pattern (Rosenman et al., 1964). Such measurement approaches have also lagged behind the development of self-report measures.

Along with developing the first multi-item self-report measure of granting forgiveness for a specific offense, Trainer (1981) also developed the first measure (see Trainer's Table E) of granting forgiveness to be completed by a trained rater. This measure was used only for validating the other scales that Trainer developed, but the approach deserves mention. Using this tool, trained raters categorize a person's degree of forgiveness for a specific offender as high, moderate, or low based on the relative presence or absence of several affective, cognitive, and behavioral qualities thought to reflect forgiveness (e.g., "relative absence of hostility and bitterness," "presence of gestures of good will," etc.).

In Chapter 9, this volume, Malcolm and Greenberg present a rating system for measuring offense-specific instances of forgiveness through analyzing psychotherapy process videotapes. This measurement system involves applying rating scales to videotapes of psychotherapy sessions during which the client and therapist discuss an interpersonal offense that the client has suffered. Though preliminary, Malcolm and Greenberg's approach is innovative and potentially quite useful—particularly for examining the unfolding of forgiveness in psychotherapy.

Behavioral Measures. Behavioral approaches whose scores might be interpreted as "forgiveness" have been explored in a variety of experimental settings. The Prisoner's Dilemma Game, for example, is a mixed-motive simulation in which two players are repeatedly faced with choosing either a cooperative or competitive strategy. The object is to win as many points as possible. If both partners cooperate in any given turn, each partner wins 3 points. If one partner cooperates while the other defects from cooperation, the "defector" wins 5 points, while the cooperator receives nothing. If both partners defect from cooperation, each wins 1 point. In this game, "forgiveness" has been operationalized as a cooperative move in response to the other player's competitive move (e.g., Axelrod, 1980a,

1980b). Rather than reciprocating a competitive move with another competitive move, the "forgiving" partner chooses to respond to the defection with a reassertion of his or her willingness to play the game in a cooperative mode.

Other behavioral measures should be considered also. Laboratory manipulations often present self-esteem threats or insults to participants and then give respondents the opportunity to behave in some way toward the person who is the source of the self-esteem threat or insult. In such laboratory settings, inhibition of aggressive responses to the transgressor may be considered an important aspect of forgiveness. Some aggressive responses that have been examined include (1) unfavorable evaluations of a transgressor that are believed to influence the transgressor's chances of winning a job or assistantship (e.g., Bushman & Baumeister, 1998; Caprara, Coluzzi, Mazzotti, Renzi, & Zelli, 1985; Kremer & Stevens, 1983; Zillman & Cantor, 1976); (2) delivery of shocks or other noxious stimuli (Caprara et al., 1985, 1987; Collins & Bell, 1997); and (3) making choices that cost the transgressor time, money, or some other resource (e.g., Brown, 1968; Caprara et al., 1987). In most instances, researchers would be interested in generalizing from participants' responses in such laboratory scenarios to interpersonal transgressions outside of the laboratory. Fortunately, it does appear that such laboratory-based measures of retaliatory aggression generalize well to nonlaboratory situations (Anderson & Bushman, 1997). Whether such behaviors in mixed-motive games or contrived laboratory scenarios actually correspond to forgiveness as measured using more conventional self-report, other-report, or observer-report measures, of course, remains to be investigated and would be substantively interesting for forgiveness theory.

Offense-Specific Measures of Seeking and Receiving Forgiveness

To date, we are aware of little research on how people seek or receive forgiveness for specific offenses (Meek et al., 1995). Meek et al. assessed (with single-item, Likert-type measures) the extent to which respondents would feel forgiven after confessing to the commission of certain transgressions. Little other published work has been completed in this domain.

Dyadic Measures of Forgiveness

At the dyadic level, a measure of forgiveness would assess a person's general tendency to forgive a particular relationship partner for interpersonal offenses that occur in the relationship. Relationship-specific assessments address questions such as the following: "Does this husband tend to seek forgiveness when he offends his wife?" "Does this employee tend to forgive his boss?" Thus, relationship-level measurements are less specific

than offense-specific measurements in that they attempt to assess a person's tendency to forgive (or seek forgiveness from) a specific relationship partner generally, rather than for a specific offense.

Currently, we are aware of only one tool that assesses forgiveness at the dyadic level. This is the Hargrave and Sells (1997) Interpersonal Relationship Resolution Scale (IRRS). This scale consists of 44 yes-no items designed to assess the extent to which a person who has received serious hurts from a specific family member (1) continues to feel pain as a result of the offenses and (2) has forgiven the offending family member for the offenses that occurred in the past. The Pain scale consists of four subscales labeled Shame, Rage, Control, and Chaos. The Forgiveness scale consists of four subscales labeled Insight, Understanding, Giving the Opportunity for Compensation, and the Overt Act of Forgiving. Internal consistencies for the eight subscales ranged from .63 to .87, and internal consistencies for the Pain and Forgiveness scales exceeded .90 (Hargrave & Sells, 1997). Individuals in a sample of outpatient psychotherapy clients with family-of-origin issues scored significantly lower on all four Forgiveness subscales (and higher on all four Pain subscales) than did a sample graduate and undergraduate students who had received no prior counseling or psychotherapy. The Forgiveness and Pain subscales also demonstrated a complex pattern of correlations with a variety of other clinically relevant self-report measures.

Further exploration of dyadic measures of forgiveness would facilitate greatly the study of forgiveness from the perspective of marital/family relationships and other close relationships. Such measures could potentially be developed in the same fashion as have other relationship-specific measures, such as the Long and Andrews (1990) dyadic perspective-taking measures, which assess partners' capacity to take the perspective of their spouse, as well as perceptions that the spouse is capable of taking the partners' perspective. Observer-report and behavioral measures of forgiveness at the relationship-specific level should be developed also.

Dispositional Measures of Forgiveness

At the dispositional level, a measure of forgiveness ostensibly assesses a person's general disposition or tendency to forgive others (or to seek forgiveness after having harmed someone else). Measures that assess forgiveness at this level are attempts to assess a general response style that transcends individual offenses, or even individual relationships.

Dispositional Measures of Granting Forgiveness

Self-Ratings. A variety of self-report measures exist that are useful for assessing the disposition to forgive others. The most widely used is a mea-

sure of the dissipation-rumination construct that Caprara and colleagues have investigated for many years (Caprara, 1986; Caprara, Barbaranelli, & Comrey, 1992; Caprara, Manzi, & Perugini, 1992; Caprara et al., 1985). Dissipation, according to Caprara, is the capacity to overcome ill feelings or the desire to retaliate after being provoked. Rumination is the tendency to maintain or even nurture one's emotional distress and desire for vengeance after being damaged by another person (Caprara & Pastorelli, 1989). The Dissipation-Rumination scale consists of 15 Likert-type items (e.g., "When somebody offends me, sooner or later I retaliate" and "I do not forgive easily once I am offended") and appears to have adequate internal consistency ($\alpha = .79$ for the Italian version and .89 for the English version). Dissipation-rumination scores appear to be positively correlated with the degree to which people retaliate against someone following a transgression or other self-esteem threat (Caprara, 1986; Caprara et al., 1985; Collins & Bell, 1997). In the last decade, similar scales have been developed as well. Mauger et al. (1992) developed a measure of the disposition to forgive other people, which they called the Forgiveness of Others Scale (FOS). Like the Dissipation-Rumination Scale, this measure consists of 15 items (in this case, in a true-false format) that assess people's desire to retaliate, hold grudges, and forgive following an interpersonal offense. Scores show adequate internal consistency. Two other dispositional scales that appear to draw from the same universe of items are the Emmons (1992) Beliefs about Revenge Questionnaire (BARQ) and the Stuckless and Goranson (1992) Vengeance Scale. The Dissipation-Rumination Scale, FOS, BARQ, and Vengeance Scale appear to have a variety of important correlates, including empathy, interpersonal trust, agreeableness, and social conformity. As well, people who are on the more "forgiving" end of scores for these measures appear to have lower irritability and emotional distress, and lower levels of anger and hostility.

The empirical relations among scores on the Dissipation-Rumination Scale, FOS, BARQ, and Vengeance Scale need to be examined empirically. Judging from the semantic similarity of their item content, it is likely that these measures share a large proportion of common variance. Intensive comparisons of the psychometric properties of these scales would help to broaden considerably researchers' and clinicians' tools for assessing the disposition to forgive.

An important measurement of self-reported disposition to grant forgiveness was developed by Hebl and Enright (1993). Their Willingness to Forgive Scale is a 16-item measure that instructs respondents to read 16 scenarios in which they imagine themselves to have been damaged by another person. Respondents choose one of 10 hypothetical responses to each offense to indicate how they (1) expect that they would respond to the offense, and how they (2) would prefer to respond to the offense,

even if they do not believe that they would choose that particular mode of responding. One of the 10 response options to each scenario is to "forgive." Scores represent the number of times that respondents choose the "forgiveness" response to the scenarios. Internal consistency reliability for this measure was estimated at $\alpha = .70$ (Hebl & Enright, 1993). Al-Mabuk, Enright, and Cardis (1995), using a 12-item form of this scale, demonstrated internal consistency coefficients in the .47-.75 range. Subjects' scores on this measure appear to increase after participation in interventions focused on encouraging forgiveness (Al-Mabuk, Enright, & Cardis, 1995; Hebl & Enright, 1993).

Aside from these measures, other approaches to assessing dispositional aspects of granting forgiveness have been developed through the years. The value that people ascribe to being "forgiving" is assessed with Rokeach's (1967) Value Survey. This scale requires respondents to place 18 terminal values and 18 instrumental values (including the value of being "forgiving") in rank order according to their relative priority in respondents' value systems. People who tend to be more religious, lower in Machiavellianism, and more traditional tend to place the value of being "forgiving" higher in their value systems than do people who are lower on each of those respective traits (for reviews see McCullough & Worthington, 1995, 1999).

Finally, Gorsuch and Hao (1993) examined a series of self-report items regarding people's tendencies to forgive and retaliate when they have been intentionally hurt by others. Gorsuch and Hao reported that a variety of measures of religious involvement differentiated between people scoring high and low on these measures of forgiveness. Because these items have been administered to a nationally representative sample, they should be looked at seriously for developing standardized measures of dispositional forgiveness.

Partner Ratings. Currently, no psychometric tools are available for assessing the disposition to forgive from the perspective of a significant other.

Observer-Report Measures. As well, no observer-report measures have been developed for assessing the disposition to forgive.

Behavioral Measures. Finally, no behavioral measures have been developed to assess dispositional forgiveness by observing and summing people's behavioral responses across a variety of real-life situations.

Dispositional Measures of Seeking-Receiving Forgiveness

We are not aware of any psychometric instruments to date that have attempted to measure people's tendencies to seek forgiveness (or receive

forgiveness) from others when they have committed an interpersonal transgression.

Summary

Clearly, the measurement of forgiveness has been progressing for nearly two decades. The energy that has been devoted to developing self-report measures of offense-specific forgiveness probably reflects the fact that most theorizing about forgiveness has focused on granting forgiveness in the aftermath of transgression or victimization. The relative emphasis on measures of granting forgiveness might also be due to the fact that research to date has not focused on the interpersonal aspects as much as the intrapersonal or mental health aspects of forgiveness. Nevertheless, the scientific study of forgiveness would benefit a great deal from research efforts devoted to increasing the breadth of available measurement methods. Whereas many self-report instruments exist for assessing transgression-specific forgiveness, few instruments exist for measuring forgiveness at the dyadic level. Development of behavioral measures has been almost completely ignored. No dispositional measures of seeking-receiving forgiveness exist. We hope that these assessment gaps will be filled in years to come.

OTHER AREAS FOR PSYCHOMETRIC DEVELOPMENT

In addition to increasing the breadth of measures available for assessing forgiveness, the entire field of research on forgiveness would be much improved if procedural changes were made in how measures of forgiveness were developed and employed in research. First and foremost, we recommend that researchers begin to augment the use of classical measurement theory with generalizability theory, which can address the complexities of forgiveness assessment in a more elegant and informative way.

Four subsidiary psychometric issues emerge from a generalizability approach to thinking about the measurement of forgiveness constructs. First, we recommend that forgiveness constructs routinely be assessed using more than one method. Second, we recommend that researchers think hard about the appropriate level of specificity at which forgiveness should be assessed to address a particular research question adequately. Third, we recommend that researchers begin to explore (and exercise care in controlling) sources of variability that can influence forgiveness scores. Fourth, we recommend that researchers specify the nomological networks underlying the assessment of forgiveness prior to collecting construct validity data. We explore each of these themes in the remainder of the present chapter.

Generalizability Theory as an Adjunct to Classical Measurement Theory

Classical measurement theory would conceptualize people's scores on a measure of forgiveness as the additive combination of a true score component and an error component. The true score reflects the people's actual standing on the construct of interest (e.g., the extent to which persons "really" forgive the transgressor they are rating, the extent to which they "really" have a forgiving disposition, etc.). The error component refers to variance in the measure that is unstable across replications, thereby making the assessment procedure an imperfect measure of people's actual standing on the construct. Reliability coefficients (e.g., coefficient alpha) are interpreted under the assumptions of classical measurement theory as the proportion of the total variance in a measure that is attributable to true scores.

In classical measurement theory, psychometricians have typically been concerned with three sources of error variance because of their impact on the reliability of measures: (1) error due to temporal instability, (2) error due to heterogeneity among items, and (3) error due to inconsistencies in how two raters might apply a rating scale to the same stimulus. These sources of error are estimated, respectively, with test-retest coefficients (e.g., interclass r), internal consistency reliability coefficients (e.g., Cronbach's α), and interrater reliability coefficients (e.g., intraclass r , see Shrout & Fleiss, 1979). The major limitation of classical measurement theory for estimating a measure's dependability is that it does not allow one to consider multiple sources of error variance simultaneously. When two or more sources of error contribute to measurement variance, classical reliability tests always overestimate the proportion of variance that is attributable to true scores (Hoyt, 1998).

Generalizability theory (GT; Cronbach, Gleser, Nanda, & Rajaratnam, 1972) is an expanded psychometric approach that is not limited to the examination of a single source of extraneous variance at any given time. GT is to classical reliability theory what factorial analysis of variance is to one-way analysis of variance (ANOVA) (Hoyt & Melby, 1999). Although GT is not a new psychometric approach, it would be a new approach for use in developing measures of forgiveness constructs. Using a GT approach, a researcher would attempt to identify all of the facets (or dimensions along which observations might be classified) that might influence scores on a forgiveness measure in a given application, and examine their relative contributions to score variance simultaneously.

In the case of a self-report, offense-specific measure of granting forgiveness (e.g., Wade's Forgiveness Scale), for example, such facets would almost certainly include items (the facet for which internal consistency re-

liability estimates in classical measurement theory are conducted) and time (the facet for which test-retest reliability estimates in classical measurement theory are conducted). Such facets can be specified and estimated explicitly and simultaneously in a GT approach to scale development. As well, the interaction of these effects can be specified and examined. What matters most, from a GT point of view, is not a participant's "true score," but, rather, the universe of conditions (or levels of facets) to which a given score might generalize.

Although facets are commonly equated with sources of error in generalizability studies, an aspect of measurement that is error in one application may be substantively important when the construct of interest is considered from another point of view (Cronbach, 1995). By advocating a flexible, context-sensitive approach to interpretation of the results of generalizability studies, GT blurs conventional distinctions between reliability and validity of measurement (Cronbach et al., 1972).

Table 4.1 illustrates a 4 (victims) \times 3 (offenders) \times 2 (offenses) fully crossed factorial design for a hypothetical generalizability study of forgiveness. The number of levels of each factor is small for illustrative purposes; to estimate variance components reliably, it would be important to include more victims and offenders, or to study multiple small groups of victims and offenders (Smith, 1978). For concreteness, readers can think of victims as students (or siblings) and offenders as teachers (or parents). All victims are assumed to be acquainted with all individuals in the group of offenders, and victims are asked to report the extent to which they forgave offenders for one major and one minor offense in the recent past. Cell entries X_{ijk} are ratings of the extent to which victim i forgave offender j following offense k (1 = major; 2 = minor).

This data set is a fully crossed factorial design, and the generalizability analysis uses the same procedures as factorial analysis of variance; thus, this generalizability study investigates seven sources of variance in ratings: three main effects, three two-way interactions, and one three-way interaction (confounded with error because, as is typical in generalizability studies, the design includes only one observation per cell). However, unlike ANOVA,

TABLE 4.1. Design for Generalizability Study of Forgiveness Behaviors

Partner (j)	1		2		3	
	Major	Minor	Major	Minor	Major	Minor
Offense (k)						
Actor (i) 1	X_{111}	X_{112}	X_{121}	X_{122}	X_{131}	X_{132}
2	X_{211}	X_{212}	X_{221}	X_{222}	X_{231}	X_{232}
3	X_{311}	X_{312}	X_{321}	X_{322}	X_{331}	X_{332}
4	X_{411}	X_{412}	X_{421}	X_{422}	X_{431}	X_{432}

which focuses on testing main effects and interactions for statistical significance, generalizability analyses focus on the magnitude of these effects, estimating the variance accounted for by each main effect and interaction in the model. These variance components indicate the relative importance of each of these seven sources in determining forgiveness.

Of special interest in connection with our concerns about determining the appropriate level of measurement of forgiveness (discussed earlier) are the victim (*i*), offender (*j*), and relationship (*ij* interaction) variance components. Examination of the relative magnitude of these components addresses the following issues (see Table 4.2):

1. To what extent is forgiveness a function of victim disposition (i.e., to what extent do victims differ in their average willingness to forgive)?
2. To what extent is forgiveness a function of offender "forgivability" (i.e., to what extent do offenders differ in the average level of forgiveness they elicit from those they have offended)?
3. To what extent is forgiveness a function of unique characteristics of the victim-offender relationship (i.e., are some victims more likely to forgive a particular offender than others)? If forgiveness in a relationship is mediated by liking, for example, relationship variance in forgiveness would be expected to be large, because liking is largely a dyadic phenomenon (Kenny, 1994).

TABLE 4.2. Interpretation of Variance Components from Generalizability Study

Component	Question/issue addressed
<i>i</i>	To what extent is forgiveness a function of actor characteristics (actor variance in the Social Relations Model)?
<i>j</i>	To what extent is forgiveness a function of partner characteristics (partner variance in Social Relations Model)?
<i>ij</i>	To what extent is forgiveness a function of characteristics of the dyad (dyadic variance in Social Relations Model)?
<i>k</i>	To what extent is forgiveness a function of the magnitude of the offense?
<i>ik</i>	Do actors vary in their willingness to forgive minor versus major offenses?
<i>jk</i>	Do partners vary in their forgiveability for minor versus major offenses?
Residual	To what extent is forgiveness a function of systematic or random factors not modeled in this study?

Our hypothetical generalizability study partitions forgiveness variance into other components, including variance attributable to the severity of offense (*k*) and to the interaction of offense severity with victim and offender effects (*ik* and *jk* interactions, respectively), along with a residual component. These components also address questions of dependability of measurement, although they are substantively interesting as well, since researchers should be aware of the extent to which a given victim's forgiveness levels are likely to generalize across offenses, both within and between relationships.

The main effect of offense examines whether forgiveness levels vary by the severity of the offense, as might be expected (e.g., Girard & Mullet, 1997). The *ik* interaction examines whether the ordering of victims on forgiveness differs by level of offense. A relatively small *ik* variance component indicates that individuals who readily forgive a trivial slight are also most willing to forgive more severe transgressions. If *ik* variance in forgiveness is large, however, the victims who are most forgiving of a minor offense may not be most forgiving following a major hurt. If this is the case, theories of forgiveness must be careful to specify the magnitude of the offense in their predictions, and offense-specific measures of forgiveness must control for severity to assess a unified construct. The *jk* interaction addresses a similar issue with respect to offender forgiveability: Are those who are most forgivable following minor transgressions also the most forgivable when they have committed a major offense? If *jk*-related variance is negligible in the generalizability study, it suggests that the answer to this question is "yes"; if not, then offense severity should be a consideration in theories and research on individual differences in forgiveability.

In addition to the facets that are specified in Tables 4.1 and 4.2, other sources of error that are of interest in the measurement of forgiveness could be investigated either separately or in conjunction with the sources considered here. Stability of forgiveness over time (i.e., test-retest reliability) is clearly an important consideration for researchers: To what extent does forgiving a relationship partner today predict relative standing on the same forgiveness construct in the future? When multi-item measures of forgiveness are used to assess this construct, generalizability of forgiveness across items (i.e., internal consistency reliability) is also of interest.

Exalting Multimethod Measurement as a Gold Standard

Recently, researchers have begun to lament the fact that most studies of forgiveness have relied almost exclusively on monomethod measurement of forgiveness based on the self-reports of only a single person: the putative forgiver. There has been very little study of the perpetrator's side of things (Baumeister, Exline, & Sommer, 1998; Exline & Baumeister, Chap-

ter 7, this volume). A serious limitation imposed by exclusive reliance on forgivers' self-reports is the unknown extent to which individual differences on these measures are confounded with response biases or differences among respondents in their interpretations of rating scales (Hoyt, 1998; Kenny, 1994). Unfortunately, variance due to response biases is reliable (i.e., it is counted as "true score" variance in conventional reliability analyses). Thus, validity coefficients in studies using self-report measures of forgiveness are attenuated to an unknown extent due to this confound, and statistical power is correspondingly decreased by an unknown amount.

To remedy the shortcomings imposed by monomethod assessment, researchers should strive to complement self-reports of forgiveness from the "forgiver's" side of things with reports from other informants (including relationship partners), rating scales completed by third parties, or behavioral measures. This is not to suggest that such measures will necessarily converge; in fact, the expectation that they might *not* produce convergent findings is precisely the reason why multimethod measurement is used. If we find, for example, that self-reported empathic affect is correlated with self-reported forgiveness (e.g., McCullough et al., 1997, 1998), but not with behavioral measures that putatively assess forgiveness, or the transgressor's perceptions of the extent to which he or she has been forgiven; then we learn something important about the limits of the empathy-forgiveness connection.

Assessing Forgiveness at the Appropriate Level of Specificity

Assessing forgiveness at the appropriate level of specificity is likely to have important consequences for the development of stable and theoretically compelling research findings regarding forgiveness. According to the aggregation principle explored initially by Fishbein and Ajzen (1974), and later by Rushton, Brainerd, and Pressley (1983), Epstein (1983), and Gorsuch (1984, 1988), scores on psychometric instruments are most likely to correlate with scores on other psychometric instruments that aggregate behaviors to the same level of specificity. A corollary of this principle is that measures assessing a single behavioral instance are most likely to be correlated with other measures that are germane to that same single behavioral instance. Moreover, measures that assess a variable (such as forgiveness) at the level of a specific relationship are most likely to correlate with other measures that assess qualities of a specific relationship. Finally, measures that assess an overall personality trait or disposition are most likely to correlate with other measures that assess an overall personality trait or disposition.

The generalizability study summarized in Tables 4.1 and 4.2 can be used to illustrate the specificity principle in the context of GT. People may be inconsistent in their responses to multiple offenses within a single rela-

tionship (variance attributed to k , ik , or jk components in Table 4.1). The inconsistency may be systematically related to the severity of the offense, which was manipulated in this example, but may also be a function of other factors (e.g., qualities of the relationship between offender and offeree) not examined in this study. People may be consistent in their tendencies to forgive within a single relationship but vary across relationships (j and ij components in Table 4.1). Forgiveness behaviors are probably at least partly a function of one or more of these facets.

Failure to pay attention to the specificity principle in selecting appropriate measures of forgiveness puts researchers at risk of failing to find important relationships. It also creates conceptual confusion. For example, imagine the following scenario. Dr. A. was interested in assessing the association of forgiveness and marital adjustment. To test his hypothesis that "forgiving promotes marital adjustment," Dr. A. administered Wade's (1989) Forgiveness Scale and the Locke-Wallace (1959) Marital Adjustment Test to a sample of 300 married persons. Dr. A. found that participants' scores on Wade's Forgiveness Scale were correlated significantly with scores on the Locke-Wallace at $r = .16$, $p < .01$. On this basis, Dr. A. concluded that forgiveness was related to marital adjustment.

Such a conclusion could potentially be both conceptually and empirically problematic. It was not forgiveness in any general sense that was measured, but rather individuals' reports of the extent to which they had forgiven their spouses for a particular interpersonal offense (presumably, one that each respondent chose without guidance from the researcher). Thus, even though Dr. A. appeared to be interested in forgiveness at the dyadic (relationship) level (i.e., the extent to which a partner forgives his or her spouse), he used an occasion-specific measure of forgiveness (the extent to which a partner forgave his or her spouse for a *single, isolated offense*). In GT terms, Dr. A.'s implicit assumption was that offense-specific forgiveness would generalize to other offenses within the relationship, that is, would be highly correlated with a good dyadic measure of forgiveness. To the extent that forgiveness is consistent across offenses within a relationship, this assumption is unproblematic. However, if actors vary in their willingness to forgive multiple offenses by a single partner, then the forgiveness assessment would have been error laden in Dr. A.'s study. The likely consequence is attenuation of the observed effect size, and reduced statistical power.³ Clearly, the obtained correlation of $r = .16$ was statistically significant, but what is the meaning of this modest correlation? Because the measure used to assess forgiveness was not at the appropriate level of specificity, Dr. A.'s hypothesis received a weak test (and was only vaguely supported). A better test of Dr. A.'s hypothesis would have been to use a less specific measure of forgiveness. To test whether forgiving marriages really are better quality marriages, Dr. A. could have aggregated several instances of forgiveness within each marriage or developed a measure that assessed forgiveness as a general relational quality.

We are not trying to say that it is never appropriate to examine the relations of variables with different levels of specificity. For example, it would be surprising if people's level of empathy for an individual who had damaged them was not related, in an empirical sense, to their dispositional capacity for empathy. As well, it is clear that narcissism—a dispositional variable—is a predictor of the extent to which people will engage in aggressive behavior toward an offender who threatens their self-esteem in a laboratory scenario (Bushman & Baumeister, 1998). Responses to such laboratory scenarios, of course, would be considered offense-specific measures in the $3 \times 2 \times 4$ taxonomy. What we *are* saying, however, is that without good theory to explain how constructs at one level of specificity are related to forgiveness constructs that are assessed at different levels of specificity, such empirical efforts (1) obscure the fact that forgiveness is likely to operate at different levels, (2) ignore the fact that adequate generalizability between levels is by no means assured, and (3) (to the extent that generalizability across levels is limited) lead to attenuated associations.

Controlling Sources of Variability in Forgiveness Ratings

Just as measuring forgiveness at the appropriate level of specificity is likely to produce pay-offs, so is the stringent control of measurement error. Usually, researchers think of controlling measurement error through random assignment to experimental conditions or through statistical controls. As well, it is likely that unintended sources of error creep into forgiveness ratings through weak control of the instructions given to research participants. In many studies (e.g., McCullough, Worthington, & Rachal, 1997, Study 1), researchers direct respondents to recall a transgression they had suffered at some point in the past, and then to indicate the extent to which they have forgiven their transgressor for this offense. This rating task probably introduces measurement error from a variety of sources.

When respondents are free to choose a transgression to rate in such a task, researchers have no idea what sorts of considerations inform respondents' choices. Certainly, some respondents will choose recent transgressions that are still salient because they are actively trying to figure out how to cope with them. Other respondents might choose long-standing sources of bitterness that are salient because they have been discussing them in psychotherapy. Still others might choose rather trivial offenses, because such offenses are the only ones that they are able to recall. Some people might choose the transgressions of a close relationship partner, while others might choose the transgression of a coworker, acquaintance, or stranger. Because factors such as the recency, salience, perceived severity, and identity of the perpetrator of the offense are likely to influence forgiveness ratings (and because these sources of error are difficult to

control even statistically), it is likely that they introduce error that interferes with our ability to understand the psychological processes at work in forgiveness.

Such sources of variation in forgiveness ratings can be controlled through two design features. First, when the researcher's goal is to assess a person's general tendency to forgive, but forgiveness is assessed in only a single relationship for each actor, variance due to actor, partner, and dyad effects are confounded, potentially creating large amounts of noise in forgiveness scores (Hoyt, 1998). One method to reduce confounding is to include multiple partners and offenses, as illustrated in Table 4.1.

Second, such sources of error can be controlled (at least in part) by examining forgiveness in the context of specific interpersonal transgressions. The most recent intervention work from Enright and his colleagues (e.g., Coyle & Enright, 1997; Freedman & Enright, 1996) exemplifies this approach. In both of these studies, sample participants were selected precisely because they had suffered similar transgressions. In the Coyle and Enright (1997) study, all the participants were men who felt that they had been damaged by a sexual partner who had decided to abort a fetus that they had helped to conceive. While the abortions had taken place at varying lengths of time in the past and probably had different levels of salience for each of the men in the study, focusing on a more or less uniform transgression probably helped to eliminate error that might have influenced the men's scores on a measure of the extent to which they had forgiven the relationship partner. Similarly, Freedman and Enright (1996) examined the extent to which adult women had forgiven male family members who had sexually abused them as children. Again, focusing on a specific type of transgression, perpetrated by people in a specific kind of relationship to the participant, probably helped to control a variety of sources of variance that otherwise would have contributed to error.

Specifying the Nomological Network Prior to Collecting "Construct Validity" Data

Construct validation is the process of determining what psychological constructs account for the performance of an instrument (Cronbach & Meehl, 1955). In theory, scores on a measure that assesses forgiveness should be related, at the proper magnitudes, to the variables that a particular theory of forgiveness specifies to be related to forgiveness. Thus, the endeavor to develop measures for assessing forgiveness that have "construct validity" presupposes that existing theoretical treatments of forgiveness specify a set of probabilistic statements about how forgiveness is related to other constructs of interest (and measures of those constructs). Cronbach and Meehl called this set of probabilistic statements the construct's "nomological net."

In their landmark paper on construct validity, Cronbach and Meehl

(1955) pointed out that concepts with short scientific histories tend to have sparse nomological nets. Early in construct development, little will be known about how the construct is related to other constructs because (1) little solid theoretical work will exist; and (2) the existing psychometric studies will be preliminary. This describes the status of forgiveness research perfectly—the existing theoretical work has not stimulated very much empirical work, and most of the existing measures of forgiveness are only loosely tied to theoretical treatments of forgiveness. Of course, this is exactly what Cronbach and Meehl would lead us to expect. But that does not mean that all is as it should be.

To move beyond this current state of psychometric development, researchers should begin to tie their conceptualizations (and measures) of forgiveness into larger conceptual networks. Based on the best theory and research, what constructs (and operationalizations of them) and variables should explain variability in forgiveness? Which of these should be causally prior to forgiveness? Which should be concomitants of forgiveness? Which should be consequences? What should be the strength of these relationships? At what levels of specificity (offense-specific, dyadic, or dispositional) should measures of forgiveness constructs be related to measures of other constructs? Currently, no comprehensive nomological net has been laid out that would permit us to determine whether the correlations of forgiveness variables with other variables of interest (e.g., hostility, blame, empathy, spirituality, hope, anxiety, etc.) that have been obtained in research to date are actually evidence “for” or “against” the construct validity of such forgiveness measures.

Another element of construct validation is specifying and examining which constructs (and measures of those constructs) should be independent of forgiveness constructs (and measures of those forgiveness constructs). Such conceptual and empirical exercises—the tasks of exploring discriminant validity—have been mostly ignored in psychometric research on forgiveness to date. All of the issues raised here should be addressed in efforts to understand the construct validity of a psychometric instrument designed to measure a forgiveness construct.

CONCLUSION

In this chapter, we have tried to bring some clarity to the existing empirical work dedicated to the assessment of forgiveness. By organizing the existing measures along three dimensions (specificity of measurement, direction of measurement, and method of measurement) we see that some psychometric work is being done in certain areas, but other areas remain weak or completely unexplored. Each of the unexplored areas remains uncharted scientific territory.

In charting this new scientific territory—and in elaborating what we already know—we recommend that researchers treat future psychometric studies on forgiveness with the same forethought and methodological rigor that is extended in other areas of social-scientific research. A generalizability framework provides, to our way of thinking, the best conceptual bridge for matching psychometric research strategies to the complexities that might underlie forgiveness in particular research applications. As well, researchers should also begin to use multiple measures of forgiveness that are appropriate to particular research applications and should choose research designs that reduce extraneous score variance. Researchers should also continue to develop more explicitly a set of theoretically derived, probabilistic statements that link particular forgiveness constructs (and their measures) to other constructs (and measures). Addressing these psychometric concerns in future studies will help increase both the breadth and depth of what we know about this set of interesting and important constructs.

NOTES

1. This third “methods” dimension is obviously somewhat plastic. Forgiveness constructs could be assessed with many more than these four methods, including qualitative analysis, content analysis, diary methods, physiological methods, archival methods, and historiometric methods. Our neglect of these methods is not meant to suggest that we do not believe them to be valid; rather, it reflects the fact that no measures currently exist using these methods, and so to include such methods in our taxonomy would have been pointless. Clearly, all of these categories represent alternative “methods” through which forgiveness could be assessed, and each of them deserves to be explored in its own right.
2. McCullough, Worthington, and Rachal (1997) also used a 5-item self-report measure of forgiveness. These items and their instructions were adapted from the EFI (Subkoviak et al., 1995), although the McCullough et al. paper failed to attribute these items to Subkoviak et al. Enright has requested that researchers refrain from using this combination of items from the EFI in their own research. Rather, he requests that interested researchers use the complete, 60-item EFI.
3. Ironically, the effect size in Dr. A.’s study would also presumably have been inflated to some extent by correlated response bias, because the same person rated forgiveness and dyadic adjustment for each relationship. The net impact of these uncontrolled upward and downward biases in effect sizes can only be estimated by conducting a multivariate generalizability study (Hoyt, 1998).

REFERENCES

- Al-Mabuk, R. H., Enright, R. D., & Cardis, P. A. (1995). Forgiveness education with parentally love-deprived late adolescents. *Journal of Moral Education, 24*, 427–444.
- Anderson, C. A., & Bushman, B. J. (1997). External validity of “trivial” experiments: The case of laboratory aggression. *General Psychology Review, 1*, 19–41.

- Axelrod, R. (1980a). Effective choice in the Prisoner's Dilemma. *Journal of Conflict Resolution*, 24(3), 3-25.
- Axelrod, R. (1980b). More effective choice in the Prisoner's Dilemma. *Journal of Conflict Resolution*, 24(3), 379-403.
- Baumeister, R. F., Exline, J. J., & Sommer, K. L. (1998). The victim role, grudge theory, and two dimensions of forgiveness. In E. L. Worthington (Ed.), *The foundation of forgiveness* (pp. 79-104). Philadelphia: Templeton Foundation Press.
- Brown, B. R. (1968). The effects of need to maintain face on interpersonal bargaining. *Journal of Experimental Social Psychology*, 4, 107-122.
- Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology*, 75, 219-229.
- Caprara, G. V. (1986). Indicators of aggression: The dissipation-rumination scale. *Personality and Individual Differences*, 7, 763-769.
- Caprara, G. V., Barbaranelli, C., & Comrey, A. L. (1992). A personological approach to the study of aggression. *Personality and Individual Differences*, 13, 77-84.
- Caprara, G. V., Coluzzi, M., Mazzotti, E., Renzi, P., & Zelli, A. (1985). Effect of insult and dissipation-rumination on delayed aggression and hostility. *Archivio di Psicologia, Neurologia e Psichiatria*, 46, 130-139.
- Caprara, G. V., Gargaro, T., Pastorelli, C., Prezza, M., Renzi, P., & Zelli, A. (1987). Individual differences and measures of aggression in laboratory studies. *Personality and Individual Differences*, 8(6), 885-893.
- Caprara, G. V., Manzi, J., & Perugini, M. (1992). Investigating guilt in relation to emotionality and aggression. *Personality and Individual Differences*, 13, 519-532.
- Caprara, G. V., & Pastorelli, C. (1989). Toward a reorientation of research on aggression. *European Journal of Personality*, 3, 121-138.
- Collins, K., & Bell, R. (1997). Personality and aggression: The dissipation-rumination scale. *Personality and Individual Differences*, 22, 751-755.
- Coyle, C. T., & Enright, R. D. (1997). Forgiveness intervention with post-abortion men. *Journal of Consulting and Clinical Psychology*, 65(6), 1042-1046.
- Cronbach, L. J. (1995). Giving method variance its due. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: A festschrift in honor of Donald Fiske* (pp. 145-157). New York: Wiley.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Darby, B. W., & Schlenker, B. R. (1982). Children's reactions to apologies. *Journal of Personality and Social Psychology*, 43, 742-753.
- Davidson, D. L., & Jurkovic, G. J. (1993, April). *Forgiveness and narcissism: Consistency in experience across real and hypothetical hurt scenarios*. Paper presented at the annual convention of the Christian Association for Psychological Studies, Kansas City, MO.
- Dreelin, E. D. (1993). *Religious functioning and forgiveness*. Unpublished dissertation, Fuller Graduate School of Psychology, Pasadena, CA.
- Epstein, S. (1983). Aggregation and beyond: Some issues on the aggregation of behavior. *Journal of Personality*, 51(3), 360-392.
- Fishbein, M., & Ajzen, I. (1974). Attitudes toward objects as predictors of single and multiple behavioral criteria. *Psychological Review*, 81, 59-74.
- Freedman, S. R., & Enright, R. D. (1996). Forgiveness as an intervention goal with incest survivors. *Journal of Consulting and Clinical Psychology*, 64, 510-517.
- Gassin, E. A. (1998). Receiving forgiveness as moral education: A theoretical analysis and initial empirical investigation. *Journal of Moral Education*, 27, 71-87.

- Girard, M., & Mullet, E. (1997). Forgiveness in adolescents, young, middle-aged, and older adults. *Journal of Adult Development*, 4, 209-220.
- Gorsuch, R. L. (1984). Measurement: The boon and bane of investigating religion. *American Psychologist*, 39, 228-236.
- Gorsuch, R. L. (1988). Psychology of religion. *Annual Review of Psychology*, 39, 201-221.
- Gorsuch, R. L., & Hao, J. Y. (1993). Forgiveness: An exploratory factor analysis and its relationships to religious variables. *Review of Religious Research*, 34(4), 333-347.
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, 23, 56-62.
- Hargrave, T. D., & Sells, J. N. (1997). The development of a forgiveness scale. *Journal of Marital and Family Therapy*, 23(1), 41-62.
- Hebl, J. H., & Enright, R. D. (1993). Forgiveness as a psychotherapeutic goal with elderly females. *Psychotherapy*, 30, 658-667.
- Hoyt, W. T. (in press). Rater bias: When is it a problem and what can we do about it? *Psychological Methods*.
- Hoyt, W. T., & Melby, J. N. (1999). Dependability of measurement in counseling psychology: An introduction to generalizability theory. *Counseling Psychologist*, 27, 325-352.
- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.
- Kremer, J. F., & Stephens, L. (1983). Attributions and arousal as mediators of mitigation's effect on retaliation. *Journal of Personality and Social Psychology*, 45(2), 335-343.
- Locke, H. J., & Wallace, K. M. (1959). Short marital-adjustment and prediction tests: Their reliability and validity. *Marriage and Family Living*, 21, 251-255.
- Long, E. C. J., & Andrews, D. W. (1990). Perspective taking as a predictor of marital adjustment. *Journal of Personality and Social Psychology*, 59, 126-131.
- Mauger, P. A., Perry, J. E., Freeman, T., Grove, D. C., McBride, A. G., & McKinney, K. E. (1992). The measurement of forgiveness: Preliminary research. *Journal of Psychology and Christianity*, 11(2), 170-180.
- McCullough, M. E. (1995). *Forgiveness as altruism: A social-psychological theory of forgiveness and tests of its validity*. Unpublished doctoral dissertation, Virginia Commonwealth University, Richmond.
- McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington, E. L., Jr., Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships II: Theoretical elaboration and measurement. *Journal of Personality and Social Psychology*, 75, 1586-1603.
- McCullough, M. E., Sandage, S. J., Rachal, K. C., & Worthington, E. L., Jr. (1997, August). *A sustainable future for the psychology of forgiveness*. Paper presented at the annual meeting of the American Psychological Association, Chicago, IL.
- McCullough, M. E., & Worthington, E. L., Jr. (1995). Promoting forgiveness: A comparison of two brief psychoeducational group interventions with a waiting-list control. *Counseling and Values*, 40, 55-68.
- McCullough, M. E., & Worthington, E. L., Jr. (2000). Religion and the forgiving personality. *Journal of Personality*.
- McCullough, M. E., Worthington, E. L., Jr., & Rachal, K. C. (1997). Interpersonal forgiving in close relationships. *Journal of Personality and Social Psychology*, 73(2), 321-336.
- Meeck, K. R., Albright, J. S., & McMinn, M. R. (1995). Religious orientation, guilt, confession, and forgiveness. *Journal of Psychology and Theology*, 23(3), 190-197.
- Park, Y. O., & Enright, R. D. (1997). The development of forgiveness in the context of adolescent friendship conflict in Korea. *Journal of Adolescence*, 20(4), 393-402.
- Rhode, M. G. (1990). *Forgiveness, power, and empathy*. Unpublished doctoral dissertation, Fuller Graduate School of Psychology, Pasadena, CA.
- Rokeach, M. (1967). *Human Values Survey*. Sunnysvale, CA: Halgren Tests.
- Rosenman, R. H., Friedman, M., Straus, R., Wurm, M., Kositchek, R., Hahn, W., & Werthessen,

- N. T. (1964). A predictive study of coronary heart disease. *Journal of the American Medical Association*, 189, 113-124.
- Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin*, 94(1), 18-38.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Smith, P. (1978). Sampling errors of variance components in small sample generalizability studies. *Journal of Educational Measurement*, 3, 319-346.
- Subkoviak, M. J., Enright, R. D., & Wu, C. (1992, October). *Current developments related to measuring forgiveness*. Paper presented at the annual meeting of the Midwestern Educational Research Association, Chicago, IL.
- Subkoviak, M. J., Enright, R. D., Wu, C., Gassin, E. A., Freedman, S., Olson, L. M., & Sarinopoulos, I. (1995). Measuring interpersonal forgiveness in late adolescence and middle adulthood. *Journal of Adolescence*, 18, 641-655.
- Trainer, M. F. (1981). *Forgiveness: Intrinsic, role-expected, expedient, in the context of divorce*. Unpublished doctoral dissertation, Boston University.
- Wade, S. H. (1987). *A content analysis of forgiveness*. Unpublished master's project, Fuller Graduate School of Psychology, Pasadena, CA.
- Wade, S. H. (1989). *The development of a scale to measure forgiveness*. Unpublished doctoral dissertation, Fuller Graduate School of Psychology, Pasadena, CA.
- Zillman, D., & Cantor, J. R. (1976). Effect of timing of information about mitigating circumstances on emotional responses to provocation and retaliatory behavior. *Journal of Experimental Social Psychology*, 12, 38-55.

PART II

Basic Psychological Research