

Issues in the Multimodal Measurement of Forgiveness

William T. Hoyt
Michael E. McCullough

Although the recommendation to include multiple indicators of a construct of interest is commonplace in research design texts (e.g., Cook & Campbell, 1979), investigators conducting basic and applied research on forgiveness have typically relied on a single measure of this construct. There are many reasons why forgiveness researchers may fail to use multimodal measurement. Certainly, inclusion of additional questionnaires or other measurement procedures poses a burden to participants as well as researchers. Perhaps investigators simply do not believe that a concomitant benefit to research validity will compensate for this additional burden. In this relatively new research area, selecting multiple indicators that overlap sufficiently to constitute measures of the same underlying construct, but not so much that they are essentially redundant, may pose a challenge. Finally, researchers may avoid including more than one measure of forgiveness because this augmentation to the research design creates challenges at the data analysis phase. Our goal in this chapter is to address each of these challenges. In the sections that follow, we review the rationale for preferring multimodal measurement, provide a conceptual framework to assist researchers and research consumers in evaluating forgiveness measures, and describe common models for data analysis using multiple measures, with illustration of their relevance to the forgiveness domain.

PERSONAL ASSUMPTIONS ABOUT FORGIVENESS

The authors come to forgiveness research from the perspective of counseling psychology, with its emphasis on positive development and enhancing human strengths. Forgiveness is of interest as a human virtue from the perspective of numerous religious and spiritual traditions (McCullough & Snyder, 2000). As scientists, counseling psychologists have typically been sensitive to the importance of environmental and contextual determinants of human behavior, and to the importance of social relationships

as an indicator of quality of life (Heppner, Casas, Carter, & Stone, 2000). Because of its relationship-enhancing potential, the capacity for forgiveness may well be an important indicator of both relational and individual health. In this chapter, we consider forgiveness as first and foremost an interpersonal process, with little attention to other aspects of forgiving (e.g., forgiveness of self; forgiveness of God or of inanimate objects or events). We conceptualize interpersonal forgiveness as a transactional process between two individuals. As such, it has multiple determinants, which include characteristics of the forgiver, the transgressor, the relationship, and the offense. As discussed below, this concept has implications for interpretation of scores on forgiveness measures and may provide a rationale in many research contexts for multimodal measurement procedures.

REVIEW OF LITERATURE: CONSTRUCT VALIDITY AND MULTIMODAL MEASUREMENT

The process of construct validation addresses the fundamental question, posed by Cronbach and Meehl (1955): What constructs account for variance in test performance? (p. 282). The use of the plural *constructs* embodies the fundamental insight that scores on any psychological test (or questionnaire, or behavioral rating scale) contain *surplus meaning*—that is, they inevitably reflect systematic variance in one or more characteristics of the respondents, in addition to standing on the construct they are intended to measure. This is easy to see in the case of domain-specific achievement tests. Reliable differences on such tests must reflect reading ability and perhaps differences in proficiency with the item formats, as well as familiarity with the content domain under investigation. These unwanted components of systematic variance, which we refer to as *bias variance*, make interpretation of scores on a test problematic.

Sources of Bias in Forgiveness Measures

In this and the following section, we develop a conceptual framework for exploring bias in forgiveness measures. In the absence of such a framework, investigators are unlikely to be motivated to make use of multiple measures and may fail to make best use of the data from multiple indicators when they do take the trouble to include them in their research designs.

Forgiveness measures (in common with most personality ratings) contain a form of bias not present in ability tests, namely, distortions in ratings due to idiosyncratic characteristics or motives of the rater. These *rater biases* are different for self- and other-ratings. Because most forgiveness measures consist of self-ratings (McCullough, Hoyt, & Rachal, 2000), we focus on self-reports here. Readers interested in biases in ratings of others are referred to previous work in this area (Hoyt, 2000; Hoyt & Kerns, 1999).

Bias in self-ratings has been studied under the rubric of *response sets*—styles of responding that contribute systematic but construct-irrelevant variance to ratings.

Well-known response sets include acquiescence and social desirability. *Acquiescence* variance is characterized by differences in the extent to which people tend to agree that descriptors (regardless of content) apply to themselves. To minimize confounding with acquiescence, many forgiveness measures include at least some reverse-scored items.

Social desirability variance is attributable to differences in the degree to which respondents modify their self-reports so as to make a good (or socially desirable) impression. Social desirability bias contributes more strongly to score variance for attributes that are culturally valued (e.g., intelligence) than for those that are value-neutral (e.g., perfectionism).

Although scholars debate about the validity of measures that purport to control for these biases (McCrae & Costa, 1983; Ones, Viswesvaran, & Reiss, 1996), the important (and uncontroversial) point is that self-reports are contaminated to an unknown degree with construct-irrelevant variance, referred to by Campbell and Fiske (1959) as *method variance*. A corollary is that high correlations between two sets of scores “might be explained as due either to basic trait similarity or to shared method variance” (p. 85), that is, to the tendency to evoke similar biases on the part of respondents. Thus, Campbell and Fiske emphasized the importance of convergent validity across methods (and discriminant validity with measures of different traits by similar methods) as criteria for construct validation.

Multiple Measures Versus Multimodal Measurement

At this point, we can define *multimodal measurement* as inclusion of two or more measures of a construct by methods (or *modes*) of measurement that are different enough to provide unique perspectives on the construct of interest. What constitutes “different enough” methods depends on the research question and will be the subject of detailed consideration in connection with the examples presented below.

As we will see, most of the studies in the forgiveness literature that have employed multiple measures have not been true multimodal studies, because the two (or more) measures (usually self-reports) share bias variance as well as construct-relevant variance. Such multiple-measure, mono-method studies are an improvement over single-measure research designs, in that they may include different facets of the complex forgiveness construct. For example, studies of transgression-related interpersonal motivations (e.g., McCullough & Hoyt, 2002) typically examine increases in benevolence motives and decreases in avoidance and revenge motives. Similarly, Enright's Forgiveness Inventory (EFI; Subkoviak, Enright, Wu, & Gassin, 1995) attempts to measure forgiveness as increases in positive affect, cognition, and behavior regarding a transgression, as well as reductions in negative affect, cognition, and behavior through a set of self-report items. However, measures such as those developed by McCullough and Hoyt (2002), and Enright (Subkoviak et al., 1995) do not do much to eliminate the confounding bias variance, because this variance is probably common to each of the subscales on these measures.

REVIEW OF LITERATURE: WHAT ARE "METHODS" OF MEASUREMENT?

Campbell and Fiske (1959) were intentionally vague in defining *method*, preferring to construe this term broadly to encompass a wide range of measurement procedures (Fiske & Campbell, 1992). Investigators seeking to study method variance have examined both fine and coarse differences in these procedures as sources of variance in personality scores. Many published studies using techniques developed by Campbell and Fiske examine method differences (e.g., different question formats in two self-report measures) that are probably too trivial to constitute different *modes* of measurement, in the sense discussed here.

Our field is at an early stage in its appreciation of the effects of method, and it is not yet possible to give comprehensive guidelines for determining when two methods are different enough to comprise a multimodal assessment of a construct such as forgiveness. The critical consideration is that the variance shared among the two measures should be valid variance, not bias variance. That is, the two measures should share construct-relevant (or *common factor*) variance but contain different sources of method (or *specific factor*) variance. In this section, we suggest two broad categories of measurement differences that could be considered by investigators seeking to use multimodal measurement: the source of the measured data and the level of measurement.

Source Differences

A method factor that is likely to contribute substantially to variance in ratings is the source of the information about forgiveness. Personality psychologists have long been exhorted to collect data from multiple sources (Cattell, 1957). Cattell's enthusiasm for a diversity of data sources has been commemorated with the LOTS acronym, reminding researchers of four broad classes of personality information on: life events data (L data); observational data (O data); test data (T data); and self-reports (S data; John & Benet-Martinez, 2000).

Life events data have the advantage that they are obtained objectively from life records. Although not used, to date, by forgiveness researchers, L data can provide one perspective on forgiveness in some populations. For example, in a study of divorced couples, compliance with alimony payment schedule may be related to level of forgiveness (among other things).

Observational data may be collected in a variety of settings and from a variety of sources. Forgiveness researchers have occasionally used O data—usually from acquaintance ratings of participants' forgivingness. For example, Brown (2003) validated a dispositional measure of forgiveness by documenting convergence with romantic partner ratings. Acquaintances have the advantage of depth of knowledge of the participants, but their ratings can be relatively unreliable because (a) acquaintances may have idiosyncratic interpretations of scale items, which usually cannot be mitigated

by rater training; (b) acquaintances are unique to each participant, which increases bias variance; and (c) it is difficult to aggregate ratings across many observers because acquaintances are difficult to recruit (Hoyt, 2000). To our knowledge, no one has made use of O data derived from trained observers, laboratory settings, and/or standardized interaction partners (confederates) to measure forgiveness, although Ripley and Worthington (2002) did use observational ratings of couple interactions to assess the ratio of positive to negative communication behaviors following a forgiveness intervention. These procedures enhance reliability of measurement, at some cost to generalizability of ratings to naturalistic interactions.

Test data are based on laboratory situations that yield objective scores on the variable of interest. T data on forgiveness could be derived from several existing laboratory tests, such as defections in the Prisoner's Dilemma game (Kassinove, Roth, Owens, & Fuller, 2002) or laboratory analogs of aggressive responding in response to a simulated provocation (e.g., Anderson & Bushman, 1997; Bushman & Baumeister, 1998).

Finally, *self-report data* are data from questionnaires (or interviews) in which participants describe their own attitudes or behavior. As already noted, self-reports are by far the most common source of data in forgiveness research. S data are also a common source for other variables that researchers believe are correlated with forgiveness. Correlations between two variables are difficult to interpret when both are measured by self-report because they reflect trait correlations but also (and to an unknown degree) method covariance. To avoid "mono-operation bias" (Cook & Campbell, 1979, p. 65), it behooves forgiveness researchers to consider augmenting their measurement of forgiveness with data from L, O, and T sources.

Levels of Analysis of Forgiveness: Offense, Relationship, and Individual

Although method variance is often described disparagingly as "nuisance" variance or statistical "noise" (e.g., Lubinski & Dawis, 1992), Campbell and Fiske (1959) left open the possibility that methods may be worthy of study in their own right (see also Cronbach, 1995). In the domain of forgiveness research, causal determinants at multiple *levels of analysis* are important for our theoretical understanding of forgiveness yet contribute to variance in forgiveness scores that in some research contexts is irrelevant to the construct of interest.

Forgiveness processes are studied at multiple levels. As a behavior, forgiveness is inherently linked to a particular event or *offense*. A person hurts or offends me in some way, and I react with changes in my feelings, attitudes, and motivations toward him or her. My (eventual) willingness to forgive this transgression is based in part on numerous situation-level details, including the severity of the offense, its intentionality, and the transgressor's willingness to apologize and make amends. Each of these factors contributes to variance in scores on transgression-based measures of forgiveness (McCullough et al., 2000).

Forgiveness may also be studied as a *relational* or *dyadic* process. Some relationships are more forgiving than others, and people's willingness to forgive a particular relationship partner is likely to be at least somewhat consistent across unique offenses that occur within that relationship.

Finally, forgiveness at the *individual* level refers to individual differences in the disposition to forgive that are at least somewhat consistent across relationships and offenses within relationships. The premise underlying the many studies of individual-level correlates of forgiveness (e.g., correlations with measures of personality, psychopathology, or well-being) is that some people are more forgiving (across a variety of relationships and specific transgressions) than are others and that this forgiving disposition is rooted in stable personality traits and has consequences for their mental health.

Considering the appropriate level of analysis is crucial to evaluating the construct validity of forgiveness measures. The *level of measurement* of the variables must correspond to the *level of analysis* of the constructs in the research hypothesis. Scores derived from a different measurement level inevitably contain substantial construct-irrelevant variance. For example, if an investigator hypothesizes that more forgiving couples are likely to be more satisfied and committed, the hypothesis is at the relationship level of analysis. What happens if he or she tests this hypothesis with variables measured at the situational level (i.e., with an offense-specific measure of forgiveness)? Forgiveness scores based on a single transgression would then serve to operationalize relational (or dyadic) forgiveness. But these scores differ from the *typical* forgiveness in the relationship to the extent that, for some or all couples, the specified offense differs from the typical offense for that couple. Any transgression-specific variance in scores, which would be valid variance for research at the level of specific offense, is nuisance variance in the context of the present research hypothesis and reduces the validity of measurement (by reducing the proportion of variance in scores that is attributable to general tendencies to forgive the relationship partner).

To enhance validity in this example, we could aggregate forgiveness scores based on two or more specific offenses. Using these two or more sets of offense-specific forgiveness scores as indicators of relationship-level forgiveness constitutes *multimodal* measurement in the sense described above: These measures share construct-relevant variance (i.e., variance in consistent tendencies to be forgiving or unforgiving toward the partner) but not method variance (i.e., deviations from these stable tendencies attributable to temporal or offense-specific factors). Although the correlation between forgiveness scores over any arbitrarily selected pair of offenses within a relationship is likely to be modest, by aggregating a number of such modestly correlated measures, we may obtain a highly reliable (and valid) composite index of relational forgiveness (Lubinski & Dawis, 1992; see also McCullough & Hoyt, 2002).

Summary of Recommendations

In this section, we have suggested that investigators wishing to use multimodal measurement could think about two different conceptual schemes that represent diverse

measurement modes for forgiveness research. One scheme (LOTS) identifies four distinct sources of data on forgiveness; including data from more than one source constitutes multimodal measurement. The other scheme identifies three levels of analysis (individual, relationship, and offense-specific) that are relevant to the study of forgiveness; an alternative strategy is to collect data of multiple types at a lower level of analysis (e.g., relationship-level data with friends, parents, and romantic partner) as a multimodal approach to assessment of forgiveness at a higher level of analysis (e.g., dispositional forgiveness, which should be consistent across these different relationship types). Below, we consider how to work with multimodal data to address questions of interest to forgiveness researchers.

MAKING BEST USE OF MULTIPLE MEASURES: ANALYTICAL STRATEGIES AND EXAMPLE STUDIES

We have already discussed one strategy for using data from multiple measures of forgiveness: aggregation to create a composite variable. This is advantageous because the composite will have a higher proportion of valid variance (and lower proportion of error variance) than its component variables. When component variables are measured by different methods (i.e., in the case of multimodal measurement), the proportion of method variance is also lower in the composite than in its components. A related approach with similar benefits is to treat forgiveness as a latent variable in a structural equation model with two or more measures as multiple indicators. In this section, we present alternatives to conventional aggregation approaches. These may be useful when different aspects of forgiveness are theorized to load on the two (or more) measures or when the measures use different methods, and the role of method variance is of substantive interest. When two or more measures embody different aspects of forgiveness, it can be instructive to analyze them separately, comparing patterns of correlations with criterion variables. When method variance is of substantive interest (Cronbach, 1995), a number of statistical models can be used to explore the role of method in determining ratings of forgiveness. We describe three such models: the multitrait multimethod matrix (MTMM), generalizability theory (GT), and the social relations model (SRM). When possible, we illustrate with an example drawn from the literature on forgiveness.

Extrinsic Convergent Validation: Exploring Patterns of Correlations with Criterion Variables

Fiske (1971) noted that a strong correlation between two measures is necessary but not sufficient evidence of their conceptual equivalence. Fiske recommended that, in addition to reporting convergent validity correlations between two measures thought to assess the same construct, investigators should examine their patterns of *extrinsic*

convergent validity (ECV)—that is, their patterns of correlation with criterion variables. When these patterns are highly similar for both measures, a strong argument can be made for empirical equivalence. (See Lubinski, 2004, p. 99, for an example of an argument via ECV analysis for the equivalence of three measures of verbal intelligence.) When two similar measures have different patterns of correlation, this can be important for elaborating the theory of the construct being assessed.

ECV analysis has the virtue that it is methodologically simple (involving straightforward comparison of correlation coefficients) but conceptually powerful. An example of the use of ECV to study forgiveness is the research program by McCullough and colleagues investigating a motivational model of forgiveness.

McCullough et al. (1998) equated forgiveness with decreases in motivation to avoid the perpetrator and to obtain revenge on him or her. McCullough et al. reported avoidance-revenge correlations between .4 and .5 in two samples, and comparable (negative) correlations with a single-item measure of forgiveness. However, subsequent studies revealed different patterns of correlation with relationship satisfaction and commitment in heterosexual couples. Most notably, male partners' revenge (but not avoidance) motives (negatively) predicted their female partners' relationship satisfaction (McCullough et al., 1998). Avoidance and revenge motives were also differentially related to transgression-relevant variables, with avoidance motives more strongly ameliorated by offender apology and victim empathy for the offender than are revenge motives (McCullough et al., 1998). When examined at the individual level of analysis, these motives were linked differentially to broad personality styles, with dispositional avoidance consistently uniquely predicted by neuroticism and dispositional revenge uniquely (negatively) predicted by agreeableness (McCullough & Hoyt, 2002). These findings confirm that, although they share considerable common variance (up to 25%), the avoidance and revenge scales measure distinct dimensions of forgiveness that have different correlates and different implications for relationship partners.

Multitrait Multimethod Matrix

The multitrait multimethod matrix (MTMM; Campbell & Fiske, 1959) has proven to be far and away the most popular tool for examining method variance in psychological measures. However, we do not recommend this method without reservation. Campbell and Fiske's guidelines for analysis are impressionistic rather than quantitative, and alternative procedures for conducting MTMM analyses using confirmatory factor analysis are temperamental, with models often failing to converge (Kenny & Kashy, 1992). Below, we discuss two generalizability-based approaches that we believe have more potential to address methodological and substantive issues for researchers using multimodal measurement.

Generalizability Theory: Decomposing Score Variance

Cronbach's inquiries into the meaning of test scores in the 1940s and 1950s eventually led to the development of another analytic framework for understanding the role of method variance in psychological measures. Although originally conceived as a liberalization of the assumptions of classical reliability theory (Cronbach, Rajaratnam, & Gleser, 1963), generalizability theory (GT) eventually emerged as a broad analytic framework for investigating the generalizability of scores across different measurement conditions, or *facets* (Cronbach, Gleser, Nanda, & Rajaratnam, 1972). As such, GT straddles the conventional boundary between classic reliability and validity theory, and provides a flexible tool for investigating sources of variance that contribute to scores derived from a variety of measurement procedures.

Probably partly as a consequence of its flexibility, GT is underutilized in contemporary psychological research. GT does not lend itself to cookbook "default" applications that are the norm in statistical software packages. Researchers wishing to use these techniques must learn to use a specialized computer application (GENOVA; Crick & Brennan, 1983; free download at <http://www.uiowa.edu/~itp/pages/SWGENOVA.SHTML>). Although helpful primers are available for using GT to estimate reliability of measurement (e.g., Hoyt & Melby, 1999; Shavelson & Webb, 1991), few attempts have been made to extend GT principles to examine score validity. We here present some initial thoughts on applicability of GT techniques to addressing questions of interest to forgiveness researchers.

In our discussion of levels of analysis for studying forgiveness, we noted that forgiveness occurs in response to discrete offenses but that responses to multiple offenses within a given dyad are likely to show some consistency, such that forgiveness can also be considered as a characteristic of the dyad or relationship. Further, it is likely that a person who is relatively forgiving in one relationship will also be forgiving in others, so that study of forgiveness at the individual (dispositional) level will also be fruitful. McCullough et al. (2000, table 4.1) showed how this three-tiered hierarchy (offenses embedded in streams of dyadic interactions embedded in individuals' responding in a variety of interpersonal contexts) could be modeled as a GT design. For each participant in the proposed study, six offenses are studied. The offenses encompass three of that person's relationships, with two of the offenses having been perpetrated by each of the three relationship partners. The data for the study are the participant's feelings of forgiveness in response to each of the six offenses.

Note that although each of the six measures is a self-report of willingness to forgive, measurement procedures in this research design are multimodal in the sense described above. Scores on each of the six forgiveness measures contain variance attributable to dispositional forgiveness, and the six measures include different subsets of method variance due to relationship partners and situations. McCullough and Hoyt (2002) carried out a version of this research design in two studies, operationalizing relationship as *relationship type* (in Study 1, for example, each participant reported forgiveness of same-sex friend, opposite-sex friend, and romantic partner) and offense

as a function of the *severity* of the transgression (so that each participant reported on one severe and one mild offense in each relationship type).

Meaning of Variance Components. The purpose of the generalizability analysis (in McCullough & Hoyt, 2002) was to *partition variance* in forgiveness ratings into variance attributable to four main factors of interest. The relative size of each variance component reflects the importance of the corresponding determinant of forgiveness. To give a sense of the types of questions that can be addressed in a GT analysis, we first describe the meaning of each variance component for this study design, then present a brief summary of our findings.

Person variance (P) reflects the extent to which forgiveness in response to a given offense is predictable from the victim's general disposition to forgive others (across offense severity and relationship type). If *P* variance is large, this indicates that people are relatively consistent in their forgivingness across relationships and offenses. If *P* variance is small, then willingness to forgive is strongly conditioned on relational or situational factors, and we look to the remaining variance components to gauge the relative importance of these factors.

Variance attributable to the *Person × Relationship Type interaction (PR)* reflects the extent to which persons' rank (on forgivingness) differs as a function of relationship type. If *PR* variance is large, then persons who are most forgiving in one relationship may not be so forgiving in another. (Knowing that someone is very forgiving of her romantic partner, for example, would not help you to predict her forgivingness of a same-sex friend.)

Variance attributable to the *Person × Severity interaction (PS)* reflects the extent to which persons' rank (on forgivingness) differs in response to moderate or severe offenses. If *PS* variance is large, then people who are the most forgiving of a minor offense may not be the most forgiving of a more severe hurt.

Finally, variance attributable to the three-way interaction (*Person × Relationship × Severity*) is notated *PRS,e* to reflect the limitation that the highest order interaction is confounded with error in any generalizability study. This is because there is only one data point per combination of these three factors, so we cannot separate stable (reliable) variance from random error. Thus, variance attributable to the highest order interaction in any generalizability study is referred to as *residual* variance and reflects variance due to random error, the three-way (in this case) interaction, and interactions with other facets (measurement conditions) not explicitly measured in this research design.

GT analyses also produce variance estimates for the main effects of relationship type and severity (i.e., *R* and *S*) and for the interaction of these facets (i.e., *RS*). These are of theoretical interest as well. The *S* main effect, for example, will be large if (as expected) people are generally less forgiving of severe offenses than of mild ones. We focus on variance components involving persons (*P*, *PR*, *PS*, and residual) because of their importance for understanding the consistency (or lack thereof) in

people's willingness to forgive—that is, the importance of dispositional forgiveness in determining responses to specific offenses.

Summary of Findings. Using the motivational framework described above to measure forgiveness, McCullough and Hoyt (2002) found that persons accounted for 25% and 44% of variance in avoidance and revenge motivation, respectively, in Study 1. Thus, people showed much more consistency in their vengeful motivations than in their avoidant impulses. Relationship type was one factor in this inconsistency (*PR* accounted for 13% of variance in both avoidance and revenge): People who were most forgiving of same-sex friends, for example, were not necessarily most forgiving of romantic partners. *PS* accounted for little variance in either measure (i.e., persons who were relatively forgiving of severe transgressions were also relatively forgiving of mild transgressions) but residual variance was large (37% and 36% for avoidance and revenge, respectively), undoubtedly reflecting, at least in part, other sources of variance in response to single offenses (e.g., apology, intentionality) that were not investigated here.

These findings are important in that they suggest that revenge motivation is more strongly influenced by dispositions (i.e., more consistent across offenses and relationships) than is avoidance motivation. Additional analyses provided guidance for future users of scenario measures, presenting G coefficients (analogous to reliability coefficients) for composites derived from aggregating forgiveness scores across scenarios, and examining congruence between forgiveness scores derived from historical offenses (discussed above) with those from fictional offenses (not described here). By conducting variance partitioning analyses, we learn about the relative importance of various determinants of forgiveness.

Social Relations Model: A Special Case of GT for Round-Robin Designs

A final model useful for analyzing multimodal measures of forgiveness is the social relations model (SRM; Kenny, 1994). Like GT, SRM is a variance partitioning model, with the advantage that it incorporates ratings from both members of a dyad. SRM data must be collected in round-robin format (each person in a group rates each other person), so it is most useful for studying naturally occurring groups, such as friendship groups, roommates, or families. Like GT, SRM requires specialized software (SOREMO; Kenny, 1987; download for a fee at <http://users.rcn.com/dakenny/srmp.htm>), which may be one reason why it is underused in social and personality research.

To highlight the potential of SRM for studying forgiveness, we present selected findings from a study by Hoyt, Fincham, McCullough, Maio, and Davila (in press; Study 1) who conducted SRM analyses on data from 94 families (father, mother, and daughter in 8th grade), each of whom reported her or his general level of forgiveness to the other two family members.

SRM analyses partition variance in dyadic ratings into actor, partner, relationship, and error variance. *Actor variance*, which accounted for more than 50% of variance in most dyads, reflects consistency in forgiving across relationships (i.e., *forgiveness*). *Partner variance* reflects individual-level differences in *forgivability*—a construct that has been little researched. Partner variance was significant for mothers and children but not for fathers. Thus, whereas there was no evidence of individual differences in forgivability among fathers, some mothers (and some children) were perceived as more forgivable than others in the eyes of other family members. *Relationship variance* indicates unique forgiveness (or unforgiveness) toward a relationship partner, controlling for the victim's general forgiveness and the offender's general forgivability. Relationship variance was high only for mothers' ratings of fathers (accounting for 75% of variance in mothers' ratings of fathers), indicating unique adjustments in wives' forgiveness toward their husbands based on relational factors.

The SRM framework allows for examination of many phenomena of interest in dyadic relationships. *Reciprocity* addresses whether people who are forgiving also tend to be forgiven. SRM examines reciprocity at both the individual level (where we found evidence of reciprocity for both mothers and children) and the dyadic level (where we found no evidence of reciprocity, as expected based on the paucity of relationship variance). In addition to collecting data on forgiveness in multiple relationships, we also examined both victim and offender perspectives on forgiveness, which allows for checks on congruence among scores from different data sources (i.e., self-reports and other-reports). Congruence correlations indicated moderate to strong levels of self-other agreement at both individual and dyadic levels of analysis.

RELEVANCE FOR CLINICAL AND APPLIED INTERVENTION

The importance of multimodal assessment will not be surprising to practitioners. Every couples counselor is aware that relationship partners have differing perspectives on both positive and negative relationship events, and that both perspectives are important. Questions about generalizability of measurement also bear formal similarity to questions about generalizability of treatments. For example, if I assist an adult victim of child sexual abuse to forgive the abuser, I may believe that this process has value in itself and that it serves as a model for how current relationships can be repaired when trust is damaged. Yet I recognize that generalizability is inherently limited and that forgiveness is a function of particularities of a given relationship as well as of the skill, perceptiveness, and willingness of the forgiver.

For practitioners relying on research-supported theories or measures of forgiveness, the limitations of studies using unimodal measurement are useful to keep in mind. Thus, when self-reported forgiveness is found to be correlated with other self-report measures (e.g., depression, well-being), readers should be aware that these correlations reflect correlated method variance as well as trait variance, so that the true strength of association between traits is obscured. On the other hand, when forgiving

for a single offense (e.g., a relationship conflict) is correlated with indicators of well-being, there is a mismatch in level of measurement between predictor and criterion variables. Because forgiveness scores contain substantial situational and dyadic variance, they will not correlate as highly with well-being as would a trait-level forgiveness measure. Thus, the source and level of measurement are important to keep in mind when interpreting both theoretical and applied research findings.

CONCLUSION

In this chapter, we reviewed the literature on construct validity, focusing on the problem of surplus meaning (or method variance), to provide a rationale for the use of multimodal measurement by forgiveness researchers. We presented two category systems (source of data and level of measurement) that may assist researchers in selecting multiple measures that contain common trait variance but distinct sources of method variance. We described the benefits of aggregation across measurement methods and presented several specialized analytic techniques to examine variations in research findings across different measures and measurement methods. We hope that this conceptual framework will be useful to both researchers and consumers of research as they consider how to design studies and interpret study findings. We hope the analytical toolkit presented here will provide a stimulus to forgiveness researchers to move beyond mono-operation bias in creative ways to advance both our psychometric and substantive knowledge about forgiveness processes.

REFERENCES

- Anderson, C. A., & Bushman, B. J. (1997). External validity of "trivial" experiments: The case of laboratory aggression. *Review of General Psychology, 1*, 19–41.
- Brown, R. P. (2003). Measuring individual differences in the tendency to forgive: Construct validity and links with depression. *Personality and Social Psychology Bulletin, 29*, 759–771.
- Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology, 75*, 219–229.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.
- Cattell, R. B. (1957). *Personality and motivation structure and measurement*. Oxford, England: World Book.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Crick, J. E., & Brennan, R. L. (1983). *GENOVA: A generalized analysis of variance system*. Iowa City, IA: American College Testing.
- Cronbach, L. J. (1995). Giving method variance its due. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: A festschrift honoring Donald W. Fiske* (pp. 145–157). Hillsdale, NJ: Lawrence Erlbaum.

- 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.
- Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, *16*, 137–163.
- Fiske, D. W. (1971). *Measuring concepts of personality*. Chicago: Aldine-Atherton.
- Fiske, D. W., & Campbell, D. T. (1992). Citations do not solve problems. *Psychological Bulletin*, *112*, 393–395.
- Heppner, P. P., Casas, J. M., Carter, J., & Stone, G. L. (2000). The maturation of counseling psychology: Multifaceted perspectives, 1978–1998. In S. D. Brown, & R. W. Lent (Eds.), *Handbook of counseling psychology* (3rd ed., pp. 3–49). New York: Wiley.
- Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, *5*, 64–86.
- Hoyt, W. T., Fincham, F., McCullough, M. E., Maio, G., & Davila, J. (in press). Responses to interpersonal transgressions in families: Forgiveness, forgivability, and relationship-specific effects. *Journal of Personality and Social Psychology*.
- Hoyt, W. T., & Kerns, M. D. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods*, *4*, 403–424.
- Hoyt, W. T., & Melby, J. N. (1999). Dependability of measurement in counseling psychology: An introduction to generalizability theory. *Counseling Psychologist*, *27*, 325–352.
- John, O. P., & Benet-Martinez, V. (2000). Measurement: Reliability, construct validation, and scale construction. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 339–369). New York: Cambridge University Press.
- Kassinove, H., Roth, D., Owens, S. G., & Fuller, J. R. (2002). Effects of trait anger and anger expression style on competitive attack responses in a wartime prisoner's dilemma game. *Aggressive Behavior*, *28*, 117–125.
- Kenny, D. A. (1987). *SOREMO: A FORTRAN program for round robin data structures* [computer program]. Storrs: University of Connecticut.
- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, *112*, 165–172.
- Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 Years after Spearman's (1904) "General intelligence, objectively determined and measured." *Journal of Personality and Social Psychology*, *86*, 96–111.
- Lubinski, D., & Dawis, R. V. (1992). Aptitudes, skills and proficiencies. In M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 3, 2nd ed., pp. 1–59). Palo Alto, CA: Consulting Psychologists Press.
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, *51*, 882–888.
- McCullough, M. E., & Hoyt, W. T. (2002). Transgression-related motivational dispositions: Personality substrates of forgiveness and their links to the Big Five. *Personality and Social Psychology Bulletin*, *28*, 1556–1573.
- McCullough, M. E., Hoyt, W. T., & Rachal, K. C. (2000). What we know (and need to know) about assessing forgiveness constructs. In M. E. McCullough, & K. I. Pargament (Eds.), *Forgiveness: Theory, research, and practice* (pp. 65–88). New York: Guilford Press.

- 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., & Snyder, C. R. (2000). Classical source of human strength: Revisiting an old home and building a new one. *Journal of Social and Clinical Psychology*, *19*, 1–10.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Washington, DC: American Council on Education.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*, 660–679.
- Ripley, J. S., & Worthington Jr., E. L. (2002). Hope-focused and forgiveness-based group interventions to promote marital enrichment. *Journal of Counseling and Development*, *80*, 452–463.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Thousand Oaks, CA: Sage.
- Subkoviak, M. J., Enright, R. D., Wu, C., & Gassin, E. A. (1995). Measuring interpersonal forgiveness in late adolescence and middle adulthood. *Journal of Adolescence*, *18*, 641–655.