Marquette University
e-Publications@Marquette

10-1-2014

Advancing the Measurement of Violence: Challenges and Opportunities

John H. Grych
Marquette University, john.grych@marquette.edu

Sherry L. Hamby
Sewanee: The University of the South

INTRODUCTION

Advancing the Measurement of Violence: Challenges and Opportunities

John Grych
Marquette University

Sherry Hamby
University of the South

Our understanding of the causes and consequences of violence depends on accurately defining and measuring the constructs we study. Although the methods used most often in violence research have led to a wealth of important findings, the field is ripe for both reflection and innovation. The purpose of this special issue is to highlight critical measurement issues in the study of violence and to describe innovative approaches that will move this research forward. In this Introduction to the special issue, we identify 3 challenges for the valid measurement of violence—defining constructs, accurately capturing responses in scoring, and diversifying measurement methods—and discuss how the 8 studies that constitute the issue address these challenges and identify promising directions for future work.

Keywords: violence, measurement, methodology, assessment, research

The study of violence was transformed in the 1970s by the realization that people would disclose perpetrating violence against and being victimized by family members on confidential self-report surveys. Early measures like the Conflict Tactics Scales (Straus, 1979; Straus, Hamby, Boney-McCoy, & Sugarman, 1996) represented major methodological advances because they provided reliable and practical tools for looking behind closed doors to document the level of violence occurring in American families. This enabled researchers to move beyond reliance on crime statistics and police reports and investigate new questions about the causes and consequences of interpersonal violence. Since then, there has been a proliferation of self-report measures of an array of constructs pertinent to violence. Experimental paradigms for studying violence also have been utilized, but psychological research on violence has been dominated by the use of retrospective self-report questionnaires. Although this approach has produced a great deal of knowledge about interpersonal violence, relying so heavily on a single type of measurement has constrained what is studied in violence research, limited understanding of important aspects of interpersonal violence, and produced unresolved controversies, such as questions about gender patterns in intimate partner violence (Hamby, 2014).

All fields require innovation to move forward, and the purpose of this special issue is to explore questions about the measurement of violence that are critical for advancing our understanding and identifying new directions for future work. The eight articles in the issue were selected from more than 50 submissions; many of these articles were excellent, and several that could not be included in the special issue will appear in future regular issues of the journal. The articles that follow include both conceptual and empirical studies and examine a range of methodological issues in researching violence, which we organize around three key challenges: defining constructs precisely, accurately capturing disclo-

John Grych, Department of Psychology, Marquette University; Sherry Hamby, Department of Psychology, University of the South.

Correspondence concerning this article should be addressed to John Grych, Department of Psychology, Cramer Hall, Marquette University, Milwaukee, WI 53201-1881. E-mail: john.grych@marquette.edu

sures of violence, and diversifying measurement strategies.

The Challenge of Operationalization: Increasing Precision

Our understanding of a phenomenon can be no more precise than the measures used to assess it. The measurement tools and strategies available to researchers determine what is studied and how the findings of those studies are interpreted. Increasing the precision of measurement enables us to assess constructs in a more nuanced and accurate way and to investigate more complex conceptual questions about specific aspects of violence. There are many underappreciated elements of the task of measurement, and one particularly underappreciated element comes before a single item is written or a single survey is administered: the task of defining or operationalizing the constructs that will be investigated. Because of time pressure and resource pressure, this step may be rushed over, with choices about measurement guided primarily by the availability of and precedence for using particular questionnaires in prior work. Although there is much to be said for building on prior research by utilizing the same or similar measurement strategies, this practice also can limit the scope of the questions that are investigated, discourage innovation, and lead to reification of the most well-established measures.

Follingstad and Bush (2014, pp. 369–383) have carved out some space to consider the shape and form of a "Cadillac" model of measurement. They did not have to look far for models. As we have written elsewhere (Hamby & Grych, 2013), different disciplines in the social and clinical sciences have much to learn from each other, and Follingstad and Bush draw from the American Psychiatric Association's working groups model that is used for the Diagnostic and Statistical Manual (DSM; American Psychiatric Association, 2013), among other projects. The DSM is another project that has created its share of controversy and that, too, offers potential for learning, both in terms of avoiding missteps and withstanding the intense gaze of large groups with competing interests. The scope of Follingstad and Bush's proposal is, frankly, daunting. However, the importance of their contribution lies in the very dauntingness of the steps they propose for creating a "gold standard" measure of IPV, which are both time and resource intensive. Few concepts in psychology, relevant to violence or not, have undergone the careful and systematic development work that Follingstad and Bush propose for assessing intimate partner violence, and their article provides a valuable illustration for how to thoughtfully and rigorously operationalize a complex construct like IPV.

Leonard, Winters, Kearns-Bodkin, Homish, and Kubiak (2014, pp. 384–398) explore the meaning of a related construct that is often studied but rarely well-defined in violence research: bidirectional or reciprocal partner violence. Violence between partners typically is termed "bidirectional" when both endorse engaging in any aggression toward the other, with the implication that both partners contribute roughly equally to the violence occurring in their relationship. This approach ignores differences in the rates, severity, and sequence of aggressive behavior between partners and results in a simplistic definition that obscures potentially important distinctions. Leonard and his colleagues offer a more nuanced and sophisticated look at this construct in a longitudinal study that assesses specific aggressive behaviors in newly married couples over the first two years of marriage. They find that "bidirectional" violence includes different patterns of interaction that vary according to gender and severity, and that most are asymmetrical rather than symmetrical in nature. That is, in most cases of "bidirectional" aggression, one partner engages in higher rates of aggression than the other. The study also highlights the value of longitudinal research in showing that these bidirectional patterns change over time, with most asymmetric groups becoming more symmetrical over the course of a year.

The introduction of new constructs often engenders controversy over their precise meaning and distinctiveness from similar concepts. Mehari, Farrell, and Le (2014, pp. 399–415) address definitional and conceptual issues regarding cyberbullying, which has emerged in a remarkably short time as a phenomenon that receives as much scholarly attention as many much more established topics of inquiry (Hamby, McDonald, & Grych, 2014). The study of cyberbullying provides another example of the time and resource pressures that can

come to bear on researchers; in a way, cyberbullying has become a victim of its own success. Because it has galvanized the attention of the scientific community and media alike, there has been a flurry of studies conducted on the topic, many of which use newly created measures that have not undergone systematic psychometric development. As Mehari et al. point out, there are significant unanswered questions about the exact nature of this phenomenon, the types of events that constitute cyberbullying, and, especially, its distinctiveness from in-person bullying. Their article is a fine-grained comparison of the features, such as anonymity and access, that vary across incidents that are perpetrated face-to-face or through the use of technology. Like Follingstad and Bush, they also envision the steps necessary to create a gold standard and to move forward with a cogent and coherent framework that guides research on this topic.

The Challenge of Scoring: Interpreting Disclosures Meaningfully

The second scientific challenge that is addressed in this special issue concerns the validity of scores obtained from measures of violence. Assessing violence is different than measuring many psychological phenomena, such as attitudes, personal characteristics, and behavioral tendencies, in that violent incidents are "count" phenomena; they are events that have occurred or not rather than qualities that people have more or less of. This approach has methodological advantages. For example, unlike many measures of individual differences, violence questionnaires have a true, meaningful zero point. As a result, behavioral data can be more readily interpretable than a score on an attitude scale (what does it mean to be a 25, e.g., on an extraversion scale?). However, this approach presents challenges, too. Some of these, such as skewed distributions due to the low base rate of violence in many samples, are well known and have statistical solutions.

A more fundamental challenge is to ensure that the behaviors reported by participants map on to the conceptualization of violence intended by the developers. Because terms such as "violence" and "abuse" can be interpreted differently by different people, questionnaires like the CTS were designed to reduce the influence of respondents' subjective perceptions by including only discrete, observable behaviors such as "hit," "kick," and "choke." The assumption behind such behavior checklists is that reports of concrete behaviors are valid indicators that those behaviors occurred, and implies that different reporters can agree on whether a behavior like a push or slap occurred. This assumption has been questioned since the first behavior checklists were developed, and the lower-than-expected concordance typically found between partner reports of relationship violence raises questions about what people are reporting when they report on the occurrence of specific behaviors. Discrepancies between partner reports typically are attributed to tendencies for perpetrators to underreport socially undesirable behaviors and memory limitations that undermine the accuracy of retrospectively recalling both perpetration and victimization, especially as participants are asked to recall events over long period of time.

Derrick, Testa, and Leonard (2014, pp. 416–431) examined partner concordance in reporting aggression using a daily diary methodology that reduces the recall burden on participants. They found that even when partners report on discrete behaviors occurring the prior day, they showed low levels of agreement. Further, they report that only participants' own reports of aggression—either perpetration or victimization—predicted their reports of relationship satisfaction and mood the following day, with events reported by both partners the strongest predictors. Their study raises questions about the extent to which reports of even seemingly straightforward behavioral items reflect the subjective perception of the reporter and suggests that the common practice of pooling reports across partners introduces more error than clarity to the measurement of violence.

Another implication of measuring the occurrence of violent events is that reports of these events are likely to be more independent than the items on a measure of attitudes or traits. For example, the items on a well-constructed measure of extraversion should be intercorrelated; if a person likes going to parties, they probably also like spending time with other people. However, just because one incident of violence has happened does not necessarily mean that another one has. This is all the more true on victimization surveys, when participants are re-

porting on the behavior of other people that the participant has encountered. Victimization surveys are, in many ways, getting at questions of the extent and chronicity of underlying vulnerabilities, and it is important to consider how these vulnerabilities might cluster to best represent the experiences of each participant. This is the question that Ford and colleagues have turned to in their article on the Adverse Childhood Experiences (ACE) scale (Ford et al., 2014, pp. 432–444). Like the concept of cyberbullying, ACE scores have captured the attention of a wide range of providers and policymakers who otherwise would not pay as much attention to violence research as we might wish. Part of the appeal likely lies with its simplicity—in its original form, the ACE scale is a simple checklist of 10 items (Felitti et al., 1998). The strong connection between reports of childhood adversity and adult health status has generated considerable interest and inspired important lines of research. However, the ACE scoring has, to date, received fairly little empirical study and questions have been raised about whether this single score of 10 items is the best way to capture childhood adversity (Finkelhor, Shattuck, Turner, & Hamby, 2013). Ford and colleagues use the large Behavioral Risk Factor and Surveillance System dataset to examine the factor structure of ACE items. They find that a 3-factor solution is a better representation of the experiences of participants than a single score. The factors that emerged—Household Dysfunction, Emotional/Physical Abuse, and Sexual Abuse—correspond to classic domains of interest in work on youth victimization and suggest that there is value to preserving some specificity as well as estimating a total adversity burden.

A third validity issue concerns the extent to which the scores produced by a measure fully capture the construct assessed. Most of the constructs studied in violence research are complex, yet questionnaires often produce a single score that ostensibly captures the meaning of the construct. Davis, Gilmore, et al. (2014, pp. 445–461) examine this issue in relation to the Sexual Experiences Study, one of the most widely used measures of sexual assault. The SES provides a wealth of detailed information about sexual assault history, including the frequency, severity, and tactics used during one or more assaults experienced by the reporter. Because the scale assesses multiple dimensions, it

can be scored in ways that emphasize different elements, and it is not clear which scoring algorithm best captures the information disclosed by respondents. Davis and her colleagues utilize an experimental design to systematically compare nine ways of scoring the SES that differentially weight dimensions such as frequency, tactics used, and outcomes. Their findings show that the scores obtained by different methods result in similar but not identical associations with other relevant variables, and indicate that scoring methods that emphasize different dimensions may be more appropriate for answering particular research questions or for use in particular samples.

The Challenge of Innovation: Diversifying Measurement Strategies

Although self-report questionnaires are the dominant method for assessing violence in contemporary research, other methodological approaches also have played important roles in studying violence. Expanding the strategies used to study violence provides opportunities to investigate new questions and to investigate old questions in new ways. The use of multiple sources of data to assess a phenomenon is a hallmark of good research design, and two of the articles in the special issue address measurement strategies that offer alternatives to structured questionnaires.

From the first punch landed on a Bobo doll to Milgram's shocking studies of obedience to the use of hot sauce as an instrument of aggression, experimental methods have provided invaluable insights into interpersonal aggression. Experimental methods offer the potential to test causal hypotheses, which is particularly important for determining which factors could be targeted effectively in interventions, but have been underutilized in the study of violence. Davis, Parrott, et al. (2014, pp. 462–476) provide a review and analysis of laboratory paradigms that have been used in violence research, addressing the potential benefits as well as conceptual, practical, and ethical challenges of using these methods. Their article describes creative approaches that provide valuable perspectives on aggression and shows how these methods can be part of an armamentarium of research strategies that, when utilized together, provide a more multifaceted look at the phenomena.

The term "measurement" often conjures images of quantitative research, and indeed quantitative measurement is the principal means by which we describe and systematize knowledge about violence (Hamby, 2014). The role of qualitative approaches to measurement is less widely appreciated, but qualitative research excels at exploring the nuances of constructs and has much to contribute to the measurement of violence. Quantitative research alone cannot, by its very nature, identify what is missing from survey content. Lerhner and Allen's article is a wonderful exemplar of what we can learn from unstructured incident narratives (Lehrner & Allen, 2014, pp. 477–490). Theirs is one of the first mixed methods studies of IPV measurement—surprising after more than four decades of research—and provides some of the most direct evidence to date that behavioral checklists such as the Conflict Tactics Scales are vulnerable to false positives; that is, endorsement of ostensibly aggressive behaviors that actually reflect joking or horseplay rather than an intent to harm. Although concerns about this issue have been discussed in the field for years, there have been very few empirical attempts to examine how closely the responses to violence questionnaires conform to the phenomena that researchers are trying to study.

The Centrality of Measurement: Strengthening the Foundation of Violence Research

As computer power has expanded and statistical methods have become increasingly sophisticated, it is easy to lose sight of some of the most basic and essential elements of science. Complex analytical models tell us little if we have not taken the time to assure that the measures generating the data are assessing what we intend them to assess. The thread that connects all of these studies is their attention to more fully and accurately representing how violence is conceptualized and measured. For the scientific study of violence to advance, we need to devote rigorous attention to how the central constructs are defined, assessed, and interpreted. The articles in this special issue identify promising directions for advancing the study of violence and provide an outline for envisioning the scientific frontier for each of the challenges described. As a group, they highlight the need to develop "gold standards" for assessing key concepts and diverse strategies for measuring complex phenomena. They argue for the importance of capturing distinct elements of multifaceted constructs and avoiding the impulse to oversimplify these constructs. To take an extreme example, many studies divide participants into dichotomous categories of "victims" versus "nonvictims" and "perpetrators" versus "nonperpetrators." Such blunt distinctions fail to capture important nuances of people's experiences or recognize the interconnections among different forms of violence. They also offer no insight into how individuals with extensive and severe victimization differ from those with the lowest levels of exposure, which has important implications for work in resilience and recovering from victimization (Grych, Hamby, & Banyard, 2014).

The public has trusted us with the disclosure of some of the most personal and challenging moments from their lives. We owe it to them to use this information wisely. These articles show that rigorous measurement will always be a fundamental aspect of violence research, as new technologies change the shape of violence, as we struggle to understand divergent findings from different data sources, as more ambitious research designs allow us to examine how these phenomena wax and wane over time. Our ultimate goals require a nuanced understanding of the phenomena of violence itself: a comprehensive understanding of the causes and consequences of violence, and the ability to reduce the incidence of violence and to ameliorate the consequences when it does occur. We cannot end up at our ultimate goals if we do not start out on the right track.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders,* 5th ed. Washington, DC: Author.

Davis, K. C., Gilmore, A. K., Stappenbeck, C. A., Balsan, M. J., George, W. H., & Norris, J. (2014). How to score the sexual experiences survey? A comparison of nine methods. *Psychology of Violence, 4,* 445–461. doi:10.1037/a0037494

Davis, K. C., Parrott, D. J., George, W. H., Tharp, A. T., Hall, G. C. N., & Stappenbeck, C. A. (2014). Studying sexual aggression: A review of the evolution and validity of laboratory paradigms. *Psy-*

chology of Violence, 4, 462–476. doi:10.1037/a0037662

Derrick, J. L., Testa, M., & Leonard, K. E. (2014). Daily reports of intimate partner verbal aggression by self and partner: Short-term consequences and implications for measurement. *Psychology of Violence, 4,* 416–431. doi:10.1037/a0037481

Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., . . . Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. *American Journal of Preventive Medicine, 14,* 245–258. doi:10.1016/S0749-3797(98)00017-8

Finkelhor, D., Shattuck, A., Turner, H., & Hamby, S. (2013). Improving the adverse childhood experiences study scale. *Journal of the American Medical Association pediatrics, 167,* 70–75. doi:10.1001/jamapediatrics.2013.420

Follingstad, D. R., & Bush, H. M. (2014). Measurement of intimate partner violence: A model for developing the gold standard. *Psychology of Violence, 4,* 369–383. doi:10.1037/a0037515

Ford, D. C., Merrick, M. T., Parks, S. E., Breiding, M. J., Gilbert, L. K., Edwards, V. J., . . . Thompson, W. W. (2014). Examination of the factorial structure of adverse childhood experiences and recommendations for three subscale scores. *Psychology of Violence, 4,* 432–444. doi:10.1037/a0037723

Grych, J., Hamby, S., & Banyard, V. L. (2014). *The Portfolio Model of Resilience: How "positive" psychology can inform the study of "negative" life events.* Sewanee, TN: Life Paths Research Program.

Hamby, S. (2014). Intimate partner and sexual violence research: Scientific progress, scientific challenges, and gender. *Trauma, Violence, & Abuse, 15,* 149–158. doi:10.1177/1524838014520723

Hamby, S., & Grych, J. (2013). *The web of violence: Exploring connections among different forms of interpersonal violence and abuse.* Dordrecht, The Netherlands: Springer. doi:10.1007/978-94-007-5596-3

Hamby, S., McDonald, R., & Grych, J. (2014). Trends in violence research: An update through 2013. *Psychology of Violence, 4,* 1–7. doi:10.1037/a0035384

Leonard, K. E., Winters, J. J., Kearns-Bodkin, J. N., Homish, G. G., & Kubiak, A. J. (2014). Dyadic patterns of intimate partner violence in early marriage. *Psychology of Violence, 4,* 384–398. doi:10.1037/a0037483

Lehrner, A., & Allen, N. E. (2014). Construct validity of the conflict tactics scales: A mixed-method investigation of women's intimate partner violence. *Psychology of Violence, 4,* 477–490. doi:10.1037/a0037404

Mehari, K. R., Farrell, A. D., & Le, A.-T. H. (2014). Cyberbullying among adolescents: Measures in search of a construct. *Psychology of Violence, 4,* 399–415. doi:10.1037/a0037521

Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics Scales. *Journal of Marriage and the Family, 41,* 75–88. doi:10.2307/351733

Straus, M. A., Hamby, S., Boney-McCoy, S., & Sugarman, D. (1996). The Revised Conflict Tactics Scales (CTS2): Development and preliminary psychometric data. *Journal of Family Issues, 17,* 283–316. doi:10.1177/019251396017003001

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at http://notify.apa.org/ and you will be notified by e-mail when issues of interest to you become available!