UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Assessing the reliability and validity of television and game violence exposure measures

Fikkers, K.M.; Taylor Piotrowski, J.; Valkenburg, P.M.

DOI
10.1177/0093650215573863

Publication date
2017

Document Version
Final published version

Published in
Communication Research

Link to publication

Citation for published version (APA):
Fikkers, K. M., Taylor Piotrowski, J., & Valkenburg, P. M. (2017). Assessing the reliability and validity of television and game violence exposure measures. *Communication Research*, *44*(1), 117-143. https://doi.org/10.1177/0093650215573863

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:16 Oct 2026

Communication Research
2017, Vol. 44(1) 117–143
© The Author(s) 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0093650215573863
journals.sagepub.com/home/crx
SAGE

Assessing the Reliability and Validity of Television and Game Violence Exposure Measures

Karin M. Fikkers[1], Jessica Taylor Piotrowski[1], and Patti M. Valkenburg[1]

Abstract

This study evaluated whether common self-report measures of television and game violence exposure represent reliable and valid measurement tools. Three self-report measures—direct estimates, user-rated favorites, and agency-rated favorites—were assessed in terms of test-retest reliability, criterion validity (their relationship with coded media diaries), and construct validity (their relationship with aggression and gender). A total of 238 adolescents participated in a two-wave survey and completed two media diaries. For game violence, the three self-report measures were reliable and valid. For television violence, only direct estimates achieved test-retest reliability and construct validity. Criterion validity could not be established for the television violence measures because the media diary was not a valid criterion for television violence. Our findings indicate that both direct estimates and favorites are valid measures for game violence, whereas for television violence, only direct estimates are valid. We conclude with a discussion about ways to further improve upon and reconceptualize media violence exposure measurement.

Keywords

aggression, games, reliability, television, validity

Media violence research has always been characterized by a certain degree of controversy, and even more so in the past decade (Busching et al., 2015; Kirsh, 2012).

[1]University of Amsterdam, The Netherlands

Corresponding Author:
Karin M. Fikkers, Amsterdam School of Communication Research (ASCoR), University of Amsterdam, P.O. Box 15791, 1001 NG Amsterdam, The Netherlands.
Email: k.m.fikkers@uva.nl

Although it is true that different studies have found different effect sizes of media violence on aggressive behavior, this range is easily surpassed by the range in *interpretations* of these effect sizes. Some researchers have compared the strength of the effect of media violence on aggression with the effect of smoking on lung cancer (Bushman & Anderson, 2001), while others maintain that media violence does not increase aggression at all (Ferguson, 2009). In this debate, researchers often refer to low quality of measurement as an explanation for small or large effects on aggression (Elson & Ferguson, 2014; Ferguson & Savage, 2012; Krahé, 2014). Many media violence studies use traditional self-report measures, which have often been criticized for their low reliability and validity. Yet, little research currently exists that can speak for or against the quality of self-report measures of media violence exposure used in this field. This lack of knowledge hinders a meaningful interpretation of and debate about the influence of media violence on aggressive behavior.

Our ability to detect and interpret effects of media violence exposure directly depends on the reliability and validity of its measurement. Reliability refers to "the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials" (Carmines & Zeller, 1979, p. 11), while validity pertains to whether a measure actually reflects the concept it is intended to measure (Carmines & Zeller, 1979). The consequences of using measures of low reliability and validity are substantial. Measures of low reliability introduce additional error variance into statistical models, which results in underestimated effect sizes or even null effects (Jordan, Trentacoste, Henderson, Manganello, & Fishbein, 2007; Lee, Hornik, & Hennessy, 2008; Prior, 2009). Low validity, in turn, makes it difficult to interpret any relationships found between an exposure variable and an outcome (Valkenburg & Peter, 2013b).

Remarkably, given the long history of the field, little research has been undertaken to assess the reliability and validity of media violence exposure measures that rely on self-report. Our current knowledge about the quality of existing measures consists mainly of published articles reporting internal consistency statistics of exposure measures, sometimes complemented by test-retest reliability statistics in longitudinal research. Few formal validation studies exist, with the notable exception of a study by Busching et al. (2015), which investigated the reliability and validity of measuring violent game exposure through favorite titles or genres. Although Busching et al. found that the measures they investigated were reliable and valid indicators of game violence exposure, no such knowledge is available for television violence exposure measures. Systematic evaluation of self-report measures of media violence exposure, both televised and game-based, is necessary in order to assess whether this field can continue with or should reconsider using current self-report measures. In order to enable critical evaluations of past studies, as well as optimal measurement in future research, this study examines and compares the reliability and validity of two of the most common types of self-report measures of exposure to television and game violence: direct estimates and favorites.[1]

Direct Estimates of Television and Game Violence Exposure

Direct estimates measure the frequency and/or duration of participants' average exposure to media violence. For example, Fikkers, Piotrowski, Weeda, Vossen, and Valkenburg (2013) used direct estimates of television and game violence exposure, asking adolescents to report (a) on how many days per week they watch violent television programs [play violent games] and (b) how much time (in hours and minutes) they spend watching violent television programs [playing violent games] on those days. The key advantage of direct estimates is that they are a quick way to obtain an estimate of a person's media exposure. For this reason, many large-scale surveys have incorporated direct estimates (Vandewater & Lee, 2009). However, arriving at a correct estimate requires several cognitive steps to be taken by the participant. Participants have to (a) understand the question and interpret it in the way intended by the researcher, (b) retrieve all relevant information from memory, then (c) integrate all this information into a single answer, and (d) report this answer correctly and truthfully (Robinson & Godbey, 1997; Schwarz & Oyserman, 2001). During this process, the accuracy of the resulting answer may be affected by two factors (Valkenburg & Peter, 2013b). First, cognitive factors (e.g., problems with recall) can affect the precision with which participants are able to recall and report time spent with violent content. This may be especially difficult for younger participants who may not have developed all necessary cognitive skills yet (Ogan, Karakus, & Kursun, 2013). Second, motivational factors (e.g., attraction to violent content) can affect participants' interpretation of what is "violent" and consequently their reporting of exposure to violent media content.

Although a number of studies have used direct estimates for media violence exposure (Fikkers et al., 2013; Fraser, Padilla-Walker, Coyne, Nelson, & Stockdale, 2012; Graber, Nichols, Lynne, Brooks-Gunn, & Botvin, 2006; Nikkelen et al., 2014; Slater, 2003; Slater, Henry, Swaim, & Cardador, 2004; Wallenius & Punamaki, 2008; Wallenius, Punamaki, & Rimpela, 2007), direct estimates are most often used to measure general exposure to media—irrespective of content. Most of the current knowledge about the reliability and validity of the direct estimates is therefore based on studies validating direct estimates of general exposure to television. These studies consistently show that general direct estimates have sufficient reliability and validity, and although general direct estimates tend to overestimate exposure, they have been shown to correlate moderately with a criterion or "gold standard" measure (e.g., media diaries in which participants report all titles of media they have used on a particular day; D. R. Anderson, Field, Collins, Lorch, & Nathan, 1985; Greenberg et al., 2005; Schmitz et al., 2004; Van der Voort & Vooijs, 1990).

Although direct estimates of general media exposure have demonstrated sufficient reliability and validity in previous research (e.g., D. R. Anderson et al., 1985; Van der Voort & Vooijs, 1990), it is unclear whether this is also true of direct estimates of *violent* media exposure. Compared with estimating one's time spent with games or television in general, estimating one's exposure to violent content in games or on television

requires the extra cognitive step of assessing the presence of violent content in those media. Participants not only have to recall when and how long they were playing games or watching television but also more specifically which kind of content they were consuming at those times, and only report those instances in which content was violent. This brings an additional cognitive task to the answering process that may affect the reliability and validity of the resulting answers. It is therefore necessary to evaluate specifically the reliability and validity of direct estimates of exposure to violence in games and on television.

Favorite Titles as Measure of Media Violence Exposure

A second common type of media violence exposure measurement is based on participants' favorite media titles. This approach was developed by C. A. Anderson and Dill (2000) to measure exposure to violence in games. Participants are asked to write down their three favorite games, and to indicate for each title (a) how often they play it (never, sometimes, often) and (b) how violent its content is. The favorites have frequently been used to measure exposure to violence both in games and on television (e.g., Coyne, Nelson, Graham-Kevan, Keister, & Grant, 2010; Ferguson, San Miguel, & Hartley, 2009; Gentile, Lynch, Linder, & Walsh, 2004). Measuring media violence via favorite games or television programs provides a simple solution to the main weakness of the direct estimates. Where direct estimates require that participants recall all instances of media violence exposure, which is cognitively demanding, the favorites only focus on the frequency and violent content of a limited number of favorite media products. Because participants only have to recall how often they play a particular game or see a particular television program, it is more likely that the frequency and content of these favorites are accurately recalled (Schwarz & Oyserman, 2001). However, the potential penalty for this improved recall is that the favorites do not capture all media violence exposure. After all, exposure to violence in non-favorite games or television programs is excluded. In addition, the favorites rely on participants' own interpretation of "violence," meaning that different participants may assign different levels of violent content to the same favorite title.

One solution to the problem associated with using "user-ratings" of violent content is to use existing media rating systems. Instead of asking participants to assess the level of violent content in their favorite media titles ("user-rated favorites"), researchers have also used "agency-rated" favorites. For example, they have used official agency-ratings such as the Entertainment Software Rating Board to determine the level of violence in games and television programs (e.g., Boxer, Huesmann, Bushman, O'Brien, & Moceri, 2009; Clemente, Espinosa, & Vidal, 2008; Coyne et al., 2010; Ferguson & Olson, 2013; Lenhart et al., 2008). While user-rated favorites are a quicker method, agency-rated favorites ensure that violent content is assessed in the same way across all participants. Although this ensures consistency across study participants, different rating systems themselves differ in what they consider "violent," and this can vary between countries (Price, Palsson, & Gentile, 2014) as well as over time (Gentile, Humphrey, & Walsh, 2005).

Busching et al. (2015) recently compared user-rated and agency-rated favorites as measures of game violence exposure using three longitudinal samples from the United States, Germany, and Singapore. Findings indicated that (a) users and agencies arrive at similar levels of violent content for the same game (inter-rater reliability), (b) there is high agreement between user-rated and agency-rated favorites (convergent validity), and (c) user-rated and agency-rated favorites both show significant positive correlations with aggressive behavior, with slightly higher correlations for the user-rated favorites (construct validity). This study revealed that agency-ratings do not necessarily result in better reliability and validity of the favorites than user-ratings of violent game content, suggesting that both can be used in research.

While Busching et al. (2015) provide useful comparative information on using different types of ratings for video game violence, their study does not provide any indication as to whether user-ratings and agency-ratings may also be reliable and valid for television violence exposure. Apart from Busching et al., knowledge about the reliability and validity of the favorites remains limited to what research articles report in their method sections (e.g., a test-retest reliability coefficient of $r = .33$, reported by Ferguson, 2011). As such, two questions related to the quality of the user-rated and agency-rated favorites remain unanswered. First, it is unclear to what extent the favorites provide a good indication of media violence exposure. Because of the favorites' specific focus on a limited number of favorite media titles, they may not be representative of total game or television violence exposure. Second, it is unclear whether user-rated and agency-rated favorites are also reliable and valid measures of exposure to violence on television. Therefore, evaluating the reliability and validity of user-rated and agency-rated favorites as an indicator of exposure to violence on television and games is a reasonable next step.

The Current Study

Although some knowledge exists regarding the reliability and validity of the direct estimates, user-rated favorites, and agency-rated favorites, this knowledge is not yet very systematic. As a result, it is difficult to know the extent to which these instruments are good measures of exposure to violence in different media. In addition, because each of these measures has its own unique weaknesses, it is difficult to establish literature-based a priori expectations about which measure(s) may be superior. Therefore, in this study, we evaluated all three measures against three measurement criteria. Specifically, this study assessed the test-retest reliability, criterion validity, and construct validity of direct estimates, user-rated favorites, and agency-rated favorites as measures of exposure to violence in games and television programs. Given that many studies investigate media violence in child and adolescent samples, our evaluation of these measures was conducted using data from a sample of early adolescents aged 10 to 14.

Test-Retest Reliability

Test-retest reliability is the extent to which a measure provides similar results when administered again after a period of time (Carmines & Zeller, 1979). It is not expected that re-administrations produce the exact same result, but rather that their results are

consistent (i.e., a participant who has a high media violence exposure score at Time 1 is also expected to have a high media violence score at Time 2). Although it has been argued that assessing test-retest reliability for media violence exposure measures is less appropriate because exposure may vary over time (Busching et al., 2015), it is relevant to describe the extent of this variability. Given that (a) more variability will result in lower test-retest reliability and (b) reliability is a prerequisite for validity, information about a measure's (lack of) variability is important for the interpretation of validation results. Therefore, the first step in this validation study was to assess the test-retest reliability of direct estimates, user-rated favorites, and agency-rated favorites by investigating the test-retest correlation between two administrations of these measures over a 4-month period.

Criterion Validity

Criterion validity is established by comparing the scores obtained with one measure with scores obtained with an already validated measure (the "criterion"; Schutt, 2012). The higher the correlation with a criterion, the more confidence we can have that the to-be-validated instrument measures what it is supposed to measure. Generally, media diaries tend to be seen as a "gold standard" in media research (Jordan et al., 2007; Juster & Stafford, 1991; Schmitz et al., 2004). Media diaries are one of the most elaborate measures of media exposure, measuring all titles of media products (e.g., games, television shows) used on a particular day, which are then coded for specific media content (e.g., violent content, see Bickham & Rich, 2006). The strength of media diaries lies in two elements that are known to improve recall. First, media diaries capitalize on the autobiographical structure of our memory. By encouraging participants to think about their day, a rich network of associations is activated, which increases the likelihood that individual episodes of media use are retrieved (Schwarz & Oyserman, 2001). Second, because media diaries tend to be filled out on the day itself or the day after, this short and recent reference period improves the likelihood of accurate recall (Schwarz & Oyserman, 2001). Indeed, a classic study by D. R. Anderson et al. (1985), in which parents kept media diaries of their child's general television exposure, showed that media diaries have both high reliability (1-month test-retest reliability: $r = .72$) and high criterion validity (when correlated with video observation: $r = .84$). Therefore, media diaries are frequently used as a criterion measure in validation studies (e.g., Greenberg et al., 2005; Özmert, Toyran, & Yurdakök, 2002; Schmitz et al., 2004; Van der Voort & Vooijs, 1990).

In media violence research, media diaries have not often been used as an exposure measure. This is mainly due to practical considerations: Using media diaries is not only expensive but it also places a high burden on both participants and researchers. Research projects often lack the time or resources for measuring media violence exposure through media diaries, and instead opt for shorter self-report measures. Indeed, when examining the empirical literature, media diaries have most often been used as a measure of general television exposure and not for other types of media such as games or for violent content in television or games (but see Bickham & Rich, 2006). However, because media diaries are expected to be more reliable than short self-report measures

due to their recall-improving characteristics, it is reasonable to expect that media diaries can also serve as a criterion measure for exposure to violent content on television and in games. Therefore, the second step in this validation study was to assess the criterion validity of the direct estimates, user-rated favorites, and agency-rated favorites by investigating their correlations with coded media diaries.

Construct Validity

Construct validity refers to "the extent to which a particular measure relates to other measures consistent with theoretically derived hypotheses" (Carmines & Zeller, 1979, p. 23). The third step in our validation process was to test theoretically and empirically based hypotheses in order to establish the construct validity of the direct estimates, user-rated favorites, and agency-rated favorites. First, based on theoretical predictions (e.g., C. A. Anderson & Bushman, 2002) as well as meta-analytic evidence (C. A. Anderson et al., 2010; Bushman & Huesmann, 2006; Ferguson & Kilburn, 2009; Paik & Comstock, 1994; Sherry, 2001), we expect valid measures of television and game violence exposure to be positively associated with aggressive behavior:

Hypothesis 1 (H1): When measuring game violence exposure, (a) direct estimates, (b) user-rated favorites, and (c) agency-rated favorites will be positively associated with aggressive behavior.
Hypothesis 2 (H2): When measuring television violence exposure, (a) direct estimates, (b) user-rated favorites, and (c) agency-rated favorites will be positively associated with aggressive behavior.

Second, several studies have found that boys are more likely than girls to select violent media exposure (e.g., Boxer et al., 2009; Coyne & Archer, 2005; Olson et al., 2007). Valid measures of game and television violence exposure should reflect this difference. Therefore, we hypothesized that direct estimates, user-rated favorites, and agency-rated favorites would indicate higher game and television violence exposure for boys than for girls:

Hypothesis 3 (H3): When measuring game violence exposure, (a) direct estimates, (b) user-rated favorites, and (c) agency-rated favorites will show more exposure for boys than for girls.
Hypothesis 4 (H4): When measuring television violence exposure, (a) direct estimates, (b) user-rated favorites, and (c) agency-rated favorites will show more exposure for boys than for girls.

Method

Participants and Procedure

After receiving approval from the sponsoring institution's Institutional Review Board, a large, private survey research institute in the Netherlands (TNS NIPO/Veldkamp)

collected the data. Adolescents were recruited through TNS NIPO's existing online panel of approximately 60,000 households that is representative of the Netherlands. Data collection consisted of two waves. The first wave consisted of an online survey in the last week of January 2012 and online media diaries in February 2012. Wave 2 consisted of an online survey in the last week of May 2012.

A total of 499 Dutch early adolescents between the ages of 10 and 14 years participated in both data waves. To be included in this validation study, participants had to have (a) scores on the direct estimates in both data waves, (b) at least one favorite television program and game that could be coded in both data waves, and (c) completed media diaries for 1 weekday and 1 weekend day, in order to establish criterion validity. A media diary day was defined as complete when all content on that day could be coded. The final sample consisted of 238 participants who met these three requirements (53.8% sibling pairs; 47.5% girls; age at Time 1: M = 11.9 years, SD = 1.5 years). All 238 participants were included in all analyses.

Measures of Television and Game Violence Exposure

Direct estimates. The direct estimates measured exposure to violent content on television and in electronic games with two items each (four items total): (a) How often do you watch television programs [play games] that contain violence? and (b) On the days that you watch television programs [play games] that contain violence, how much time do you spend on this per day? Participants were given the following definition of violence: "All violence (e.g., fighting and shooting) that living beings (e.g., humans and monsters) do to each other." Games referred to all types of games (video games, but also casual games played on mobile phones or websites). Response categories for the first item ranged from 0 (never) to 7 (7 days per week). The second item was an open-ended question, answered by filling in hours and minutes. The two items for each medium were multiplied to calculate the number of hours per week of violent television and violent game exposure. Adolescents in our sample reported an average of 1.99 hours per week (SD = 2.64) of television violence exposure (girls: M = 1.27, SD = 1.53; boys: M = 2.63, SD = 3.22). For game violence exposure, participants reported an average of 3.48 hours per week (SD = 6.64; girls: M = 0.60, SD = 1.62; boys: M = 6.08, SD = 8.22).

Favorites. Participants were asked to write down the titles of their three favorite television programs and games. For each title, they indicated (a) how often they watch this program [play this game] and (b) how much violence the program [game] contains. Participants were given the same definition of violence as used for the direct estimates. Response categories for the frequency item were (1) never or almost never, (2) less than once a week, (3) once or twice a week, (4) 3 or 4 times a week, and (5) almost every day or daily. Response categories for the violent content item were (1) no violence, (2) some violence, (3) much violence, and (4) very much violence.

For the calculation of the user-rated favorites, the frequency and violent content items for each title were multiplied with each other; these scores were then averaged

to provide an indication of the degree to which participants consume violence on television or in games (if participants had only provided one or two codable titles, we used the score of the one title or the average of the two titles). For television violence, participants reported an average of 5.15 (SD = 2.20) on the user-rated favorites (girls: M = 4.82, SD = 1.94; boys: M = 5.46, SD = 2.39; observed range = 1-15). For game violence, participants reported an average of 5.85 (SD = 3.77) on the user-rated favorites (girls: M = 3.63, SD = 1.34; boys: M = 7.85, SD = 4.12; observed range = 1-20).

For the agency-rated favorites, we multiplied the frequency item with agency-ratings of violent content (see "Content Coding" section), which were then averaged over all titles provided by that participant. For television violence, participants reported an average of 4.87 (SD = 2.44) on the agency-rated favorites (girls: M = 5.15, SD = 2.91; boys: M = 4.61, SD = 1.89; observed range = 1-15). For game violence, participants reported an average of 6.24 (SD = 4.13) on the agency-rated favorites (girls: M = 4.02, SD = 2.01; boys: M = 8.25, SD = 4.51; observed range = 1-20).

Criterion Validity: Media Diaries

Participants filled out online media diaries on 1 random weekday and 1 random weekend day in the month of February 2012. Media diaries were filled out in the evening, after 8:00 p.m.; participants reported all titles of television programs and games used in the 24 hours before (except between midnight and 7:00 a.m.). Following common diary practices, all participants received paper-and-pencil versions on which they could write down titles of media products during the day. They could then refer to this document when completing the online diary in the evening. In the online media diaries, participants were first asked whether they had seen any television programs or played any games during each of five specified parts of the day (8:00 p.m.-bedtime, 7:00 a.m.-noon, noon-3:00 p.m., 3:00-6:00 p.m., and 6:00-8:00 p.m.). When participants answered yes, they were asked for the name of the television program(s) or game(s) that they watched/played during that part of the day. Participants then indicated how long they watched each program or played each game by selecting from a list the 30-minute time intervals (e.g., 6:00-6:30 p.m., 6:30-7:00 p.m., etc.) during which that program or game had been used.

Content Coding

The titles provided by the media diaries and the favorites were coded for violent content using two official agency-rating systems. For television programs, the Dutch rating system "Kijkwijzer" ("Viewing guide") was used (Valkenburg, Beentjes, Nikken, & Tan, 2002). This rating system advises parents about the potential adverse effects of television programs and movies on children. It assigns both an age rating (suitable for all ages/6+/12+/16+) and a content rating (contains violence/scary content/sexual content/discrimination/drugs or alcohol abuse/coarse language). For games, the international Pan European Game Information (PEGI; 2012) rating system was used. This system also assigns both an age rating (3+/7+/12+/16+/18+) and a content rating,

which is the same as in the Kijkwijzer. In both systems, a violent content rating in combination with a higher age rating indicates more severe violent content.

Content coding took place in three steps. First, trained coders coded the television and game titles by looking up their ratings in the online Kijkwijzer and PEGI databases. Second, titles that were not in these online databases were coded by our coders following the official Kijkwijzer/PEGI guidelines. Third, some entries in the media diary consisted of television channels instead of actual titles. For these entries, coders looked in the television guide to see which programs were broadcast at that time on that channel. These programs were then also coded and included in the data set. For each of these steps, reliability was evaluated by double-coding at least 25% of the unique titles in the data set. The media diary resulted in a total of 6,760 television entries, of which 92% could be coded, and 3,737 game entries, of which 86% could be coded. At Time 1, the favorites resulted in a total of 1,995 television entries, of which 87% could be coded, and 1,858 game entries, of which 84% could be coded. Coding reliability was high (Kappas ranged from .83 to .94).

We calculated television and game violence exposure based on the media diary in three steps. First, we selected all television programs and games with a violent content rating. Then, we summed the number of 30-minute intervals associated with these violent television programs or games, which resulted in a number of hours of violent television/game exposure per diary day. Lastly, in order to arrive at an estimate of exposure in hours per week, we multiplied the violence exposure on the random weekday by 5 and the violence exposure on the random weekend day by 2, and then summed these two outcomes. This resulted in an average television violence exposure of 2.01 hours per week (SD = 2.77; girls: M = 2.14, SD = 3.18; boys: M = 1.90, SD = 2.34). For game violence exposure, an average of 4.57 hours per week (SD = 8.42) was reported by the participants (girls: M = 1.49, SD = 4.47; boys: M = 7.36, SD = 10.05).

For the agency-rated favorites, the categories of degree of violent content were kept similar to that of the user-rated favorites. Recall that the user-ratings consisted of the following categories: no/some/much/very much violence. The agency-ratings were mapped onto those categories in the following way: A violent content rating in combination with the lowest age rating (i.e., all ages or 3+) was considered “no violence”; violent content in combination with an age rating of 6+/7+ was considered “some violence”; violent content in combination with 12+ was considered “much violence”; and violent content in combination with 16+ and 18+ was considered “very much violence.” These agency-ratings of violent content were multiplied with the frequency rating of the favorites, and then averaged within television and game titles.

Construct Validity: Aggressive Behavior

Adolescents’ direct aggression was measured with six items adapted from the Direct and Indirect Aggression Scale (Björkqvist, Lagerspetz, & Kaukiainen, 1992). This measure has been used in adolescent samples with good reliability and validity (e.g., Hale, VanderValk, Akse, & Meeus, 2008). Adolescents were asked how often in the past 6 months they had done the following things to another adolescent: (a) call names,

(b) push in a rough way, (c) kick or hit, (d) threaten to beat up, (e) fought with, and (f) tripped on purpose. Response categories were (1) never, (2) 1 time in the past 6 months, (3) 2 to 3 times in the past 6 months, (4) about 1 time per month, (5) about 1 time per week, and (6) about every day. Scores were averaged to create scales ($\alpha = .84$), with higher scores indicating greater aggressive behavior. The mean score on aggressive behavior in our sample was 1.52 ($SD = 0.72$).

Analytic Approach

The main analyses consisted of three steps. First, test-retest reliability of the direct estimates, user-rated favorites, and agency-rated favorites was assessed by investigating the correlation between scores on the same measure at Time 1 and Time 2. Second, to assess criterion validity, correlations of the direct estimates, user-rated favorites, and agency-rated favorites with the media diary were investigated. Third, for construct validity, we investigated bivariate correlations between the direct estimates, user-rated favorites, and agency-rated favorites and aggressive behavior and gender at Time 1.

Pearson's correlation coefficients were not appropriate in this study for two reasons. First, all variables were non-normally distributed. As can be expected, many early adolescents had (very) low scores for media violence exposure, leading to a skewed distribution with a long tail. Second, approximately half of our sample (53.8%) consisted of sibling pairs, meaning that the assumption of independent observations is violated, which can result in overestimated coefficients. To account for these two characteristics of the data, we calculated Kendall's tau-a correlations in combination with the clustering option in Stata 12. In addition, we present asymmetric 95% confidence intervals for each Kendall's tau coefficient. Kendall's tau is a non-parametric correlation between two ranked variables (Newson, 2002). A positive value of tau represents the probability of agreement between two variables over the probability of disagreement between the same variables. Stata 12 enables a conversion of Kendall's tau into an approximation of Pearson's r using Greiner's relation (Newson, 2002). Although Kendall's tau is the more appropriate coefficient given the non-normal and clustered nature of our data, we also include Pearson's r to aid interpretation of the results.

Interpreting Reliability and Validity Coefficients

In a validation study, the size of a correlation coefficient is an indicator of the extent to which reliability or validity is achieved. Yet, few guidelines exist for the interpretation of the correlation coefficients. To aid interpretation of the Pearson's r coefficients in our study, we relied on the extant literature to establish reasonable guidelines for evaluating measurement reliability and validity. First, we consider a Pearson's r coefficient in the range of .40 or higher to be evidence for test-retest reliability (cf. Dahlberg, Toal, Swahn, & Behrens, 2005, p. 3). Second, we consider a Pearson's r coefficient in the range of .50 or higher to be evidence of criterion validity (cf. Jordan et al., 2007; Van der Voort & Vooijs, 1990). Third, to achieve construct validity as predicted by H1 to H4, the exposure measures should correlate with aggressive behavior and gender in

a way that is consistent with the theoretical and empirical literature. Meta-analyses show that, in survey research, the average bivariate correlations between aggressive behavior and game violence range between .08 (Ferguson & Kilburn, 2009) and .20 (C. A. Anderson et al., 2010; Greitemeyer & Mügge, 2014), and falls in the range of .19 for television violence (Paik & Comstock, 1994; Savage & Yancey, 2008).[2] Based on these results, we consider construct validity to be achieved when a television/game violence measure shows a positive bivariate correlation with aggressive behavior in that range (H1 and H2).[3] Furthermore, related to H3 and H4, we consider construct validity to be achieved when a measure correlates with gender at .40 or higher, with higher violent media exposure for boys than girls (cf. Boxer et al., 2009; Coyne & Archer, 2005; Olson et al., 2007).

Results

Game Violence Exposure

The upper panel of Table 1 presents the correlation coefficients and 95% confidence intervals for the measures of game violence exposure. The direct estimates of game violence exposure achieved test-retest reliability (tau = .52, r = .73, $p < .001$). For criterion validity, the correlation of the direct estimates with the media diaries came close to, but did not exceed our guideline of .50 (tau = .28, r = .43, $p < .001$). The two hypotheses related to the direct estimates' construct validity (H1a and H3a) were confirmed: Direct estimates of game violence exposure were positively associated with aggressive behavior and showed more exposure for boys than for girls (aggression: tau = .17, r = .27, $p < .001$; gender: tau = .34, r = .52, $p < .001$).

The user-rated game favorites also achieved test-retest reliability (tau = .47, r = .67, $p < .001$). For criterion validity, the correlation of the user-rated favorites with the media diary also came close to, but did not exceed, our guideline of .50 (tau = .29, r = .44, $p < .001$). Hypotheses H1b and H3b, related to the construct validity of the user-rated favorites, were confirmed. The associations of the user-rated favorites with aggression and gender were in the expected direction (aggression: tau = .18, r = .29, $p < .001$; gender: tau = .35, r = .52, $p < .001$).

The agency-rated game favorites met all guidelines for reliability and validity. This measure of game violence exposure achieved test-retest reliability (tau = .52, r = .73, $p < .001$) and criterion validity (tau = .37, r = .55, $p < .001$). Construct validity (H1c and H3c) was confirmed as the associations between agency-rated favorites with aggression and gender were significant and in the expected direction (aggression: tau = .15, r = .24, p = .002; gender: tau = .31, r = .47, $p < .001$).

Television Violence Exposure

The lower panel of Table 1 presents the correlation coefficients and 95% confidence intervals for the measures of television violence exposure. The direct estimates of television violence exposure achieved test-retest reliability (tau = .34, r = .51,

Table 1. Test-Retest Reliability, Criterion Validity, and Construct Validity of Game and Television Violence Exposure Measures.

							Construct validity[b]					
	Test-retest reliability			Criterion validity[a]			Aggression			Gender[c]		
Type of measure	Tau	95% CI	*r*	Tau	95% CI	*r*	Tau	95% CI	*r*	Tau	95% CI	*r*
Game violence exposure												
Direct estimates	.52*	[.45, .58]	.73*	.28*	[.19, .37]	.43*	.17*	[.09, .26]	.27*	.34*	[.29, .39]	.52*
User-rated favorites	.47*	[.40, .54]	.67*	.29*	[.21, .37]	.44*	.18*	[.10, .27]	.29*	.35*	[.30, .40]	.52*
Agency-rated favorites	.52*	[.45, .58]	.73*	.37*	[.29, .44]	.55*	.15*	[.06, .24]	.24*	.31*	[.25, .36]	.47*
Television violence exposure												
Direct estimates	.34*	[.26, .42]	.51*	.11*	[.03, .19]	.18*	.19*	[.10, .27]	.29*	.14*	[.06, .20]	21*
User-rated favorites	.33*	[.25, .41]	.50*	.08*	[.00, .16]	.13*	.02	[−.06, .11]	.04	.08*	[.00, .15]	.12*
Agency-rated favorites	.31*	[.22, .39]	.47*	.12*	[.05, .20]	.19*	−.07	[−.15, .02]	−.11	−.00	[−.09, .08]	−.00

Note. Pearson's *r* was derived from Kendall's tau using Greiner's relation in Stata (Newson, 2002). CI = confidence interval.
[a]Criterion validity involves the relationship of the self-report measures with coded media diaries.
[b]Construct validity involves the relationship of the self-report measures with aggressive behavior and gender.
[c]Girls = 0; boys = 1.
*$p < .05$.

$p < .001$). Criterion validity could not be established for the direct estimates of television violence exposure, because its correlation with the media diaries was well below our guideline of a Pearson's r of .50 (tau = .11, $r = .18$, $p = .006$). The two hypotheses related to the direct estimates' construct validity (H2a and H4a) were confirmed: Direct estimates of television violence exposure were positively associated with aggressive behavior (tau = .19, $r = .29$, $p < .001$) and showed more exposure for boys than for girls (tau = .14, $r = .21$, $p < .001$). For gender, the correlation did not meet our guideline of a Pearson's r of .40, but it was significant and in the expected direction.

For the user-rated television favorites, only test-retest reliability could be established (tau = .33, $r = .50$, $p < .001$). Neither criterion validity nor construct validity was achieved for this measure. For criterion validity, the correlation between the user-rated favorites and the diary was well below our guideline of a Pearson's r of .50 (tau = .08, $r = .13$, $p = .038$). For construct validity, no significant correlation with aggression was found (tau = .02, $r = .04$, $p = .575$), thus rejecting H2b. H4b, which hypothesized more exposure for boys than for girls, was confirmed (tau = .08, $r = .12$, $p = .040$). However, because the correlation was well below our guideline of .40, we do not consider construct validity with gender achieved for the user-rated television favorites.

For the agency-rated television favorites, too, only test-retest reliability could be established (tau = .31, $r = .47$, $p < .001$). Neither criterion validity nor construct validity was achieved for this measure. For criterion validity, the agency-rated television favorites did not meet our guideline of .50 (tau = .12, $r = .19$, $p = .002$). For construct validity, no significant correlations with aggression and gender were found (aggression: tau = −.07, $r = -.11$, $p = .120$; gender: tau = −.00, $r = -.00$, $p = .941$), thus rejecting H2c and H4c.

Post Hoc Assessment of the Media Diary

After completing the main analyses, we opted to conduct a post hoc evaluation of the media diaries for two reasons. First, no studies have used media diaries to measure exposure to violence in games and on television, and thus, our assumption of the media diary as an appropriate criterion for exposure to violent content may be incorrect. And second, in particular, the correlations between the three self-report measures of television violence exposure and the media diaries were remarkably low. Test-retest reliability for the media diaries was assessed by investigating the correlation coefficient between the two media diary days used in this study. Construct validity was assessed in the same way as for the direct estimates, user-rated favorites, and agency-rated favorites, that is, via correlations with aggressive behavior and gender. Table 2 presents the correlation coefficients and 95% confidence intervals for the media diary measures of game and television violence exposure.

For game violence exposure measured with the media diary, test-retest reliability was achieved (tau = .30, $r = .45$, $p < .001$) as was construct validity (aggression: tau = .10, $r = .15$, $p < .001$; gender: tau = .23, $r = .35$, $p < .001$). However, for television violence exposure measured with the media diary, test-retest reliability could not be established (tau = −.01, $r = -.01$, $p = .762$). Construct validity was also not achieved:

Table 2. Test-Retest Reliability and Construct Validity of the Media Diaries.

				Construct validity					
	Test-retest reliability			Aggression			Gender[a]		
Media diary	Tau	95% CI	*r*	Tau	95% CI	*r*	Tau	95% CI	*r*
Game violence exposure	.30*	[.22, .37]	.45*	.10*	[.01, .18]	.15*	.23*	[.17, .29]	.35*
Television violence exposure	−.01	[−.07, .05]	−.01	−.10*	[−.19, −.02]	−.16*	.01	[−.07, .08]	.01

Note. Pearson's *r* was derived from Kendall's tau using Greiner's relation in Stata (Newson, 2002). CI = confidence interval.
[a]Girls = 0; boys = 1.
*$p < .05$.

Television violence exposure measured with the media diary correlated negatively with aggressive behavior (tau = −.10, $r = -.16$, $p = .015$) and did not correlate with gender (tau = .01, $r = .01$, $p = .863$).

As the post hoc analysis revealed that the media diary was not a valid criterion measure for television violence exposure, we also assessed the intercorrelations among the three self-report measures as an additional way of assessing their validity. High agreement between the self-report measures suggests that they are measuring the same concept, which is indicative of convergent validity (Busching et al., 2015). For game violence exposure, the direct estimates strongly correlated with the user-rated favorites (tau = .55, $r = .76$, $p < .001$) and the agency-rated favorites (tau = .48, $r = .69$, $p < .001$). The user-rated and agency-rated game favorites were also strongly related to one another (tau = .63, $r = .84$, $p < .001$). For television violence exposure, the direct estimates correlated moderately with the user-rated favorites (tau = .37, $r = .54$, $p < .001$) and low with agency-rated favorites (tau = .21, $r = .33$, $p < .001$). The user-rated and agency-rated television favorites showed high agreement (tau = .50, $r = .70$, $p < .001$).

Discussion

The aim of this study was to evaluate whether three commonly used self-report measures of television and game violence exposure are reliable and valid measurement tools. To this end, we assessed the test-retest reliability, criterion validity, and construct validity of direct estimates, user-rated favorites, and agency-rated favorites as measures of exposure to violence in games and on television in an early adolescent sample.

Game Violence Exposure

Results indicated that for game violence exposure, all three measures achieved test-retest reliability and construct validity; that is, they showed stability over a 4-month

period and correlated as expected with aggressive behavior and gender (thereby confirming H1 and H3). Criterion validity (the degree to which the direct estimates and the favorites corresponded with coded media diaries) was highest for the agency-rated favorites. Correlations between the media diaries and direct estimates and user-rated favorites (.43 and .44, respectively) did not exceed our self-established guideline of .50. However, because we considered correlations in range of .50 as sufficient, we also consider these measures as having achieved criterion validity. Furthermore, our post hoc analysis indicated high agreement between the three self-report measures, which also supports these measures' convergent validity. Therefore, we consider each of the direct estimates, user-rated favorites, and agency-rated favorites to be reliable and valid measures of game violence exposure.

Television Violence Exposure

For television violence exposure, test-retest reliability was achieved for the direct estimates, user-rated favorites, and agency-rated favorites. Construct validity could only be established for the direct estimates (confirming H2a, H4a); neither the user-rated favorites nor the agency-rated favorites achieved construct validity as a measure of television violence exposure (rejecting H2b and H2c, H4b and H4c). Criterion validity was not achieved with any of the three measures.

Given these relatively surprising findings for criterion validity, we conducted post hoc analyses of the media diary to ascertain its own test-retest reliability and construct validity. These analyses revealed that the media diary could not be considered a reliable and valid measure of television violence exposure in this study. Consequently, comparing the direct estimates, user-rated favorites, and agency-rated favorites to the media diary is not a valid way of assessing criterion validity for television violence exposure measures in this study. Therefore, only the results for test-retest reliability and construct validity can be used to evaluate the three self-report measures. The findings of these analyses show that only the direct estimates are a reliable and valid measure of television violence exposure. Based on the lack of construct validity for both the user-rated and agency-rated favorites, combined with the lower correlations of the user-rated and agency-rated favorites with the direct estimates, we cannot consider either favorites measure to be a valid measure of television violence exposure.

Implications for Using Current Media Violence Measures in Future Research

Our study findings provide three relevant implications for researchers who are thinking about using direct estimates, favorites, or media diaries in their own work. First, our study indicates that media diaries may not be a "gold standard" measure when it comes to measuring exposure to specific content on television. The highly specific nature of media diaries—that is, the practice of filling out media diaries on one or two specific days—may render it particularly susceptible to the day-to-day variation that is inherent to media use (cf. Jordan et al., 2007). In other words, using a media diary on

one or two specific days may not be representative of the television content that a person is exposed to on average. Instead, it seems more appropriate to use media diaries as a measure of time spent with television in general, irrespective of content.

Second, our findings revealed that the violence exposure measures used in this study (including the media diary) consistently worked better for games than for television. Although several differences in the nature of television and game use may be put forward to explain this pattern, our data suggest that the most likely reason is a higher variability in television viewing versus game playing. In our study, for example, adolescents reported almost twice as many television titles compared with game titles in their media diaries. Moreover, television violence exposure was not very common in our sample of typically developing early adolescents (ca. 2 hours per week on average). This combination of high variability of television viewing and low frequency of violent content means that using media diaries on 2 days decreases the chance that certain programs (such as violent programs) are captured. This may explain why the media diaries and favorites worked less well in our study as measures of television violence exposure than as measures of game violence exposure. Future validation research may investigate whether measurement reliability can be enhanced by increasing the number of diary days or favorite titles reported, under the assumption that a larger number is more representative of total television violence exposure. Furthermore, our findings point to the importance of conducting validation studies when adapting measurement tools. The favorites measure, for example, was originally developed as a measure of game violence exposure (C. A. Anderson & Dill, 2000). The results presented here and elsewhere (Busching et al., 2015) support the reliability and validity of this measure for violent game exposure. However, our results do not support the use of favorites for measuring television violence exposure. This suggests that researchers should be cautious when applying a measurement approach designed for one medium to other types of media. Differences between media may affect measurement in unanticipated ways, making the assumption that one measure will also work for other media a hazardous one.

Third, our results support the conclusion by Busching et al. (2015) that both user-ratings and agency-ratings are reliable and valid ways of assessing the level of violence in games. From a utility perspective, this indicates that researchers need not spend the time and resources on having titles content-analyzed, at least not for general violent content in games. Interesting next steps would be to investigate whether user-ratings of more specific types of media violence, such as indirect aggression, are also reliable and valid indicators. The ability to use viewer interpretations, especially for concepts that are difficult to content-code, would provide a range of new opportunities for media violence research. For violent content in television programs, our findings indicate that neither user-rated nor agency-rated favorites were valid. However, it is important to note that although these measures were not found valid, our study cannot conclude whether this is a consequence of the *ratings* being invalid. It is possible that the television favorites were invalid because the low number of favorites used was not representative of exposure to violent television content, and not because users or agencies incorrectly assessed the violent content in those favorite television shows. Before

fully rejecting user-ratings and agency-ratings of television content, future research should compare different types of raters for violent television content to more fully assess the suitability of user-ratings or agency-ratings in television research.

Moving Forward: Reconsidering Media Violence Exposure Measurement

Although our study shows that direct estimates (for television and game violence exposure) and user-rated and agency-rated favorites (for game violence exposure) are reliable and valid exposure measures, the modest reliability and validity coefficients obtained in this study indicate that there is room for improvement. It seems that the field of media violence research has reached an important crossroad. On the one hand, our study suggests that we may continue using direct estimates and favorites and accept that, as "reasonably valid" ways of capturing media violence exposure, they are good enough. Or, we can make an effort to advance the field through collective "disciplinary self-reflection" about how we measure media violence exposure (Valkenburg & Peter, 2013b). In our view, this self-reflection is critical and should consider several aspects of the measurement process, such as the conceptualization of media violence exposure, the evaluation of media exposure measures, and anticipating challenges of self-report measures.

Conceptualizing media violence exposure. An important first step toward improving media violence exposure measurement is to think more carefully about what we consider "media violence exposure." Remarkably, this question has received little attention. Although definitions of "violent media" have been put forward (e.g., media "that depict intentional attempts by individuals to inflict harm on others," C. A. Anderson & Bushman, 2001, p. 354), these definitions lack precision about the type of violence we are interested in and ignore the issue of what *exposure* to such content is. A clearer conceptualization of media violence exposure will not only help us understand whether our current measures are capturing what we want to measure (Allen, 1981; Clarke & Kline, 1974) but will also encourage researchers to be more precise in their theoretical predictions about media violence (Jordan et al., 2007; Slater, 2004).

There are two ways in which we can improve our thinking about media violence exposure. First, we should consider the fundamental question of what *is* exposure to media violence. Looking at the current stock of media violence exposure measures, we see that exposure is operationalized as frequency, or "how often" people are exposed to violent content, in most exposure measures (including the favorites). Yet, frequency measures are unlikely to fully capture a media exposure experience. Consider the difference between two adolescents, both of whom play violent games every day of the week, but while one plays the game for only 5 minutes at a time, the other plays for 2 hours each day. Although the frequency of exposure is the same, these media experiences are likely to be different due to their differences in *duration*. On the one hand, differences in duration may reflect different motivations for using such content (e.g., ritualized media use—playing a game to pass the time waiting for a bus—versus instrumental use—satisfying more intrinsic needs and motivations; cf. Przybylski,

Rigby, & Ryan, 2010; Rubin, 2009). On the other hand, this duration difference likely influences how the user cognitively, affectively, and physiologically responds to the content (e.g., Krcmar & Lachlan, 2009), which is hypothesized to be a mediating route to media effects (e.g., General Aggression Model, C. A. Anderson & Bushman, 2002; Differential Susceptibility to Media-Effects Model, Valkenburg & Peter, 2013a). Yet, despite the added value of assessing both frequency and duration, few self-report measures take into account both aspects of exposure (with the exception of the direct estimates). As a field, it is important to think about what it is that we want to measure in terms of "exposure" to violent content and whether measures that favor frequency of media violence exposure, duration, or both are best able to capture this.

That said, we also should take the concept of "exposure" one step further and consider whether measuring "time spent" with media violence is sufficient. Such measures only focus on capturing "encounters" with media violence but not whether media violence was actually attended to (cf. Potter, 2008; Slater, 2004). The implicit assumption that time spent with media violence (frequency, duration, or both) equates with actual attention and cognitive effort to such content may be problematic in an environment where media multitasking is increasingly common (Rideout, Foehr, & Roberts, 2010). Researchers have argued that the *kind* of exposure, rather than the degree of it, may be what matters (e.g., automatic, attentive, or transported exposure; Potter, 2008; Valkenburg & Peter, 2013a). Relevant future research should reflect on the concept of media violence exposure as well as try to identify the best ways of capturing such exposure.

In addition to reconsidering how we define exposure, we should also be more specific about the kind of violent content in media that we are interested in. Currently, our self-report measures are characterized by a simplified view of violence by treating such content as present or absent (Ferguson, Garza, Jerabeck, Ramos, & Galindo, 2013). Yet, just like "exposure," media violence can differ in kind as well as degree (Tamborini, Weber, Bowman, Eden, & Skalski, 2013). Within media-effects research, theoretical models such as Social Cognitive Theory (Bandura, 2001) propose that some types of violent content (e.g., rewarded, justified, conducted by attractive perpetrators) may be more influential than other types of violence (e.g., punished, unjustified). In addition, experimental studies have shown that different types of violence can result in different responses (e.g., Bartsch & Mares, 2014). Similarly, the uses-and-gratifications paradigm would suggest that media users may have preferences for different kinds of violence, perhaps as a result of developmental level (e.g., cartoon or fantasy violence vs. more realistic violence; Valkenburg & Cantor, 2000) or disposition (e.g., Greene & Krcmar, 2005). Yet, current measures attempt to capture all violent content, instead of focusing on the type(s) of violence that are theoretically relevant. Thus, a second important way to improve media violence exposure measurement is to develop measures that are more sensitive to different types of violence and provide a better match with theoretical expectations about media violence.

Evaluating media exposure measures. Apart from more systematic attention to the conceptualization and operationalization of media violence exposure, if we hope to

improve our measurement, we also must put forth evaluation criteria for when a measure is considered reliable and valid. The current lack of such guidelines was a challenge for our study as well as for other researchers who may wish to evaluate their measure before using it. Although a multitude of handbooks exists to define and explain the different types of reliability and validity, it remains unclear as to "how high" a particular coefficient must be before reliability or validity is achieved. In our study, we provided guidelines based on a mix of theoretical and empirical work as a way of providing transparency about how we reached our conclusions. However, these guidelines are neither perfect nor indisputable. For example, we used average effect sizes found in meta-analyses as a guideline for the interpretation of construct validity of the media violence exposure measures. However, meta-analyses have their own limitations associated with publication bias and methodological differences between studies, which may result in overestimation of effect sizes (see Savage & Yancey, 2008). As such, some researchers may feel other guidelines are more appropriate when interpreting correlations between media violence measures and aggression. Moving forward, establishing more standardized guidelines for the evaluation of media exposure measures is critical. Doing so will help to provide more clarity about which measures achieve reliability and validity in which contexts (e.g., age groups, types of media). Moreover, a move toward more standardized use of the measures themselves would facilitate direct comparisons between studies investigating media violence.

In addition to deciding on the guidelines for reliability and validity coefficients, it is vital for future validation research to think about how reliability and validity can be assessed in the best possible way. For example, each of the three criteria used in this study may be improved in future validation research. First, test-retest reliability coefficients reflect random measurement error as well as change in the behavior over time (Heise, 1969; Lee et al., 2008). Because the variable nature of media use may prevent clear interpretations of test-retest reliability coefficients, relevant future validation research should also assess *stability* of the behavior in addition to the reliability of the measure. Second, our study used media diaries to assess criterion validity but found that this method was not a "gold standard" for television violence exposure. Future validation research should think about what other criterion measures may be used to evaluate short self-report measures. It is possible that there is, in fact, no "gold standard" available. In that case, researchers can assess a measure's convergent validity by comparing the to-be-validated measure with a range of other existing (and preferably already validated) measures. Third, in this study, we used aggression and gender to assess the exposure measures' construct validity. However, in a field where researchers passionately disagree about whether media violence is related to aggression at all, it would be advisable to include additional constructs that are theoretically expected to relate to media violence exposure (e.g., sensation seeking).

Anticipating challenges of self-report measures. Of course, no matter how well we conceptualize, operationalize, and evaluate media violence exposure measures, there will always be weaknesses inherent to self-report measures common in this and other fields of communication research. Given the costly and time-consuming nature of many

"gold standard" measures (e.g., media diaries, behavioral measures, video observation), the reality is that most media researchers need to rely on self-report measurement. It is therefore paramount that researchers consider the inherent weaknesses of such measures at study onset. For example, an important threat to the validity of self-report studies is that they often rely on single respondents for all variables. This could result in particular response patterns due to mischievous responding or single-responder bias. For example, when relying on user-rated favorites, asking participants to first rate the violent content of their favorite games, and then report on aggressive behavior, may set up demand characteristics that result in spurious effects. Researchers can take several steps to prevent this. First, researchers can triangulate data using multiple respondents (e.g., teens, their parents, and agency-ratings for violent content). Second, when using different respondents is not feasible, researchers can embed the relevant questions in a larger survey as a way of preventing respondents to guess the aim of the study. Third, researchers can use multiple measures of media exposure and outcomes. These need not only be the traditional self-report measures. Rather, researchers may also consider turning to newer methods, such as implicit measures (see Hefner, Rothmund, Klimmt, & Gollwitzer, 2011, for a review), as a way of capturing behaviors in a less obvious way. By anticipating the weaknesses of self-report measures, as well as looking for other ways of measuring media violence exposure, media violence researchers can further improve the quality of their future work.

Conclusion

In all, the results of this study support the reliability and validity of direct estimates as a measure of both television and game violence exposure, as well as the reliability and validity of user-rated and agency-rated favorites of game violence exposure. These measures are, therefore, appropriate for use in future media violence studies. However, it is important to recognize that the reliability and validity coefficients for each of these measures were modest, which indicates that there is room for improvement. Ultimately, we believe that the future of this field does not lie in producing more studies using the current stock of measures. Instead, in order to truly move forward, scholars should systematically reflect on several aspects of media violence exposure measurement, such as what we mean by "media violence exposure," when a measure is reliable and valid, and how to anticipate challenges associated with using self-report measures. Such collective self-reflection should result in a better understanding of media violence exposure.

Acknowledgments

The authors are thankful to Dr. Peter Lugtig for his help with the analytic process. We also thank Jesse van Amerongen, Hester Brakel, Luuc Brans, Kris Compiet, Raúl Inzaurralde, Niek Melchers, Steven Ramaker, Monique Vogelzang, Dr. Helen Vossen, and Moniza Waheed for their involvement in the media content coding process.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a grant to the third author from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement [AdG09 249488-ENTCHILD].

Notes

1. Although the term *direct estimate* could be interpreted as suggesting that favorite titles are an "indirect" measure, we use the term because it is most common in the literature. We do not suggest that the favorites are an "indirect measure."
2. Although meta-analyses provide an indication of the average effect size in a particular area of research, they are not free from limitations or subjectivity. As discussed by Savage and Yancey (2008), meta-analyses may provide overestimations of effect sizes due to problems such as publication bias, mixed quality of the studies included, problems with statistical reporting, and studies using post hoc comparisons. However, in the absence of other information on average effect sizes in this field, we use the most consistent meta-analytic evidence as an indicator for a bivariate relationship between media violence and aggression.
3. As the goal of construct validity is to assess whether one measure correlates with other theoretically relevant variables in the expected direction, we focus on bivariate correlations in this study. Studies that are interested in assessing *effects* of media violence on aggression or other outcomes would clearly include relevant control variables as a way of ruling out spurious relationships.

References

Allen, R. L. (1981). The reliability and stability of television exposure. *Communication Research*, *8*, 233-256. doi:10.1177/009365028100800205

Anderson, C. A., & Bushman, B. J. (2001). Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: A meta-analytic review of the scientific literature. *Psychological Science*, *12*, 353-359. Retrieved from http://www.jstor.org/stable/40063648

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, *53*, 27-51. doi:10.1146/annurev.psych.53.100901.135231

Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology*, *78*, 772-790. doi:10.1037//0022-3514.78.4.772

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., . . . Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin*, *136*, 151-173. doi:10.1037/a0018251

Anderson, D. R., Field, D. E., Collins, P. A., Lorch, E. P., & Nathan, J. G. (1985). Estimates of young children's time with television: A methodological comparison of parent

reports with time-lapse video home observation. *Child Development*, *56*, 1345-1357. doi:10.1111/j.1467-8624.1985.tb00202.x

Bandura, A. (2001). Social cognitive theory of mass communication. *Media Psychology*, *3*, 265-299. doi:10.1207/S1532785XMEP0303_03

Bartsch, A., & Mares, M. (2014). Making sense of violence: Perceived meaningfulness as a predictor of audience interest in violent media content. *Journal of Communication*, *64*, 956-976. doi:10.1111/jcom.12112

Bickham, D. S., & Rich, M. (2006). Is television viewing associated with social isolation? Roles of exposure time, viewing context, and violent content. *Archives of Pediatrics & Adolescent Medicine*, *160*, 387-392.

Björkqvist, K., Lagerspetz, K. M. J., & Kaukiainen, A. (1992). Do girls manipulate and boys fight? Developmental trends in regard to direct and indirect aggression. *Aggressive Behavior*, *18*, 117-127. doi:10.1002/1098-2337(1992)18:2<117::AID-AB2480180205>3.0.CO;2-3

Boxer, P., Huesmann, L. R., Bushman, B. J., O'Brien, M., & Moceri, D. (2009). The role of violent media preference in cumulative developmental risk for violence and general aggression. *Journal of Youth and Adolescence*, *38*, 417-428. doi:10.1007/s10964-008-9335-2

Busching, R., Gentile, D. A., Krahé, B., Möller, I., Khoo, A., Walsh, D. A., & Anderson, C. A. (2015). Testing the reliability and validity of different measures of violent video game use in the USA, Singapore, and Germany. *Psychology of Popular Media Culture*, *4*, 97-111. doi:10.1037/ppm0000004

Bushman, B. J., & Anderson, C. A. (2001). Media violence and the American public: Scientific facts versus media misinformation. *American Psychologist*, *56*, 477-489.

Bushman, B. J., & Huesmann, L. R. (2006). Short-term and long-term effects of violent media on aggression in children and adults. *Archives of Pediatrics & Adolescent Medicine*, *160*, 348-352. doi:10.1001/archpedi.160.4.348

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.

Clarke, P., & Kline, F. G. (1974). Media effects reconsidered: Some new strategies for communication research. *Communication Research*, *1*, 224-240. doi:10.1177/009365027400100205

Clemente, M., Espinosa, P., & Vidal, M. A. (2008). The media and violent behavior in young people: Effects of the media on antisocial aggressive behavior in a Spanish sample. *Journal of Applied Social Psychology*, *38*, 2395-2409. doi:10.1111/j.1559-1816.2008.00397.x

Coyne, S. M., & Archer, J. (2005). The relationship between indirect and physical aggression on television and in real life. *Social Development*, *14*, 324-338.

Coyne, S. M., Nelson, D. A., Graham-Kevan, N., Keister, E., & Grant, D. M. (2010). Mean on the screen: Psychopathy, relationship aggression, and aggression in the media. *Personality and Individual Differences*, *48*, 288-293. doi:10.1016/j.paid.2009.10.018

Dahlberg, L. L., Toal, S. B., Swahn, M., & Behrens, C. B. (2005). *Measuring violence-related attitudes, behaviors, and influences among youths: A compendium of assessment tools* (2nd ed.). Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.

Elson, M., & Ferguson, C. J. (2014). Twenty-five years of research on violence in digital games and aggression: Empirical evidence, perspectives, and a debate gone astray. *European Psychologist*, *19*(1), 33-46. doi:10.1027/1016–9040/a000147

Ferguson, C. J. (2009). Media violence effects: Confirmed truth or just another X-file? *Journal of Forensic Psychology Practice*, *9*, 103-126. doi:10.1080/15228930802572059

Ferguson, C. J. (2011). Video games and youth violence: A prospective analysis in adolescents. *Journal of Youth and Adolescence*, *40*, 377-391. doi:10.1007/s10964-010-9610-x

Ferguson, C. J., Garza, A., Jerabeck, J., Ramos, R., & Galindo, M. (2013). Not worth the fuss after all? Cross-sectional and prospective data on violent video game influences on aggression, visuospatial cognition and mathematics ability in a sample of youth. *Journal of Youth and Adolescence*, *42*, 109-122. doi:10.1007/s10964-012-9803-6

Ferguson, C. J., & Kilburn, J. (2009). The public health risks of media violence: A meta-analytic review. *Journal of Pediatrics*, *154*, 759-763. doi:10.1016/j.jpeds.2008.11.033

Ferguson, C. J., & Olson, C. K. (2013). Video game violence use among "vulnerable" populations: The impact of violent games on delinquency and bullying among children with clinically elevated depression or attention deficit symptoms. *Journal of Youth and Adolescence*, *43*, 127-136. doi:10.1007/s10964-013-9986-5

Ferguson, C. J., San Miguel, C., & Hartley, R. D. (2009). A multivariate analysis of youth violence and aggression: The influence of family, peers, depression, and media violence. *The Journal of Pediatrics*, *155*, 904-908.e3.

Ferguson, C. J., & Savage, J. (2012). Have recent studies addressed methodological issues raised by five decades of television violence research? A critical review. *Aggression and Violent Behavior*, *17*, 129-139. doi:10.1016/j.avb.2011.11.001

Fikkers, K. M., Piotrowski, J. T., Weeda, W. D., Vossen, H. G. M., & Valkenburg, P. M. (2013). Double dose: High family conflict enhances the effect of media violence exposure on adolescents' aggression. *Societies*, *3*, 280-292. doi:10.3390/soc3030280

Fraser, A. M., Padilla-Walker, L. M., Coyne, S. M., Nelson, L. J., & Stockdale, L. A. (2012). Associations between violent video gaming, empathic concern, and prosocial behavior toward strangers, friends, and family members. *Journal of Youth and Adolescence*, *41*, 636-649. doi:10.1007/s10964-012-9742-2

Gentile, D. A., Humphrey, J., & Walsh, D. A. (2005). Media ratings for movies, music, video games, and television: A review of the research and recommendations for improvements. *Adolescent Medicine Clinics*, *16*, 427-446. doi:10.1016/j.admecli.2005.02.006

Gentile, D. A., Lynch, P. J., Linder, J. R., & Walsh, D. A. (2004). The effects of violent video game habits on adolescent hostility, aggressive behaviors, and school performance. *Journal of Adolescence*, *27*, 5-22. doi:10.1016/j.adolescence.2003.10.002

Graber, J. A., Nichols, T., Lynne, S. D., Brooks-Gunn, J., & Botvin, G. J. (2006). A longitudinal examination of family, friend, and media influences on competent versus problem behaviors among urban minority youth. *Applied Developmental Science*, *10*, 75-85.

Greenberg, B. S., Eastin, M. S., Skalski, P., Cooper, L., Levy, M., & Lachlan, K. (2005). Comparing survey and diary measures of internet and traditional media use. *Communication Reports*, *18*, 1-8. doi:10.1080/08934210500084164

Greene, K., & Krcmar, M. (2005). Predicting exposure to and liking of media violence: A uses and gratifications approach. *Communication Studies*, *56*, 71-93. doi:10.1080/0008957042000332250

Greitemeyer, T., & Mügge, D. O. (2014). Video games do affect social outcomes: A meta-analytic review of the effects of violent and prosocial video game play. *Personality and Social Psychology Bulletin*, *40*, 578-589. doi:10.1177/0146167213520459

Hale, W. W., III, VanderValk, I., Akse, J., & Meeus, W. (2008). The interplay of early adolescents' depressive symptoms, aggression and perceived parental rejection: A four-year community study. *Journal of Youth and Adolescence*, *37*, 928-940. doi:10.1007/s10964-008-9280-0

Hefner, D., Rothmund, T., Klimmt, C., & Gollwitzer, M. (2011). Implicit measures and media effects research: Challenges and opportunities. *Communication Methods & Measures*, *5*, 181-202. doi:10.1080/19312458.2011.597006

Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review*, *34*, 93-101. Retrieved from http://www.jstor.org/stable/2092790

Jordan, A., Trentacoste, N., Henderson, V., Manganello, J., & Fishbein, M. (2007). Measuring the time teens spend with media: Challenges and opportunities. *Media Psychology*, *9*, 19-41. doi:10.1080/15213260709336801

Juster, F. T., & Stafford, F. P. (1991). The allocation of time. Empirical findings, behavioral models, and problems of measurement. *Journal of Economic Literature*, *29*, 471-522.

Kirsh, S. J. (2012). *Children, adolescents, and media violence: A critical look at the research* (2nd ed.). Thousand Oaks, CA: Sage.

Krahé, B. (2014). Restoring the spirit of fair play in the debate about violent video games: A comment on Elson and Ferguson (2013). *European Psychologist*, *19*(1), 56-59. doi:10.1027/1016–9040/a000165

Krcmar, M., & Lachlan, K. (2009). Aggressive outcomes and videogame play: The role of length of play and the mechanisms at work. *Media Psychology*, *12*, 249-267. doi:10.1080/15213260903052257

Lee, C., Hornik, R., & Hennessy, M. (2008). The reliability and stability of general media exposure measures. *Communication Methods and Measures*, *2*, 6-22. doi:10.1080/19312450802063024

Lenhart, A., Kahne, J., Middaugh, E., MacGill, A., Evans, C., & Mitak, J. (2008). *Teens, video games and civics: Teens' gaming experiences are diverse and include significant social interaction and civic engagement*. Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/files/old-media//Files/Reports/2008/PIP_Teens_Games_and_Civics_Report_FINAL.pdf.pdf

Newson, R. (2002). Parameters behind "nonparametric" statistics: Kendall's tau, Somers' D and median differences. *The Stata Journal*, *2*(1), 45-64.

Nikkelen, S. W. C., Vossen, H. G. M., Valkenburg, P. M., Velders, F. P., Windhorst, D. A., Jaddoe, V. W. V., & Tiemeier, H. (2014). Media violence and children's ADHD-related behaviors: A genetic susceptibility perspective. *Journal of Communication*, *64*, 42-60. doi:10.1111/jcom.12073

Ogan, C., Karakus, T., & Kursun, E. (2013). Methodological issues in a survey of children's online risk-taking and other behaviours in Europe. *Journal of Children and Media*, *7*, 133-150. doi:10.1080/17482798.2012.739812

Olson, C. K., Kutner, L. A., Warner, D. E., Almerigi, J. B., Baer, L., Nicholi, A. M., II, & Beresin, E. V. (2007). Factors correlated with violent video game use by adolescent boys and girls. *Journal of Adolescent Health*, *41*, 77-83. doi:10.1016/j.jadohealth.2007.01.001

Özmert, E., Toyran, M., & Yurdakök, K. (2002). Behavioral correlates of television viewing in primary school children evaluated by the Child Behavior Checklist. *Archives of Pediatrics & Adolescent Medicine*, *156*, 910-914.

Paik, H., & Comstock, G. (1994). The effects of television violence on antisocial behavior: A meta-analysis. *Communication Research*, *21*, 516-546. doi:10.1177/009365094021004004

Pan European Game Information. (2012). *Annual report*. Retrieved from http://www.pegi.info/NL/index/id/media/pdf/390.pdf

Potter, W. J. (2008). The importance of considering exposure states when designing survey research studies. *Communication Methods and Measures*, *2*, 152-166. doi:10.1080/19312450802062299

Price, J., Palsson, C., & Gentile, D. A. (2014). What matters in movie ratings? Cross-country differences in how content influences mature movie ratings. *Journal of Children and Media*, *8*, 240-252. doi:10.1080/17482798.2014.880359

Prior, M. (2009). Improving media effects research through better measurement of news exposure. *Journal of Politics*, *71*, 893-908. doi:10.1017/S0022381609090781

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology*, *14*, 154-166. doi:10.1037/a0019440

Rideout, V. J., Foehr, U. G., & Roberts, D. F. (2010). *Generation M²: Media in the lives of 8- to 18-year-olds*. Retrieved from http://kaiserfamilyfoundation.files.wordpress.com/2013/01/8010.pdf

Robinson, J. P., & Godbey, G. (1997). *Time for life: The surprising ways Americans use their time*. University Park: Pennsylvania State University Press.

Rubin, A. M. (2009). Uses-and-gratifications perspective on media effects. In J. Bryant & M. B. Oliver (Eds.), *Media effects: Advances in theory and research* (3rd ed., pp. 165-184). New York, NY: Routledge.

Savage, J., & Yancey, C. (2008). The effects of media violence exposure on criminal aggression: A meta-analysis. *Criminal Justice and Behavior*, *35*, 772-791. doi:10.1177/0093854808316487

Schmitz, K. H., Harnack, L., Fulton, J. E., Jacobs, D. R., Gao, S. J., Lytle, L. A., & Van Coevering, P. (2004). Reliability and validity of a brief questionnaire to assess television viewing and computer use by middle school children. *Journal of School Health*, *74*, 370-377. doi:10.1111/j.1746-1561.2004.tb06632.x

Schutt, R. K. (2012). *Investigating the social world: The process and practice of research* (7th ed.). Thousand Oaks, CA: Sage.

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, *22*, 127-160. doi:10.1177/109821400102200202

Sherry, J. L. (2001). The effects of violent video games on aggression: A meta-analysis. *Human Communication Research*, *27*, 409-431. doi:10.1093/hcr/27.3.409

Slater, M. D. (2003). Alienation, aggression, and sensation seeking as predictors of adolescent use of violent film, computer, and website content. *Journal of Communication*, *53*, 105-121. doi:10.1093/joc/53.1.105

Slater, M. D. (2004). Operationalizing and analyzing exposure: The foundation of media effects research. *Journalism & Mass Communication Quarterly*, *81*(1), 168-183.

Slater, M. D., Henry, K. L., Swaim, R. C., & Cardador, J. M. (2004). Vulnerable teens, vulnerable times: How sensation seeking, alienation, and victimization moderate the violent media content-aggressiveness relation. *Communication Research*, *31*, 642-668. doi:10.1177/0093650204269265

Tamborini, R., Weber, R., Bowman, N. D., Eden, A., & Skalski, P. (2013). "Violence is a many-splintered thing": The importance of realism, justification, and graphicness in understanding perceptions of and preferences for violent films and video games. *Projections*, *7*(1), 100-118. doi:10.3167/proj.2013.070108

Valkenburg, P. M., Beentjes, H., Nikken, P., & Tan, E. (2002). Kijkwijzer: The Dutch rating system for audiovisual productions. *Communications*, *27*, 79-102. doi:10.1515/comm.27.1.79

Valkenburg, P. M., & Cantor, J. (2000). Children's likes and dislikes of entertainment programs. In D. Zillmann & P. Vorderer (Eds.), *Media entertainment: The psychology of its appeal* (pp. 135-152). Hillsdale, NJ: Lawrence Erlbaum.

Valkenburg, P. M., & Peter, J. (2013a). The differential susceptibility to media effects model. *Journal of Communication*, *64*, 221-243. doi:10.1111/jcom.12024

Valkenburg, P. M., & Peter, J. (2013b). Five challenges for the future of media-effects research. *International Journal of Communication*, *7*, 197-215.

Van der Voort, T. H. A., & Vooijs, M. W. (1990). Validity of children's direct estimates of time spent television viewing. *Journal of Broadcasting & Electronic Media*, *34*, 93-99.

Vandewater, E. A., & Lee, S. (2009). Measuring children's media use in the digital age: Issues and challenges. *American Behavioral Scientist*, *52*, 1152-1176. doi:10.1177/0002764209331539

Wallenius, M., & Punamaki, R. (2008). Digital game violence and direct aggression in adolescence: A longitudinal study of the roles of sex, age, and parent-child communication. *Journal of Applied Developmental Psychology*, *29*(4), 286-294. doi:10.1016/j.appdev.2008.04.010

Wallenius, M., Punamaki, R., & Rimpela, A. (2007). Digital game playing and direct and indirect aggression in early adolescence: The roles of age, social intelligence, and parent-child communication. *Journal of Youth and Adolescence*, *36*, 325-336. doi:10.1007/s10964-006-9151-5

Author Biographies

Karin M. Fikkers (MA, Radboud University Nijmegen, the Netherlands) is a PhD candidate at the Amsterdam School of Communication Research (ASCoR) at the University of Amsterdam. Her research focuses on the relationship between media violence exposure and aggressive behavior in adolescents. Specifically, she studies which adolescents are particularly vulnerable to the effects of media violence exposure.

Jessica Taylor Piotrowski (PhD, Annenberg School for Communication, University of Pennsylvania) is an Associate Professor at ASCoR at the University of Amsterdam. She is also the Director of the Center for Research on Children, Adolescents, and the Media (CcaM) at ASCoR. Her research investigates how children process and comprehend media content, with specific attention to the potential benefits of media. Her work has been presented at numerous conferences and published in communication, psychology, and education journals.

Patti M. Valkenburg (PhD, Leiden University, the Netherlands) is a distinguished university professor at the University of Amsterdam, and research professor at ASCoR. Her research involves the interplay between child development, media-specific parenting, and media effects. She has published more than 120 theoretical and empirical articles on the cognitive, emotional, and social effects of media and technology on children and adolescents. Her work has received numerous awards, including an Advanced Grant from the European Research Council and the Spinoza prize, the highest academic award in the Netherlands.