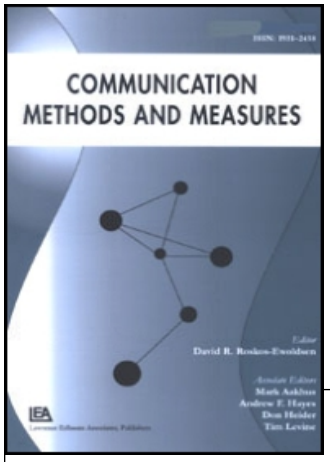

Communication Methods and Measures
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653633

The Reliability and Stability of General Media Exposure Measures

Chul-joo Lee [a]; Robert Hornik [a]; Michael Hennessy [b]
[a] Annenberg School for Communication, University of Pennsylvania,
[b] Annenberg Public Policy Center, University of Pennsylvania,

Online Publication Date: 01 January 2008

To cite this Article: Lee, Chul-joo, Hornik, Robert and Hennessy, Michael (2008) 'The Reliability and Stability of General Media Exposure Measures', Communication Methods and Measures, 2:1, 6 — 22

To link to this article: DOI: 10.1080/19312450802063024
URL: http://dx.doi.org/10.1080/19312450802063024

Communication Methods and Measures, *2*(1–2), 6–22, 2008

ISSN: 1931-2458 print / 1931-2466 online
DOI: 10.1080/19312450802063024

Routledge
Taylor & Francis Group

The Reliability and Stability of General Media Exposure Measures

Chul-joo Lee and Robert Hornik
Annenberg School for Communication, University of Pennsylvania

Michael Hennessy
Annenberg Public Policy Center, University of Pennsylvania

Relatively little research has examined the reliability and stability of media exposure measures. To address this issue, this study examines the reliability and stability of general media exposure measures using three rounds of the National Survey of Parents and Youth (NSPY), an in-home nationally representative panel survey of around 8,000 children (aged 9–18) and 6,000 of their parents. Using a structural equation model first suggested by Wiley and Wiley (1970), we find that the measures for youth show moderate reliability with latent variables showing moderate to high stability. For parents, reliability is moderate and but latent variable stability is high. The implications of the findings for future research in this area are discussed.

In the field of media effects research, a fundamental problem is the quality of media exposure measures. Most media effects studies examine the effects that individual differences in media exposure have on diverse cognitive, attitudinal, and behavioral consequences. Insofar as the media exposure measures employed in these studies are problematic, it is difficult to assess outcomes of individuals' media exposure behaviors. Therefore, quality assessments of media exposures from respondents are a critical precursor for accurate media effects research (McGuire, 1986; Slater, 2004).

Correspondence should be addressed to Chul-joo Lee, Annenberg School for Communication, 3620 Walnut Street, Philadelphia, PA 19104. E-mail: clee@asc.upenn.edu

There are four standards of media exposure measures: *validity*, *fineness of distinctions*, *reliability*, and *stability*. Validity captures the match between conceptualization of a variable and its measure and fineness of distinctions refers to the ability of a measure to capture causally meaningful differences (but not unimportant differences) among respondents (Carmines & Zeller, 1979). There is much research on the validity of media exposure measures (e.g., Allen, 1981; Allen & Taylor, 1985; Johnson, 1982; Rimmer & Weaver, 1987; Salomon & Cohen, 1978; Slater, 2004; Van Mierlo & van den Bulck, 2004) but less about issues of fineness of distinctions.

This paper does not deal with validity and fineness of distinction in a central way. Rather, we focus on the second pair of issues: *reliability* and *stability*. This study first examines how these two standards are conceptually different. Second, based on our critical review of previous studies in this area, we adopt a structural equation model first proposed by Wiley and Wiley (1970) in order to separately assess reliability and stability of general media exposure measures.

RELIABILITY

Reliability refers to the notion that re-elicitation of the same measure from a respondent should produce an identical response as long as the underlying true exposure is unchanged. Although there is universal recognition of the need for reliable measures, the reliability issue arises in the media effects research arena because exposure measures are likely to include measurement errors. The measurement errors in studies of theoretical variables may alter the substantive conclusion of the research (Carmines & Zeller, 1979; McGuire, 1986). Unreliable measurements of media exposure always attenuate exposure's relationship to other variables and almost always lead to underestimation of exposure effects on outcomes. For example, Bartels (1993) attributed the negative findings and nonfindings in the field of electoral political communication, in part, to carelessness regarding the inferential consequences of measurement errors of campaign exposure. He showed that adjusting for measurement error increased the effect of media exposure on opinion change in a presidential campaign. Also, Henderson (2006) only found it possible to detect prospective effects of television watching on obesity by explicitly incorporating measurement error in the analyses. In contrast, Ritchie, Price, and Roberts (1987) indicated that when they took into consideration the unreliability of their measures other than TV use via structural equation models, the apparent relationships between TV use and reading achievement almost disappeared. These findings clearly show that considering the measurement errors of media exposure measures in media effects research can increase the validity of research findings.

STABILITY

Stability refers to the tendency for the underlying variable to be unchanged over time (for an overview, see Allen, 1981). Stability is important for assessment of measurement quality for two reasons. If one uses a test-retest association to estimate measure reliability, then one needs to be able to assess stability of the underlying measure. The obtained test-retest correlation coefficient reflects at least two sources of influence: random error in the measurement and a real change of the behavior over time. The first reflects the reliability of the measurement instrument, whereas the second taps the stability of the underlying behavior (Heise, 1969; Wiley & Wiley, 1970). Stability defines the upper limit of the correlation and thus the estimates of reliability need to be adjusted accordingly.

Stability is also important for substantive reasons. Many theories of media effects are about the effects of media exposure across time. However, studies testing those theories often measure exposure only at one time point but use it as if it were an estimate of longer-term exposure. They assume that such exposure is influential because the measure represents likely exposure before and after the moment of measurement. If a one-time measure is being used to capture stable exposure, then it is important to show that the measure *does* represent longer term exposure, that is, that it provides an accurate estimate of a stable exposure. If there is high stability in the variable, we can assume that measurements of media exposure at one time point also reflect media exposure at later time points. Then, we can be assured that "the responses are neither capricious nor so highly linked to the immediate situation that they have no predictive power" (Allen, 1981, p. 237).

Thus, an assessment of the usefulness of an exposure measure needs to include how reliable it is and how well it captures a stable behavior. The assessment of these two characteristics of a measure is the main goal of this paper.

ASSESSING RELIABILITY AND STABILITY OF MEDIA EXPOSURE MEASURES

Communication researchers have adopted diverse approaches to media exposure measurement. Some (e.g., Allen & Taylor, 1985; Stamm & Jacoubovitch, 1980) have focused narrowly on specific contents or genres, discarding the previous time-spent measures for media in general. Others go beyond exposure and include some proxies such as attention or knowledge (e.g., Chaffee & Schleuder, 1986; Drew & Weaver, 1990; Price & Zaller, 1993; cited in Chernin, Fishbein, Bleakley, Davis, & Stevens, 2006). Still others have experimented with different time periods for their media exposure questions and suggest that more recent or shorter time spans, such as *past week*, are more valid and reliable ways to capture

individuals' media use (Price, 1993). Finally, a number of media researchers have used methods other than respondent self-reports such as observations and diaries (for an overview, see Chernin et al., 2006; Webster & Wakshlag, 1985).

Previous studies have tested the convergent validity of these different media exposure measures by using the correlations among them. It has been found that, compared with presumably more precise measures (i.e., observation or diary), self-reported media exposure estimates are systematically biased upwards (Price, 1993). The media use levels that people tend to report are higher than reported levels that correspond to more objective measures, such as surveillance cameras and diaries. Despite this bias, a number of previous validation efforts strongly suggest there exist moderate to strong correlations among these alternative media exposure assessments (Alexander, Wartella, & Brown, 1981; Anderson, Field, Collins, Lorch, & Nathan, 1985; Bechtel, Achelpohl, & Akers, 1972; Kay, 1972; Price, Ritchie, Roberts, & Lieberman, 1986; Ritchie et al., 1987; Rubin, 1976; Southwell, Barmada, Hornik, & Maklan, 2002; van der Voort & Vooijs, 1990).

Researchers have also used a different strategy for measure validation by using as their validity criterion the association between their media exposure measures and their presumed outcomes, such as political knowledge, public affairs knowledge, and health knowledge (e.g., Culbertson & Stempel, 1986; Dutta-Bergman, 2004; McLeod et al., 1996; Romantan, Hornik, Weiner, Price, & Cappella, 2005; Schooler, Chaffee, Flora, & Roser, 1998).

Relatively little is known about either the reliability or the stability of many studies' media exposure measures, particularly if one demands that reliability requires assessment of consistency of response over time. It is customary to define reliability in terms of internal consistency of multiple items at a single point in time (i.e., using Cronbach's alpha or other related summary statistics; see Streiner, 2003). However, this is clearly second best; single time point measures are subject to common extraneous influences, which exaggerate internal consistency.

In general, media effects researchers have addressed only tangentially the unreliability of their media exposure measures, regardless of what definition or operationalization that they have used. Chernin et al. (2006, p. 3) stated that "this is particularly surprising given that McGuire (1986) and others (e.g., Bartels, 1993; Zaller, 1996) have argued that unreliable exposure measures may account for many of the field's 'limited effects' findings."

Moreover, existing media effects studies operate under the prevailing assumption that media effects are cumulative and that exposure measures obtained at one point incorporate information about long-term exposure; however, researchers have paid little attention to the stability of media exposure behavior itself. If media exposure behavior varies greatly from one sampled time interval to the next, the use of values obtained at any single time point as a person's characteristic behavior can be problematic (Allen, 1981; Allen & Taylor, 1985; Price, 1993).

Although small in number, a few studies address the reliability and the stability of media exposure measures. Allen (1981), for instance, assessed the reliability and the stability of his television exposure measures and found that the indicators of television exposure are moderately reliable, whereas the television exposure construct is substantially unstable. Allen and Taylor (1985) examined the reliability and the stability of their measures of exposure to newspaper and television public affairs, and demonstrated that measures of newspaper public affairs exposure are more reliable than those of television public affairs exposure. With regard to stability, their findings pointed to the opposite pattern. That is, the television public affairs construct was found to be more stable than the newspaper public affairs construct. More recently, Chernin and her colleagues (2006) estimated the reliability of four measures of media exposure across six media (television, Internet, music, video games, newspapers, and magazines). Their measures include a frequency measure (never, rarely, sometimes, often), an open-ended measure, a global measure, and a check-box measure. They concluded that all the four measures of media exposure demonstrate consistently high test-retest reliability.

These studies are limited in some respects. First, researchers have conducted only a handful of studies to test whether media exposure behaviors can be regarded as individuals' stable characteristics (i.e., traits). Second, it seems that most studies did not separate the reliability of their measures from the stability of the measures' underlying constructs. Given that most of these studies were limited to two rounds of data collection, they could only rely on test-retest associations. As noted above, test-retest reliability estimates based on the test-retest correlation may underestimate true reliability because the estimate is affected by two factors: measurement errors and temporal instability of the underlying behavior (Heise, 1969; Wiley & Wiley, 1970). Third, many of these studies (e.g., Allen, 1981; Chernin et al., 2006) are not based on a nationally representative sample. Therefore, their results may not represent reliability and stability in the population.

In addition, although a few media scholars (e.g., Hawkins & Pingree, 1997; Shah, McLeod, & Yoon, 2001) have raised concerns about commonly used Internet measures (i.e., how frequently or how many hours people use the Internet), not many studies have assessed the reliability and the stability of Internet exposure measures. Han and his colleagues (2006) and Newhagen and Rafaeli (1996) discussed the difference of Internet use from other mass media consumption in greater detail. They contended that the Internet is able to tailor its contents according to audience's interests, whereas mass media diffuse the same content to everyone. Thus, in comparisons with other mass media audiences, Internet users can play more active roles. Based on these differences, scholars (e.g., Hawkins & Pingree, 1997; Shah et al., 2001) have argued that simple time use measures of the Internet are problematic in terms of reliability and validity.

To redress these issues, this paper assesses the reliability and the stability of a particular class of general media exposure measures. We use *self reports of media exposure*, which are typical of the most popular media exposure measures in existing media effects literature. Although researchers assume that direct observations or diaries of mass communication behavior are stronger in terms of reliability and validity, these measurements are generally "cost-prohibitive and impractical in most survey settings" (Price, 1993, p. 615). Therefore, researchers most commonly ask people about their general media exposure, such as how many hours (or how frequently) they use a certain medium on an "average day" or in a "typical week." Our study focuses on these general media exposure measures and implicitly assumes but does not test, the validity of these measures if they are shown to be reliable and stable.

METHODS

Sample

This paper uses data from the National Survey of Parents and Youth (NSPY), an in-home survey designed to be representative of children aged 9–18 and their parents. The purpose of NSPY was to assess the effectiveness of the National Youth Anti-Drug Media Campaign, which targeted youth aged 9–18, their parents, and other influential adults. NSPY questionnaires were administered in respondents' homes using touch-screen laptop computers.

We use a three-round dataset collected from youth 9 to 18 and one of their parents. During the first round of data collection NSPY enrolled three separate nationally representative samples of youth aged 9 to 18 and their parents. Sampling of eligible youth in those baseline waves was designed to produce approximately equal-sized samples within three age subgroups (9–11, 12–13, and 14–18). One or two youth were randomly selected from each eligible sample household. One parent was randomly chosen from each eligible household. A second parent was selected in the rare event when two youths who were not siblings were sampled.

Wave 1 included 3,298 youth aged 9–18 years old and 2,284 of their parents, who were interviewed between November 1999 and May 2000; Wave 2 included 2,361 youth and 1,632 of their parents interviewed between July and December 2000; Wave 3 included 2,458 youth and 1,682 of their parents interviewed between January and June 2001.

All of these youths and parents who could be recontacted and were still in the eligible age range were then interviewed approximately one, two, and three years after the baseline measurement wave. The second round involved interviews with 6,518 youth and 4,834 parents; round three included 5,854 youth and 4,261

parents. The NSPY achieved a response rate of 65% for youth and 63% for parents in round one, with follow-up response rates of 86% and 84% for youth and parents in round two, and with 91–93% of remaining eligible youth and parents in round three.

Measures

Here we employed six general media exposure measures for both parents and youths and added one general media use measure only for the parent sample. We asked youths about their weekday TV use, weekend TV use, weekday radio use, weekend radio use, magazine reading, and Internet use. Only teens (aged 12–18) were asked about their exposure to radio, magazines, and the Internet; children (aged 9–11) were asked only about their TV use. In the parent sample, we asked for their frequency of newspaper reading as well as of weekday TV use, weekend TV use, weekday radio use, weekend radio use, magazine reading, and Internet use.

With regard to TV use and radio use, we used time-spent measures. Respondents including parents and youths were asked how many hours they watched TV on an average weekday (and an average weekend) and how many hours they listened to radio on an average weekday (and an average weekend). In order to tap respondents' weekday TV use and weekday radio use, they were asked to rate on a nine-point scale (1 = none, 2 = half-hour or less, 3 = about 1 hour, 4 = about 2 hours, 5 = about 3 hours, 6 = about 4 hours, 7 = about 5 hours, 8 = a bout 6 hours, 9 = 7 hours or more) but for weekend TV use and weekend radio use, we asked respondents to present themselves on an eight-point scale (1 = none, 2 = less than 1 hour, 3 = 1–2 hours, 4 = 3–4 hours, 5 = 5–6 hours, 6 = 7–8 hours, 7= 9–10 hours, 8 = 11 hours or more).

Respondents in both the parent and the youth sample were asked how often they had used the Internet in the last 6 months. They marked their responses on a five-point scale (1 = never, 2 = a few times a year, 3 = once or twice a month, 4 = at least once a week, 5=every day or almost every day). Likewise, we asked both parents and youths how often they read magazines. Then, we presented them with a five-point scale (1 = never, 2 = a few times a year, 3 = once or twice a month, 4 = at least once a week, 5 = every day or almost every day).

Only parents were asked how often they read newspapers. They answered this question on a five-point scale (1 = never, 2 = a few times a year, 3 = once or twice a month, 4 = at least once a week, 5 = every day or almost every day).

Analysis Strategy

With single-indicator measures, test-retest reliability has been the most widely used method. This method is based on repeated employment of the same measures over time with the same individual, and shows the extent to which the measures can

produce consistent values over time (Anastasi, 1982; The Standards for Educational and Psychological Testing, 1999). However, test-retest reliability may not be true reliability because it reflects not only measurement errors but also the temporal instability of the behavior or the phenomenon that the measures are expected to tap. Recognizing this limitation, some researchers have adopted shorter time intervals between measurements in order to eliminate the effect of the true scores' temporal instability (Anderson et al., 1985; Chernin et al., 2006; Schmitz et al., 2004). Assuming that the behaviors of their interests are less likely to change in the short term, these researchers attained reliability estimates based on the correlation coefficients of the values over a short test-retest interval (e.g., a week). In contrast, stability estimates are obtained by conducting repeated measurements over a considerably longer period, ranging from 3 months to 2 years (Heise, 1969), and Webb, Redman, Gibberd, and Sanson-Fisher (1991) applied both of the methods to the same population and obtained estimations of both measurement error and change in behavior over time.

However, because the measurement interval in our data spans about one year, use solely of the test-retest method for calculating the reliability of our general media exposure measures seems inappropriate. Moreover, the test-retest reliability test is limited in that the repeated employment of same measures over time likely confounds reliability with the effects of respondents' memory. That is, respondents' memory rather than true reliability may explain observed correlation coefficients (Carmines & Zeller, 1979; Heise, 1969; Nunnally, 1978).

Based on these considerations, we adopted the path model proposed by Wiley and Wiley (1970). This model distinguishes the effects of reliability from those of stability by using a path model when same measures are used at least three times. Wiley and Wiley's model assumes that observed scores are affected by the true score and measurement error, and that the relationship between true scores over time are also affected by external factors that may reduce stability. As a result, these measurement errors and external factors may decrease correlations among observed scores across waves. Figure 1 shows the generic path model.

To estimate the reliability and stability coefficients with this path model, it is necessary to make some simplifying assumptions to ensure identifiability of the path model because the total degrees of freedom are only six (i.e., three variances and three covariances). Following Wiley and Wiley (1970), we assume that errors of measurement are independent of true scores and independent of each other over time. These assumptions lead to the expectation that error variance (but not reliability) will be constant across measurement time points, and that there will be no association of errors across time. In addition, this model assumes that all effects of exposure at time 1 affect exposure at time 3 through exposure at time 2, so that the stability of time 1 to time 3 is the product of the two estimated stability regression coefficients, $\gamma 21$ and $\gamma 32$. Reliability in the path model is the percent of variance of the indicator that is due to the true score. For single indicator

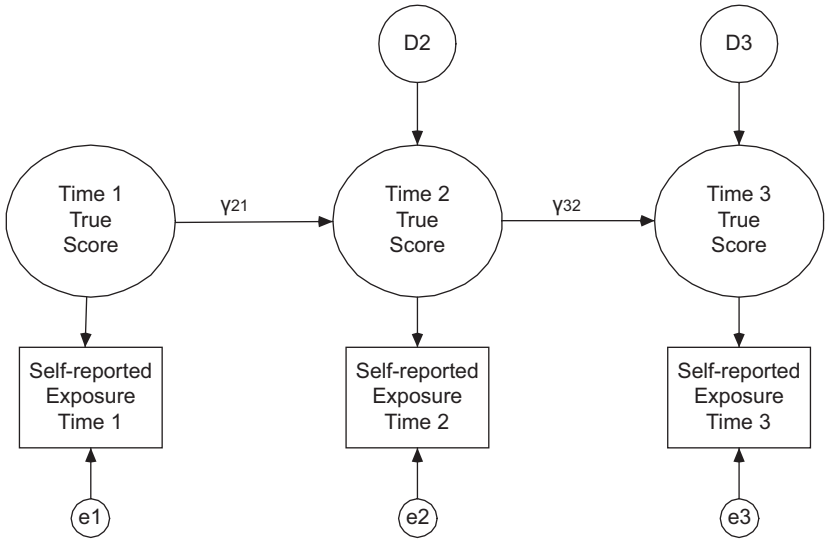

FIGURE 1 Generic Model of Exposure over Time (Wiley & Wiley, 1970).

measurement models like this one, this is the R^2 of the indicator when predicted by the true score (Kline, 2005). Using this path model and the assumptions, Wiley and Wiley described a formulaic approach to obtaining separate estimates of the reliability and the stability of repeated measures. The coefficients can be estimated algebraically without use of a structural model estimating program. The relevant set of equations and their solution appears in Table 1.

However, because we wanted to attain confidence intervals for all coefficients, we used Mplus to estimate the reliability and stability coefficients and their confidence intervals, which were estimated using the bootstrap method.

RESULTS

Table 2 indicates the means, standard deviations, and sample sizes as well as the correlation/covariance matrix. We treated the ordinal measures of exposure as if they were interval, to avoid making speculative assignments of interval equivalents to ordinal categories. However, we found that the pattern of correlations was substantially unchanged even when we did assign such values.

Table 3 displays the reliability estimates and the stability estimates. In general, we found that the reliability estimates of our general media exposure measures were low to moderate, between .54 and .66 for youth and from .49 to .87 for parents, Internet use among parents was measured more reliably than any other

TABLE 1
Solving for Reliability and Stability (Wiley & Wiley, 1970)

Structural Equation model:

$X_1 = \theta_1 + \varepsilon_1$
$X_2 = a_{21}\theta_1 + \theta_2 + \varepsilon_2$
$X_3 = a_{32}(a_{21}\theta_1 + \theta_2) + \theta_3 + \varepsilon_3$

Where Xi is exposure at time I; θ_1 is true score at time 1, and θ_2, and θ_3 represent new true exposure at times 2 and 3. ε_i is error at time i; a_{ij} is the stability effect of prior true score on subsequent exposure.

If error variance is constant, and error is uncorrelated with true score and with error at subsequent waves, these structural equations produce the following equalities:

$V(X_1) = V(\theta_1) + V(\varepsilon)$
$C(X_1X_2) = a_{21}\, V(\theta_1)$
$C(X_1X_3) = a_{21}\, a_{32} V(\theta_1)$
$V(X_2) = a^2_{21}\, V(\theta_1) + V(\theta_2) + V(\varepsilon)$
$C(X_2X_3) = a_{32}[a^2_{21}\, V(\theta_1) + V(\theta_2)]$
$V(X_3) = a^2_{32}[a^2_{21}\, V(\theta_1) + V(\theta_2)] + V(\theta_3) + V(\varepsilon)$

Where V( ) refers to the Variance and C( ) to the Covariance.

These equalities can then be solved for the needed unknowns:

Parameter	*Estimator*
a_{32}	$C(X_1X_3)/C(X_1X_2)$
$V(\varepsilon)$	$V(X_2) - [C(X_2X_3)/a_{32}]$
$V(\theta_1)$	$V(X_1) - V(\varepsilon)$
a_{21}	$C(X_1X_2)/V(\theta_1)$
$V(\theta_2)$	$V(X_2) - [a_{21}C(X_1X_2) + V(\varepsilon)]$
$V(\theta_3)$	$V(X_3) - [a_{32}C(X_2X_3) + V(\varepsilon)]$

behavior (from .85-.87) but we think the reliability of that measure reflects the large number of parents (about 28.7% on average over time) who did not use the Internet while youth nonusers of the Internet were fewer (12.0% nonusers on average over time). For all of the other measures, consistent zero use was rare, and as a result consistent responses were harder to come by. In most cases, reliability was fairly stable across measurement occasions. Parents produced higher reliabilities for weekday TV and magazine use, but youth were a little more reliable in reporting weekend radio.

Table 3 also displays the stability measures. Youth reported moderately stable behavior across media uses (in the range of .62–.80 between times 1 and 2, and .82–.90 between times 2 and 3). Parents reported very stable behavior, with some increase in stability for the 2–3 period versus the 1–2 period; it was only for the parents sample that the confidence intervals of stability overlapped with unity (for weekend radio and magazines). For both youth and parents, stability between times 1 and 3 was less than that between the single-round lags because it is a function of the definition of the stability between times 1 and 3. Indeed, some of

TABLE 2
Sample Means, Standard Deviations, Correlations, and Covariances for Reported General Media Exposure Measures at Three Rounds of the NSPY Survey

	Youth					*Parent*				
	M	*SD*	*R1*	*R2*	*R3*	*M*	*SD*	*R1*	*R2*	*R3*
Weekday TV										
R1	4.6	1.98	3.91	1.71	1.51	4.11	1.85	3.42	1.82	1.68
R2	4.60	1.96	.44	3.83	2.02	4.07	1.76	.56	3.11	1.85
R3	4.55	1.93	.40	.54	3.73	4.15	1.77	.52	.59	3.12
N			5,558					3,731		
Weekend TV										
R1	4.80	1.77	3.13	1.35	1.14	4.29	1.63	2.66	1.46	1.41
R2	4.73	1.78	.43	3.17	1.55	4.32	1.63	.55	2.67	1.55
R3	4.65	1.75	.37	.50	3.07	4.35	1.61	.54	.59	2.59
N			5,559					3,732		
Weekday radio										
R1	3.65	2.17	4.70	1.99	1.64	3.99	2.36	5.57	2.79	2.37
R2	3.91	2.23	.41	4.98	2.34	4.09	2.33	.51	5.43	2.75
R3	3.98	2.27	.33	.46	5.15	3.93	2.23	.45	.53	4.98
N			3,143					3,730		
Weekend radio										
R1	3.46	1.87	3.49	1.69	1.43	3.08	1.72	2.94	1.46	1.33
R2	3.79	1.95	.47	3.80	2.06	3.18	1.74	.49	3.03	1.50
R3	3.84	2.00	.38	.53	3.99	3.11	1.65	.47	.52	2.72
N			3,143					3,729		
Newspaper										
R1	–	–	–			3.90	1.26	1.58	1.05	.96
R2	–	–	–	–		3.91	1.27	.66	1.61	1.05
R3	–	–	–	–	–	3.96	1.24	.61	.67	1.55
N			–					3,731		
Magazine										
R1	3.05	1.12	1.26	.54	.48	3.34	1.15	1.33	.72	.69
R2	3.05	1.10	.43	1.22	.64	3.30	1.12	.56	1.26	.76
R3	3.03	1.11	.39	.52	1.24	3.28	1.13	.53	.60	1.28
N			3,143					3,732		
Internet										
R1	3.39	1.42	2.01	.82	.72	3.08	1.69	2.84	2.00	1.73
R2	3.77	1.33	.44	1.77	.95	3.34	1.63	.72	2.67	2.00
R3	3.88	1.33	.38	.95	1.77	3.51	1.58	.65	.78	2.49
N			3,119					3,721		

Notes: 1) R1 refers to time 1, R2 for time 2, and R3 for time 3.
2) All the Pearson correlation coefficients are statistically significant at the level of .001.
3) For youths, we did not measure their frequency of newspaper reading.
4) Values on the diagonal are variances, above are covariances and below are correlations.

TABLE 3
Reliability and Stability Estimates for the General Media Exposure Measures

	Youth						*Parent*					
	Reliability			*Stability*			*Reliability*			*Stability*		
Exposure Measure	*R1*	*R2*	*R3*	*γ21*	*γ32*	*γ31*	*R1*	*R2*	*R3*	*γ21*	*γ32*	*γ31*
Weekday TV	.61 (.05)	.60 (.05)	.59 (.05)	.72 (.05)	.88 (.05)	.63 (.07)	.67 (.04)	.64 (.05)	.64 (.05)	.79 (.05)	.93 (.06)	.73 (.07)
Weekend TV	.57 (.05)	.58 (.04)	.57 (.05)	.75 (.06)	.85 (.06)	.64 (.08)	.60 (.04)	.60 (.04)	.59 (.04)	.92 (.07)	.96 (.05)	.87 (.09)
Weekday radio	.54 (.08)	.57 (.07)	.58 (.07)	.78 (.10)	.82 (.08)	.64 (.12)	.61 (.05)	.60 (.05)	.56 (.05)	.83 (.07)	.85 (.06)	.70 (.10)
Weekend radio	.61 (.07)	.64 (.06)	.66 (.06)	.80 (.08)	.85 (.07)	.68 (.11)	.51 (.07)	.54 (.06)	.49 (.07)	1.02 (.06)	.89 (.06)	.90 (.16)
Newspaper		–			–		71 (.04)	.72 (.04)	.71 (.04)	.94 (.05)	.91 (.06)	.86 (.06)
Magazine	.60 (.06)	.59 (.06)	.59 (.06)	.71 (.06)	.90 (.08)	.64 (.09)	.65 (.03)	.63 (.04)	.64 (.04)	.84 (.06)	.96 (.05)	.81 (.07)
Internet	.66 (.06)	.61 (.06)	.61 (.06)	.62 (.05)	.87 (.07)	.54 (.07)	.87 (.02)	.86 (.03)	.85 (.03)	.81 (.03)	.87 (.03)	.71 (.03)

Notes: γ 21 refers to the stability between times 2 and 1; γ 32 the stability between times 3 and 2; γ 31 the stability between times 3 and 1. All confidence intervals (± term in parentheses) estimate the 95% probability range of plausible values.

the greater stability between the second pair of measurements compared to the first pair of rounds may reflect the somewhat shorter time period of the second pair. The average months between times 1 and 2 measures was 15 months but for times 2 and 3 it was 12 months.

It should be noted that some of our findings are not consistent with Allen and Taylor's (1985) findings. Allen and Taylor showed that the indicators of the newspaper construct had moderately high reliability, whereas the television indicators yielded lower reliability estimates. However, our results showed only small differences in reliability estimates between TV and newspaper. Both of the measures in this study seemed to perform moderately well.

More important, we found that Internet use measures had high reliability estimates, in contrast to the existing concerns that time spent or frequency measures of the Internet give rise to significant measurement errors. As noted above, this may reflect the presence of stable nonusers in the sample. In addition, our findings regarding the stability of the TV measure and of the newspaper measure are not consistent with the related findings of Allen and Taylor (1985). Whereas Allen and Taylor found that the television construct was more stable than the newspaper construct, we found just the opposite.

DISCUSSION

This study examined the reliability and the stability of general media exposure measures. For youth, the measures showed moderate reliability, with the underlying variables showing moderate to high stability, particularly between times 2 and 3. For parents, measure reliability was moderate, but underlying variable stability was higher. For youth, media exposure behaviors may be in the formative stages. In contrast, it is quite natural that parents exhibit patterns of media exposure behavior that are relatively stable and well-established. Indeed, insofar as youths have relatively little experience using diverse print media such as magazines, the moderately high levels of stability observed in this study are even surprising. In summary, we conclude that general media exposure measures perform only moderately in terms of their reliability, regardless of whether respondents are parents or youths. If the association of these measures with outcomes is to be tested, any observed association might be a substantial underestimate of the true association. For example, if the reliability of a media exposure measure is .60, and of the outcome measure is .80, and the true correlation between them is .30, the observed correlation will be .21, an underestimate by nearly one-third ($.30 \times \sqrt{.60} \times \sqrt{.80}$). If such measures are to be used, the need for adjusting estimates through the incorporation of measurement models is obvious.

The stability estimates support a more optimistic interpretation. These suggest that once these general exposure measures are corrected for unreliability, the

underlying variables are reasonably stable, particularly for adults. This suggests that these measures are not merely capturing immediate behavior but reflect longer-term exposure. They are useful estimators for use in stable effects theories, which are tested with single time measures, once estimates are adjusted for unreliability. They do seem to capture differences in exposure among individuals that is maintained over time.

With regard to the Internet specifically, although a group of media scholars (e.g., Han et al., 2006; Hawkins & Pingree, 1997) argued that time-spent measures or frequency measures of the Internet are problematic in terms of measurement quality, our results showed that both the reliability and the stability of Internet measures for adults, at least, are quite acceptable. This finding assures us that we can rely on our simple, general measure of people's Internet use. Our findings can free new media scholars from unreasonably pessimistic views of their simple Internet use measures. Of course, we do not mean to imply that it is not valuable to develop alternative Internet use measures (e.g., using log files on the Web) for some purposes. A general exposure measure will be of limited use if the content of Internet exposure is poorly understood.

LIMITATIONS

Our findings are based on a nationally representative sample of youth 9–18 and of their parents studied as a cohort for three to four years. This is a real strength. However, we have not corrected our estimates for the complex sample design. Therefore, we cannot claim that we were fully exploiting the representativeness of these samples.

Second, these parents are not meant to be representative of all adults – they are a particular subgroup of adults who are in a limited age range (i.e., half of our respondents were from 37 to 46) defined by the fact that they have children aged between 9 and 18. It is also possible that their media use behavior is affected not only by their own ages but also because they live with youth of this age. We probably cannot generalize to all adults. Similarly, our youth sample includes only children 9–18 obviously excluding younger children.

Third, we are looking only at a narrow class of measures of general media use; nothing we say can be applied to other measures – for example, of content-specific use.

Finally, all of our estimates depend on the credibility of the assumptions. The Wiley and Wiley (1970) assumptions, while they reflect classical measurement theory, may not be correct. Our stability estimates were relatively high. This finding may partly reflect the incorporation of Wiley and Wiley's assumption about no correlated measurement error. They and we assumed that there exist no serial correlations among the measurement errors. It is possible, however, that if

respondents tend to over- or underestimate their media use at one point in time, that the tendency is maintained at a second point in time. This will produce an overestimate of stability, since we cannot separate the effects of such a bias from true stability in the underlying behavior. With additional measures of the same construct at one point in time, it would be possible to test these assumptions and see if a less constrained path model produces different reliability and stability estimates.

ACKNOWLEDGEMENT

Data collection for this article was supported by the National Institute on Drug Abuse through contract 3-N01-DA085063-002.

REFERENCES

Alexander, A., Wartella, E., & Brown, D. (1981). Estimates of children's television viewing by mother and child. *Journal of Broadcasting, 25,* 243–251.

Allen, R. L. (1981). The reliability and stability of television exposure. *Communication Research, 8,* 233–256.

Allen, R. L., & Taylor, B. F. (1985). Media public affairs exposure: Issues and strategies. *Communication Monographs, 52,* 186–201.

American Education Research Association, American Psychological Association, and the National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: Macmillan.

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.

Bartels, L. M. (1993). Messages received: The political impact of media exposure. *The American Political Science Review, 87,* 267–285.

Bechtel, R. B., Achelpohl, C., & Akers, R. (1972). Correlates between observed behavior and questionnaire responses on television viewing. In E. A. Rubinstein, G. A. Comstock, & J. P. Murray (Eds.), *Television and social behavior: Vol. 4. Television in day-to-day life: Patterns of use*. Washington, DC: U.S. Government Printing Office.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment.* Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-017. Newbury Park, CA: Sage.

Chaffee, S. H., & Schleuder, J. (1986). Measurement and effects of attention to media news. *Human Communication Research, 13,* 76–107.

Chernin, A., Fishbein, M., Bleakley, A., Davis, E., & Stevens, R. S. (2006, June). *Test-retest reliability of four measures of media exposure*. Paper presented at the annual meeting of the International Communication Association in Dresden, Germany.

Culbertson, H., & Stempel, G. (1986). How media use and reliance affect knowledge level. *Communication Research, 13,* 38–47.

Drew, D., & Weaver, D. (1990). Media attention, media exposure, and media effects. *Journalism Quarterly, 67,* 740–748.

Dutta-Bergman, M. J. (2004). Primary sources of health information: Comparisons in the domain of health attitudes, health cognitions, and health behaviors. *Health Communication, 16*, 273–288.

Han, J. Y., Hawkins, R., Shaw, B., Pingree, S., McTavish, F., & Gustafson, D. (2006, August). *Unraveling uses and effects of an interactive cancer communication system*. Paper presented at the annual meeting of the Association for Education in Journalism & Mass Communication in San Francisco.

Hawkins, R. P., & Pingree, S. (1997). *Measuring time frames of communication behaviors in computer use*. Paper presented at the annual meeting of the International Communication Association in Toronto, Canada.

Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review, 34*, 93–101.

Henderson, V. (2006). *Exploring TV viewing and overweight in pre-adolescent and adolescent girls*. Doctoral dissertation, University of Pennsylvania. Retrieved March 22, 2007, from http://repository.upenn.edu/dissertations/AAI3225469/

Johnson, J. D. (1982). The dimensionality of readership measures. *Communication Research, 9*, 607–614.

Kay, H. (1972). Weaknesses in the television-causes-aggression analysis by Eron et al. *American Psychologist, 27*, 970–973.

Kline R. (2005). *Principles and practice of structural equation modeling*. New York: Guilford.

McGuire, W. J. (1986). The myth of massive mass media impact: Savagings and salvagings. In G. Comstock (Ed.), *Public communication and behavior* (Vol. 1, pp. 173–257). New York: Academic Press.

McLeod, J. M., Daily, K., Guo, A., Eveland, W. P. Jr., Bayer, J., & Yang, S. (1996). Community integration, local media use, and democratic process. *Communication Research, 23*, 179–209.

Newhagen, J., & Rafaeli, S. (1996). Why communication researchers should study Internet. *Journal of Communication, 46* (1), 1–13.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Price, V. (1993). The impact of varying reference periods in survey questions about media use. *Journalism Quarterly, 70*, 615–627.

Price, V., Ritchie, D., Roberts, D. F., & Lieberman, D. (1986, May). *The Stanford reading and television study: A progress report*. Paper presented at the International Communication Association in Chicago.

Price, V., & Zaller, J. (1993). Who gets the news? Alternative measures of news reception and their implications for research. *Public Opinion Quarterly, 57*, 133–164.

Rimmer, T., & Weaver, D. (1987). Different questions, different answers? Media use and media credibility. *Journalism Quarterly, 64*, 28–44.

Ritchie, D., Price, V., & Roberts, D. F. (1987). Television, reading, and reading achievement: A reappraisal. *Communication Research, 14*, 292–315.

Romantan, A., Hornik, R., Weiner, J., Price, V., & Cappella, J. (2005, May). *Learning about cancer: A comparative analysis of the performance of complementary and alternative measures of incidental exposure to the mass media*. Paper presented at the annual meeting of the International Association of Communication in New York.

Rubin, A. M. (1976). *A developmental examination of the uses of television by children and adolescents*. Doctoral dissertation, University of Illinois-Urbana Champaign.

Salomon, G., & Cohen, A. A. (1978). On the meaning and validity of television viewing. *Human Communication Research, 4*, 265–270.

Schmitz, K. H., Harnack, L., Fulton, J. E., Jacobs, D. R., Gao, S., & Lytle, L. A. (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.

Schooler, C., Chaffee, S. H., Flora, J. A., & Roser, C. (1998). Health campaign channels: Tradeoffs among reach, specificity, and impact. *Human Communication Research, 24*, 410–432.

Shah, D. V., McLeod, J. M., & Yoon, S. H. (2001). Communication, context, and community: An exploration of print, broadcast, and Internet influences. *Communication Research, 28*, 464–506.

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

Southwell, B. G., Barmada, C. H., Hornik, R. C., & Maklan, D. M. (2002). Can we measure encoded exposure? Validation evidence from a national campaign. *Journal of Health Communication, 7*, 445–453.

Stamm, K. R., & Jacoubovitch, M. D. (1980). How much do they read in the daily newspaper: A measurement study. *Journalism Quarterly, 57*, 234–242.

Streiner, D. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*, 99–103.

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.

Van Mierlo, J., & Van den Bulck, J. (2004). Benchmarking the cultivation approach to video game effects: A comparison of the correlates of TV viewing and game play. *Journal of Adolescents, 27*, 97–111.

Webb, G. R., Redman, S., Gibberd, R. W., & Sanson-Fisher, R. W. (1991). The reliability and stability of a quantity-frequency method and a diary method of measuring alcohol consumption. *Drug and Alcohol Dependence, 27*, 223–231.

Webster, J. G., & Wakshlag, J. (1985). Measuring exposure to television. In D. Zillman & J. Bryant (Eds.), *Selective exposure to communication* (pp. 35–62). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Wiley, D. E., & Wiley, J. A. (1970). The estimation of measurement error in panel data. *American Sociological Review, 35*, 112–117.

Zaller, J. R. (1996). The myth of massive media impact revived: New support for a discredited idea. In D. C. Mutz, P. M. Sniderman, & R. M. Brody (Eds.), *Political persuasion and attitude change* (pp. 17–78). Chicago: Chicago University Press.