Published in final edited form as:

J Adolesc Health. 2009 March ; 44(3): 214–221. doi:10.1016/j.jadohealth.2008.09.002.

Test-Retest Reliability of Self-Reported HIV/STD-related Measures Among African-American Adolescents in Four U.S. Cities

Peter A. Venable,
Syracuse University

Michael P. Carey,
Syracuse University

Jennifer L. Brown,
Syracuse University

Ralph J. DiClemente,
Emory University

Laura F. Salazar,
Emory University

Larry K. Brown,
Brown University

Daniel Romer,
University of Pennsylvania

Robert F. Valois,
University of South Carolina

Michael Hennessy, and
University of Pennsylvania

Bonita F. Stanton
Wayne State University

Abstract

Purpose—To evaluate the test-retest reliability of self-reported sexuality-related data in a sample of African-American adolescents residing in four U. S. cities.

Methods—Using audio computer-assisted self-interviewing (ACASI), 156 African-American teens (M age = 15.5 years) provided data on lifetime and recent sexual behavior, HIV/STD testing, and theoretical antecedents of sexual risk behavior on two occasions separated by two weeks.

Results—Most self-reports of lifetime sexual behavior and STD/HIV testing were stable across the two assessment points. Test-retest agreement was substantial for dichotomous indices of lifetime sexual behaviors and STD testing (kappas ranging from .61 – .87), and for 3 month recall of vaginal sex (kappa = .72) and number of sexual partners (ICC = .68). Lower reliability estimates emerged

Corresponding Author: Peter A. Venable, Ph.D. Center for Health and Behavior, Syracuse University, 430 Huntington Hall, Syracuse, New York 13244, Email: pvanable@syr.edu, Phone: (315) 443-1210.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

for count data of unprotected vaginal sex occasions ($ICC = .44$). Test-retest reliability estimates for antecedents of sexual risk behavior were highest for a validated measure of HIV-related knowledge ($r = .73$), but somewhat lower for peer norms ($r = .58$) and condom use self-efficacy ($r = .50$).

Conclusions—Although variability in the stability of self-report data was observed, findings confirm that most sexual behavior, STD and HIV testing history, and psychosocial measures can be assessed reliably among adolescents. Research should continue to identify strategies to enhance the reliability of self-report sexual behavior data from youth at risk for HIV and other STIs.

Research on the prevention of HIV and other sexually transmitted infections (STIs) relies on self-report to document private attitudes and behaviors. Gathering accurate self-report data is challenging due to (a) the cognitive demands of recalling past behaviors and (b) motivational biases that can lead people to mis-report their behavior [1–3]. In addition to providing data on past sexual behavior, self-report is used to assess other behaviors of relevance to sexual health (e.g., STI testing) and to document changes in theoretically-hypothesized antecedents of sexual behavior. Given the central role of self-report data for applied research on sexual behavior, investigation of the reliability and validity of sexual health measures is vital.

Establishment of measurement reliability is a prerequisite to establishing its validity. Nonetheless, research to evaluate whether sexual behavior measures elicit reliable self-reports is rare. A review of studies conducted since 1990 identified only 15 test-retest studies involving recall of sexual risk behavior [1]. Although the review provides evidence to indicate that sexual behavior can be reported consistently [4–7], wide variability in reliability estimates was observed. Moreover, this review focused on adult samples that typically included only a narrow range of outcome measures. A major conclusion from the review was the need for more methodologically rigorous test-retest studies to establish measurement reliability in populations at greatest risk for HIV.

Adolescents may be particularly prone to inconsistent reporting of health behaviors [8]. Indeed, a minority of youth acknowledge dishonesty in survey research [9] and is not uncommon for teens who initially report being sexually active to later rescind their reports [10,11]. On the other hand, because sexual activity is often episodic among adolescents [12], sexual activity may be reported with a high degree of consistency because sexual occasions are highly salient [13].

Several studies have provided data on self-report consistency among adolescents. In one study, teens completed the CDC-Youth Risk Behavior Survey twice separated by a two-week interval; in this sample, kappa reliability estimates ranged from .40 to .90 for four dichotomous items assessing lifetime sexual experiences [14]. In a second study, African-American and Latina girls (aged 12 to 14) were interviewed; reliability estimates measured by kappa ranged from .34 to 1.0 for reports of (dichotomous) sexual milestones, with the lowest reliabilities observed for items assessing kissing or hugging and “spending time alone with a boyfriend” [15]. A third study also reported high test-retest reliability for a single item assessing lifetime occurrence of “ever having sex” ($kappa = .88$) [16]. For studies that have included continuous data regarding sexual behavior, findings have been mixed. In a sample of adolescent girls, exact agreement for frequency counts of sexual intercourse occasions was only fair ($kappa = .48$) [17]. In contrast, Schrimshaw and colleagues [18] interviewed 156 older gay, lesbian, and bisexual teens and reported strong test-retest reliability results for both dichotomous and continuous measures.

Although prior research provides modest evidence of test-retest reliability for self-report of sexual behavior among adolescents, three caveats are warranted. First, most studies have focused on a narrow range of behaviors. Second, although social-cognitive determinants of sexual behavior are often used as outcomes in intervention trials [19], evidence of their

reliability is lacking (for exceptions see [4,20]). Third, only one study [16] provides data on the reliability of self-report data collected using audio-computer assisted self-interviews (ACASI). Establishing the reliability of self-reported sexual behavior assessed with ACASI is important because assessment mode has been shown to have important influences on how people report on sensitive behaviors [1,2]. Reliability studies of sexual behavior in adolescents have relied primarily on interviewer administration of questionnaires, a format that reduces literacy and item comprehension concerns and yields higher reliability estimates compared to self-administered questionnaires [21–23]. However, interviewer administered surveys also elicit lower self-reports of sensitive behaviors compared to self-administered formats, presumably because the presence of an interviewer prompts greater self-presentation concerns [24]. ACASI combines elements of both interviewer and self-administered survey modes [25]. By maximizing privacy while still providing helpful audio prompts to overcome low literacy skills, ACASI may provide an optimal approach to assessing private sexual behaviors and beliefs. An important, and as yet unanswered, question concerns the degree to which ACASI assessments yield reliable test-retest estimates for sexual health measures.

To address these limitations, the present study reports on the reliability of sexuality-related data in a sample of adolescents residing in four mid-sized U. S. cities. Our study was conducted in preparation for an intervention trial investigating the impact of a mass media campaign on sexual risk behavior among African-American adolescents [26], a population identified as disproportionately affected by HIV and other STIs [27,28]. We report reliability data on continuous and dichotomous sexual risk behavior indices, STI and HIV testing, and measures of HIV and STI knowledge, condom use self-efficacy, and perceptions of peer norms for sexual activity. We predicted that ACASI administered questionnaires would yield moderate to high levels of test-retest reliability across measures of sexual behavior, HIV and STI testing, and the social cognitive determinants of sexual risk behavior.

Method

Participants

A total of 187 participants completed the Time 1 (T1) assessment. Of these, 156 (83%) completed the identical Time 2 (T2) assessment. African-American adolescents were recruited from two northern cities (Syracuse, NY and Providence, RI) and two southern cities (Macon, GA and Columbia, SC). Among participants who completed both questionnaires, 26% ($n = 41$) were from Syracuse, 23% ($n = 36$) were from Providence, 25% ($n = 39$) were from Macon, and 26% ($n = 40$) were from Columbia. There were more female ($n = 89$) than male respondents ($n = 67$); participants ranged in age from 14 to 17 years ($M = 15.5$, $SD = 1.4$). Participants' self-identified racial background was 93% African-American, 3% multiracial, 1% American Indian, and 3% "other." In addition, 6% of the participants were Hispanic. The majority (94%) reported that they were living in their family's home or apartment, most often with their mother (85%) and less frequently with their father (21%). Most participants (71%) qualified for a free or reduced price lunch at school; 75% reported that they were in high school.

Procedures

Youth 14 to 17 years old were recruited during after-school programming at partnering community-based organizations (CBOs). Identical ACASIs were administered in a group setting on two occasions, separated by a two-week interval. All participants and their parents or guardians provided written, informed assent and consent. The study protocol was approved by the Institutional Review Boards of the participating universities.

Measures

Sexual behavior—Using items from previous research [29–31], we assessed sexual activity status for all participants. To assess lifetime oral sex, all participants responded to two dichotomous items: (a) whether they had ever “put your mouth on a partner’s private parts (some people call this ‘giving oral sex’)” and (b) whether “a partner ever put their mouth on your private parts (some people call this ‘receiving oral sex’)?” Responses to these two items were combined to form a single indicator of lifetime oral sex. To assess lifetime vaginal sexual intercourse, all participants were asked whether they had “...ever had vaginal sex (that is, when a boy put his penis in a girl’s vagina).” Thus, all participants provided data for these two sexual behavior indicators.

Among sexually active participants, eight additional items were asked: age at which they first had vaginal sex (one item), the number of vaginal sex partners in their lifetime and in the past three months (two items), the total number of vaginal sexual encounters as well as the number of unprotected vaginal sex occasions in the past 3 months (two items), and whether a condom was used during their most recent episode of vaginal sex (one item). Respondents were also asked to indicate whether they consumed alcohol or drugs during the most recent sexual intercourse occasion (two items).

Pregnancy and STI testing history—Three indicators of lifetime consequences of sexual behavior were assessed. Pregnancy experience was assessed by asking participants to indicate the number of times they or their partner had been pregnant. Responses were dichotomously coded to indicate whether participants reported ever experiencing a pregnancy or a partner pregnancy. Two additional items were used to assess lifetime history of HIV and STI testing (e.g., “Have you ever been tested for HIV?”).

HIV knowledge—Knowledge about HIV transmission and prevention was assessed using an 18-item version of the HIV Knowledge Questionnaire (HIV-KQ) [32]. For each item, participants indicated whether they believed the statement was “mostly true” or “mostly false”; if they were unsure, they were instructed to respond with “don’t know” rather than guessing. Each correct item added one point to the total score. This measure has been evaluated with a variety of populations, but not adolescents.

STI knowledge—Knowledge of STI transmission and treatment was measured with a 10-item scale validated with college students [20]. For each question, participants indicated whether the statement was “mostly true” or “mostly false.” If a participant did not know, s/he was instructed to respond with “don’t know” rather than guessing. Each correct item added one point to the total score.

Condom use self-efficacy—Five items measured participants’ self-efficacy to use condoms in different situations (e.g., when using another form of birth control; [33]). For each situation, participants indicated their self-efficacy to use a condom using a four-point Likert scale (i.e., “not at all sure”, “somewhat sure”, “sure”, “extremely sure”). If participants had never had sex before, they were asked to respond by imagining that they were in the hypothetical scenarios. Internal consistency for the scale was high (T1 data, $\alpha = .86$).

Sex refusal self-efficacy—Seven items measured the extent to which participants could refuse vaginal sex under various conditions (e.g., if your partner refuses to use a condom [34]). For each item, participants rated their confidence level to refuse sex during the next month using a four-point Likert scale (i.e., “definitely could not say no”, “probably could not say no”, “probably could say no”, “definitely could say no”). Internal consistency was high ($\alpha = .86$).

Peer norms—Three items assessed perceived peer norms for engaging in vaginal sex [35]. Using a six-point Likert scale (“none” to “almost all of them”), participants reported the number of close friends, male peers, and female peers who they believed had vaginal sexual intercourse. The scale demonstrated excellent internal consistency ($\alpha = .86$).

Data Analyses

Reliability coefficients can be strongly influenced by even a single outlier [1,4]. To address this concern, we took four steps: (a) scatter plots were examined to identify extreme outliers; (b) difference scores between the two self-reports were computed; (c) mean difference score and standard deviation (SD) were calculated for each variable; and (d) cases with difference scores > 3 SDs from the mean were deleted.

Reliability was calculated in three ways [36,37]: (a) for the dichotomous behavioral variables (e.g., history of HIV testing), we used Cohen’s kappa statistic; (b) for the behavioral count data (e.g., number of sexual partners), we used the intraclass correlation coefficient (ICC); and (c) for the theoretical antecedents (e.g., knowledge), we used Pearson’s correlation coefficient. Cohen’s kappa and ICCs indicate absolute agreement between two data points and are appropriate for evaluating the stability of self-reports of data from overlapping time periods [1]. We interpreted the strength of test-retest agreement for kappa and ICC estimates using these established benchmarks (a) below .40 is poor; (b) .40 to .59 indicates a fair to moderate degree of agreement; (c) .60 to .79 indicates substantial agreement; and (d) .80–.99 indicates nearly perfect agreement [cf. 38,39].

Results

Dichotomous Indices of Sexual Behavior and STI/HIV Testing

As shown in Table 1, moderate to high levels of consistency between the two assessments were observed for dichotomous indices of lifetime sexual behaviors (kappa ranged from .67 to .85). Kappa coefficients for self-reports of lifetime engagement in any vaginal sex and for reports of engaging in vaginal sex in the last three months were .67 and .72, respectively. Self-reported occasions of vaginal intercourse without a condom (past 3 months) were in the “moderate” range (kappa = .47). Among sexually active teens reporting on their most recent occasion of sexual activity, there was substantial agreement for self-reported condom use between the two assessments (kappa = .62) and for a single item assessing drug use prior to the most recent occasion of sex (kappa = .64). Exact agreement for reports of alcohol use prior to the most recent sexual occasion was poor (kappa = .34).

High levels of consistency were observed for self-reports of STI testing (kappa = .79) and HIV testing (kappa = .77). STI test results were reported with a high degree of consistency (kappa = .85). Lifetime history of pregnancy was moderately reliable (kappa = .61).

Continuous Outcome Measures of Sexual Behavior and Age of Sexual Debut

Prior to finalizing reliability estimates for the continuous outcome measures, bivariate distributions of T1 and T2 responses were examined to explore the influence of outliers resulting from extremely discrepant responses. Two outlier cases emerged for an item assessing counts of number of sexual partners; a single outlier emerged for each of three items assessing lifetime number of sexual intercourse partners, age of sexual debut, and the number of occasions of unprotected vaginal sex.

Reliability estimates for the four sexual behaviors assessed using continuous data are shown in Table 1. Bivariate outliers strongly influenced the reliability estimates based on ICCs. As an illustration, Figure 1 shows the scatter plot of number of lifetime sexual partners reported

at T1 and T2. A single male participant indicated that he had 9 lifetime sexual partners at T1 but, at T2, he indicated that he had 110 lifetime partners. It is unknown whether this discrepancy reflects intentional responding, or was the results of an inadvertent data entry error (e.g., perhaps he had intended to key in either 10 or 11 – both plausible values given his earlier entry of 9 – but accidentally hit one extra keystroke making the value 110, a highly implausible value). With the single outlier case included in the reliability estimate, the ICC was .32, indicating only low to moderate reliability. When the single outlier was removed, the resulting ICC was .81, indicating substantial test-retest reliability. Similarly, a marked improvement was noted for the reliability estimate for number of partners in the previous 3 months. With outliers included, the ICC was .34; however, when the two extreme outliers were removed, the ICC improved to .68, indicating substantial agreement between T1 and T2. Reliability for the age of sexual debut item equaled .79 with one outlier included (ICC = .79); however, with the outlier removed, the ICC improved to .87.

Although the reliability estimate for the count of unprotected sexual encounters during the past three months improved with an outlier removed, reliability estimates were in the poor to fair range (ICC = .44). With the outlier included in the analysis, the ICC was .22.

Hypothesized Antecedents of Sexual Risk Behavior

Test-retest reliability for the five theoretical antecedents of sexual risk behavior was calculated using Pearson's correlation coefficient (see Table 2). Across these measures, there was evidence of test-retest reliability. Reliability estimates were highest for HIV Knowledge ($r = .73$) and STI Knowledge ($r = .62$). For the self-efficacy measures, self-reports of condom use self-efficacy were reliable between assessments ($r = .50$); however, self-efficacy to refuse sexual intercourse was low ($r = .27$). In addition, the Pearson correlation for a measure of peer norms for sexual behavior from T1 to T2 was .58.

Discussion

In this study, we evaluated the test-retest reliability of lifetime and recent sexual behavior data, self-reported HIV and STD testing history, and self-report measures of theoretical antecedents of HIV risk behavior in a sample of African-American teens. Based on established benchmarks [38,39], 85% of the indices reported on in this study showed moderate to excellent levels of agreement between T1 and T2. Similar to earlier studies [14,15], adolescents in our sample were consistent in reporting the number of sexual partners they had in their lifetime and in the past 3 months, as well their age of sexual debut. Moreover, dichotomous indices of oral and vaginal sex showed moderate to high levels of reliability. For condom use, reliability estimates were highest for the last occasion of vaginal sex and somewhat lower for 3 month measures of any unprotected sex and number of unprotected occasions of vaginal sex. Participants were moderately consistent in their self-reported drug use prior to sexual activity. However test-retest reliability was poor for self-reported alcohol use prior to sex.

Participants were highly consistent in reporting on past STI and HIV testing. These findings are encouraging, particularly given the importance of testing as part of a broader strategy of containing the spread of STIs and identifying persons who are infected so that they may receive treatment and prevention services [40]. Test-retest reliability estimates for antecedents of sexual risk behavior were highest for measures of HIV and STI-related knowledge. Participants were also moderately consistent in responding to items assessing peer norms and condom use self-efficacy, whereas test-retest reliability for a measure of sex refusal self-efficacy was only fair.

As others have observed [8], findings point to some variability in reliability across self-report domains. For example, adolescents were only moderately consistent in their reporting of the

frequency of unprotected sexual encounters and were less consistent in their reports of alcohol use prior to sex. We speculate that these lower reliability estimates for counts of unprotected vaginal sex occasions and (especially) for reports of alcohol use prior to sex are due in part to having two weeks of non-overlapping time between the test and retest sessions. In addition, items assessing the frequency of unprotected sex may have yielded lower reliability estimates because of differences among participants in what they consider to be sex “without a condom.” Recent findings suggest that some teens use a condom for only a portion of an intercourse occasion, whereas others use a condom from start to finish [41]. To improve measurement reliability, measures of unprotected sex may benefit from specifying that “protected sex” refers to using a condom throughout the entire duration of a sexual occasion.

The observed variability in reliability estimates across measures reinforces the importance of including multiple measures to assess related behaviors and constructs, because there is no “gold standard” for assessing sensitive sexual attitudes and behaviors. The kappa and ICC coefficients reported here evaluate the degree of absolute agreement in responses to past sexual behavior that, optimally, should be assessed for the same recall period. In this context, reliability data reported here provide a conservative estimate of reliability because the recall periods did not overlap exactly. Adolescent sexual behavior and condom use tend to be episodic [12]. As such, counts of unprotected sexual occasions and reports of alcohol use prior to sex may have varied for some youth based on actual differences in behavior that arose during the retest interval. This highlights the methodological challenge of choosing an optimal period of time between assessments for test-retest studies. Findings also highlight that a single outlier produced by discrepant self-reports can dramatically lower reliability estimates. One participant who reports a small number of sexual encounters at T1 but later reports having many more encounters can reduce reliability estimates to unacceptably low levels. However, as others have argued [1], outliers such as these should not be considered when computing reliability estimates, as they likely represent key stroke errors or blatant (infrequent) non-compliance with a study protocol. In the context of longitudinal prevention trials, it will be less straightforward to identify instances of non-compliance because differences between assessments could represent real change rather than aberrant reporting. Nonetheless, findings point to the importance of carefully considering the veracity of responses that deviate markedly from one assessment occasion to the next.

Most studies of test-retest reliability among adolescents have relied on face-to-face interviews rather than self-administered questionnaires. The present study is among the first to provide test-retest data for adolescents using ACASI, an administration mode that is growing in popularity because it allows people with low literacy skills to participate, encourages greater candor in disclosure of sensitive information [25,42], and allows for efficient data collection and “entry.” Findings from this study offer initial evidence of the reliability of data collected with ACASI. Future research can advance knowledge by comparing self-administered questionnaires, face-to-face interviews, and ACASI to determine whether one mode of administration optimizes reliability.

Our sample consisted of African-American adolescents recruited from neighborhoods with high rates of poverty, under-resourced public schools, and poor high school graduation rates. Although not formally assessed as part of study participation, literacy concerns were presumably common to a significant subset of study participants and may have influenced the reliability of responses. The use of simple audio instructions and item prompts, along with visual presentation of each survey item on laptop computers reduces concern about item comprehension due to low literacy levels. Nonetheless, cognitive limitations among lower literacy participants may have contributed to inconsistent responding due to difficulties with item comprehension [2,43] or concentration. Indeed, even with carefully worded items, sexual behavior questions can be interpreted in ways that differ from researcher’s intent [10,44,45]

and such differences are difficult to detect with ACASI administered surveys. Our use of detailed descriptions of what was meant by terms such as “vaginal sex” (i.e., “when a boy put his penis in a girl’s vagina) presumably helped to provide clarity. Nonetheless, such descriptions remain somewhat technical. It is possible that adolescents would respond more consistently to sexual behavior items if surveys also employed culturally normative “street” terms to describe specific sexual behaviors. Whether the use of slang terms affect assessment reliability as well as evaluator/researcher credibility and trustworthiness warrant research attention. Ultimately, choices with regard to optimal approaches to item construction and mode of assessment involve trade-offs that must be carefully weighed in terms of their impact on reliability, participant candor, and efficiency of data collection.

The ecological validity of our findings is enhanced because (a) study participants were recruited from four U. S. cities with elevated STI prevalence rates and (b) data collection took place in the context of busy community-based settings. Although reliability estimates may be optimized by working individually with study participants in a controlled laboratory setting, our study methods more closely mirror those that are employed in real-world, community-based research and practice.

Four limitations of the study should be acknowledged. First, sample sizes were reduced for several reliability estimates. Estimates obtained with fewer participants should be interpreted with caution. Second, the size of our sample did not allow for an exploration of potential differences in reliability estimates among subgroups of participants. An important direction for future research would be to explore whether response consistency is influenced by participant characteristics such as gender, age, and literacy levels. Third, use of a two-week recall period may have suppressed reliability estimates for some items because of a modest amount of non-overlapping time for the recall periods. Fourth, we did not collect data that allow us to address the validity of self-reports. Clearly, there is no gold standard for assessing the validity of sexual behavior self-reports; however, studies that employ daily diary data offer a reasonable approach to assessing the validity of self-reports [46]. Test-retest studies offer one form of information concerning the veracity of self-report data that can be augmented by research that uses data from contemporaneous methods (e.g., diaries).

Apart from the use of biological endpoints such as STI testing, there are few alternatives to the use of self-report as a means of assessing sexual behavior and related constructs. (Although important, biological endpoints are not a “gold standard” either, because they reflect both dyadic risk behavior and partner disease status.) Our findings affirm that sexual behavior can be assessed with a moderate to high degree of reliability among African-American adolescents. The fact that reliability estimates vary among measures in the present study and other recent reports affirms the need for continued research on strategies that can be used to optimize the reliability of self-reports in sexual behavior assessments.

References

1. Schroder KE, Carey MP, Vanable PA. Methodological Challenges in Research on Sexual Risk Behavior: II. Accuracy of Self-Reports. *Annals of Behavioral Medicine* 2003;26:104–23. [PubMed: 14534028]
2. Catania JA, Gibson DR, Chitwood DD, Coates TJ. Methodological problems in AIDS behavioral research: influences on measurement error and participation bias in studies of sexual behavior. *Psychological Bulletin* 1990;108:339–62. [PubMed: 2270232]
3. Weinhardt LS, Forsyth AD, Carey MP, Jaworski BC, Durant LE. Reliability and validity of self-report measures of HIV-related sexual behavior: Progress since 1990 and recommendations for research and practice. *Archives of Sexual Behavior* 1998;27:155–80. [PubMed: 9562899]

4. Carey MP, Carey KB, Maisto SA, Gordon CM, Weinhardt LS. Assessing sexual risk behaviour with the Timeline Followback (TLFB) approach: continued development and psychometric evaluation with psychiatric outpatients. *Int J STD AIDS* 2001;12:365–75. [PubMed: 11368817]
5. Stone VE, Catania JA, Binson D. Measuring change in sexual behavior: Concordance between survey measures. *Journal of Sex Research* 1999;36:102–8.
6. Needle R, Fisher DG, Weatherby N, Chitwood D, Brown B, Cesari H, et al. Reliability of self-reported HIV risk behaviors of drug users. *Psychology of Addictive Behaviors* 1995;9:242–50.
7. McElrath K, Chitwood DD, Griffin DK, Comerford M. The consistency of self-reported HIV risk behavior among injection drug users. *American Journal of Public Health* 1994;84:1965–70. [PubMed: 7998638]
8. Brener ND, Billy JO, Grady WR. Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: evidence from the scientific literature. *J Adolesc Health* 2003;33:436–57. [PubMed: 14642706]
9. Siegel DM, Aten MJ, Roghmann KJ. Self-reported honesty among middle and high school students responding to a sexual behavior questionnaire. *J Adolesc Health* 1998;23:20–8. [PubMed: 9648019]
10. Palen LA, Smith EA, Caldwell LL, Flisher AJ, Wegner L, Vergnani T. Inconsistent reports of sexual intercourse among South African high school students. *J Adolesc Health* 2008;42:221–7. [PubMed: 18295129]
11. Upchurch DM, Lillard LA, Aneshensel CS, Fang Li N. Inconsistencies in reporting the occurrence and timing of first intercourse among adolescents. *J Sex Res* 2002 Aug;39:197–206. [PubMed: 12476267]
12. Leigh BC, Morrison DM, Trocki K, Temple MT. Sexual behavior of American adolescents: results from a U.S. national survey. *J Adolesc Health* 1994;15:117–25. [PubMed: 8018684]
13. Jaccard J, Wan CK. A paradigm for studying the accuracy of self-reports of risk behavior relevant to AIDS: Empirical perspectives on stability, recall bias, and transitory influences. *Journal of Applied Social Psychology* 1995;25:1831–58.
14. Brener ND, Kann L, McManus T, Kinchen SA, Sundberg EC, Ross JG. Reliability of the 1999 youth risk behavior survey questionnaire. *J Adolesc Health* 2002 Oct;31:336–42. [PubMed: 12359379]
15. Hearn KD, O’Sullivan LF, Dudley CD. Assessing reliability of early adolescent girls’ reports of romantic and sexual behavior. *Archives of Sexual Behavior* 2003 Dec;32:513–21.
16. Romer D, Hornik R, Stanton B, Black M, Li X, Ricardo I, et al. “Talking” computers: A reliable and private method to conduct interviews on sensitive topics with children. *The Journal of Sex Research* 1997;34:3–9.
17. Sieving R, Hellerstedt W, McNeely C, Fee R, Snyder J, Resnick M. Reliability of self-reported contraceptive use and sexual behaviors among adolescent girls. *J Sex Res* 2005 May;42:159–66. [PubMed: 16123846]
18. Schrimshaw EW, Rosario M, Meyer-Bahlburg HF, Scharf-Matlick AA. Test-retest reliability of self-reported sexual behavior, sexual orientation, and psychosexual milestones among gay, lesbian, and bisexual youths. *Archives of Sexual Behavior* 2006 Apr;35:225–34.
19. Ajzen, I.; Albarracín, D. Predicting and changing behavior: a reasoned action approach. In: Ajzen, I.; Albarracín, D.; Hornik, R., editors. *Prediction and Change of Health Behavior*. Mahwah: Lawrence Erlbaum; 2007. p. 1–22.
20. Jaworski BC, Carey MP. Development and psychometric evaluation of a self-administered questionnaire to measure knowledge of sexually transmitted diseases. *AIDS and behavior* 2007 Jul; 11:557–74. [PubMed: 17016760]
21. Williams ML, Freeman RC, Bowen AM, Zhao Z, Elwood WN, Gordon C, et al. A comparison of the reliability of self-reported drug use and sexual behaviors using computer-assisted versus face-to-face interviewing. *AIDS Education and Prevention* 2000;12:199–213. [PubMed: 10926124]
22. Dowling-Guyer S, Johnson ME, Fisher DG, Needle R, Watters J, Andersen M, et al. Reliability of drug users’ self-reported HIV risk behavior and validity of self-reported drug use. *Assessment* 1994;1:383–92.
23. McLaws M-L, Oldenburg B, Ross MW, Cooper DA. Sexual behavior in AIDS-related research: Reliability and validity of recall and diary measures. *Journal of Sex Research* 1990;27:265–81.

24. Des Jarlais DC, Paone D, Milliken J, Turner CF, Miller H, Gribble J, et al. Audio-computer interviewing to measure risk behaviour for HIV among injecting drug users: A quasi-randomised trial. *Lancet* 1999;353:1657–61. [PubMed: 10335785]
25. Brown JL, Vanable PA, Eriksen MD. Computer-assisted self-interviews: a cost effectiveness analysis. *Behavior research methods* 2008 Feb;40:1–7. [PubMed: 18411521]
26. Vanable, PA.; Carey, MP.; Bostwick, RA.; Brown, LK.; Valois, RF.; DiClemente, RJ., et al. Community Partnerships in Adolescent HIV Prevention Research: The Experience of Project iMPPACS. In: Stanton, B.; Galbraith, J., editors. *The Uncharted Path From Clinic-Based to Community-Based Research*. Hauppauge; Nova: in press
27. CDC. Youth Risk Behavior Surveillance; United States, 2005. Centres for Disease Control and Prevention; 2006.
28. CDC. Racial/Ethnic Disparities in Diagnoses of HIV/AIDS --- 33 States, 2001–2004; 2006.
29. Vanable PA, McKirnan DJ, Buchbinder SP, Bartholow BN, Douglas JM Jr, Judson FN, et al. Alcohol use and high-risk sexual behavior among men who have sex with men: the effects of consumption level and partner type. *Health Psychology* 2004 Sep;23:525–32. [PubMed: 15367072]
30. Brown JL, Vanable PA. Alcohol use, partner type, and risky sexual behavior among college students: Findings from an event-level study. *Addictive behaviors* 2007 Dec;32:2940–52. [PubMed: 17611038]
31. Carey MP, Carey KB, Maisto SA, Gordon CM, Schroder KE, Vanable PA. Reducing HIV-risk behavior among adults receiving outpatient psychiatric treatment: results from a randomized controlled trial. *J Consult Clin Psychol* 2004 Apr;72:252–68. [PubMed: 15065959]
32. Carey MP, Schroder KE. Development and psychometric evaluation of the brief HIV Knowledge Questionnaire. *AIDS Education and Prevention* 2002;14:172–82. [PubMed: 12000234]
33. Galavotti C, Cabral RJ, Lansky A, Grimley DM, Riley GE, Prochaska JO. Validation of measures of condom and other contraceptive use among women at high risk for HIV infection and unintended pregnancy. *Health Psychol* 1995 Nov;14:570–8. [PubMed: 8565932]
34. Zimmerman RS, Sprecher S, Langer LM, Holloway CD. Adolescents' perceived ability to say "no" to unwanted sex. *Journal of adolescent research* 1995 Jul;10:383–99. [PubMed: 12290754]
35. Stanton B, Black M, Feigelman S, Ricardo I, Galbraith J, Li X, et al. Development of a culturally, theoretically and developmentally based survey instrument for assessing risk behaviors among African-American early adolescents living in urban low-income neighborhoods. *AIDS Educ Prev* 1995 Apr;7:160–77. [PubMed: 7619645]
36. Nunnally, JC.; Bernstein. *Psychometric theory*. Vol. 3. New York: McGraw-Hill; 1994.
37. Allen, MJ.; Yen, WM. *Introduction to measurement theory*. Long Grove, IL: 1979.
38. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977 Mar;33:159–74. [PubMed: 843571]
39. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment* 1994;6:284–90.
40. Janssen RS, Holtgrave DR, Valdiserri RO, Shepherd M, Gayle HD, De Cock KM. The Serostatus Approach to Fighting the HIV Epidemic: prevention strategies for infected individuals. *American journal of public health* 2001 Jul;91:1019–24. [PubMed: 11441723]
41. Sznitman SR, Romer D, Brown LK, DiClemente RJ, Valois RF, Vanable PA, et al. Prevalence, Correlates and STI Risk Related to Coitus Interruptus (Withdrawal) among African-American Adolescents. Manuscript submitted for publication
42. Schroder KE, Carey MP, Vanable PA. Methodological challenges in research on sexual risk behavior: I. Item content, scaling, and data analytical options. *Annals of Behavioral Medicine* 2003 Fall;26:76–103. [PubMed: 14534027]
43. Al-Tayyib AA, Rogers SM, Gribble JN, Villarroel M, Turner CF. Effect of low medical literacy on health survey measurements. *American journal of public health* 2002 Sep;92:1478–80. [PubMed: 12197979]
44. Sanders SA, Reinisch JM. Would you say you "had sex" if...? *Jama* 1999 Jan 20;281:275–7. [PubMed: 9918484]
45. O'Sullivan LF. Challenging assumptions regarding the validity of self-report measures: the special case of sexual behavior. *J Adolesc Health* 2008 Mar;42:207–8. [PubMed: 18295127]

46. Durant LE, Carey MP. Self-administered questionnaires versus face-to-face interviews in assessing sexual behavior in young women. Archives of sexual behavior 2000;29:309–22. [PubMed: 10948721]

Figure 1.
Scatter plot of Test and Retest Self-Reports for the Number of Sexual Partners

Table 1
Test-Retest Reliability Estimates for Self-Reported Sexual Behavior

Self-Report Variable	Type of Data [*]	N	Test-retest reliability index
<i>Lifetime</i>			
Oral Sex	dichotomous	156	.75
Vaginal Sex	dichotomous	156	.67
Tested for STIs	dichotomous	156	.79
Diagnosed with an STI	dichotomous	41	.85
Tested for HIV	dichotomous	156	.77
Pregnancy history	dichotomous	156	.61
Sex partners (number)	count	79	.81
Age of sexual debut (years)	count	79	.87
<i>Last Three Months</i>			
Vaginal sex	dichotomous	156	.72
Vaginal sex without a condom	dichotomous	156	.47
Unprotected vaginal sex (number)	count	100	.44
Vaginal sex partners (number)	count	78	.68
<i>Most Recent Vaginal Sex</i>			
Condom use	dichotomous	79	.62
Drug use prior to sex	dichotomous	79	.64
Alcohol use prior to sex	dichotomous	79	.34

* Note. Dichotomous variables evaluated using kappa; count variables evaluated using the intraclass correlation coefficient; N = number of participants providing data.

Table 2

Test-Retest Reliability Estimates for Theoretical Antecedents of Sexual Risk Behavior (N = 156)

Theoretical Antecedent	r
HIV Knowledge	.73
STI Knowledge	.62
Condom Use Self-Efficacy	.50
Sex Refusal Self-Efficacy	.27
Descriptive Norms for Sexual Behavior	.58

HIV = human immunodeficiency virus; STI = sexually transmitted infection.