Published in final edited form as:

AIDS Behav. 2008 March ; 12(2): 321. doi:10.1007/s10461-007-9272-1.

Validating an Index of Adolescent Sexual Behavior Using Psychosocial Theory and Social Trait Correlates

Michael Hennessy, Amy Bleakley, Martin Fishbein, and Amy Jordan

All with the Public Policy Center, Annenberg School for Communication, University of Pennsylvania, 3620 Walnut Street, Philadelphia, PA 19104

Abstract

Using a web-based survey of adolescents 14–16 years of age, a hierarchical index of heterosexual behavior was developed with excellent psychometric properties. The easiest sexual behavior to perform was “deep kissing” and the most difficult was “receiving anal sex” for females and “giving anal sex” for males. The index was validated with data that show increased sexual activity with being older and of minority status, with social traits such as physical development, having a romantic partner, and sensation seeking, and with psychosocial variables known to be associated with sexual behavior such as attitudes, norms, self-efficacy and intentions.

Keywords

adolescent sexual behavior; Mokken scaling; media effects; sensation seeking

INTRODUCTION

Sexual portrayals in the media have concerned child advocates, policymakers, and parents from the first instance of mass media marketed to children. The issue is one that has led to policy initiatives such as the V-chip and industry self-regulatory responses such as television and movie ratings (Bushman and Cantor, 2003) and video and music parental advisory notices (Hemphill, 2003). Thus, there is concern that what children and adolescents see, hear, and read in the media may have negative influences on their social-sexual development and behavior. Research is able to document sex in the media if not to clearly identify the effects of such exposure. We know that the amount of sex on television is increasing (Kunkel et al., 1996; Kunkel et al., 2001; Kunkel et al., 2003) and content analytic research suggests that there may be a basis for assuming that sex in the media might be shaping adolescents' beliefs, attitudes, norms and intentions to have sex (Ward and Friedman, 2006). Moreover, such exposure might be providing information that is misleading and inaccurate, and may lead to the formation of beliefs, attitudes, perceived norms, and perceptions of control or self-efficacy that put adolescents at risk for AIDS and other negative health outcomes (Ward, 2003). In this study we develop a multi-category behavioral index of adolescent sexual behavior that captures various types of sexual activity and is an appropriate outcome for investigating the relationship between sexual behavior and exposure to sexual content in the media.

The Annenberg Sex and Media Study (ASAMS) is a multi-year investigation of the relationship between sex in the media and self-reported sexual behavior in adolescents. It was designed to investigate whether sexual content in the media shapes adolescents' sexual behavior. In

ASAMS, the analytic variables used are guided by a particular social-cognitive theory – the *integrative model of behavioral prediction* (Fishbein, 2000) - a combination of the theories of reasoned action, planned behavior, the health belief model and social cognitive theory. Thus, ASAMS measures the psychosocial variables identified by theories of behavioral prediction and behavior change that are assumed to be most relevant for predicting and understanding behavior (e.g., attitudes, self-efficacy and measures of normative pressure).

In this paper, we first provide a rationale for the scaling of sexual behaviors. Then we discuss past efforts at scaling behaviors and present three methods of validating our index. After a discussion of the measures used, we present the scaling results, show the validity correlations, and use regression to display predicted values of the index score by the major demographic groups of respondents: race, gender, and age.

Rationale for Sex Behavior Difficulty Scaling

We desire a multi-category behavioral outcome because a focus on just a single aspect of sexual behavior such as vaginal sex limits analysis of the relationship between sex in the media and adolescent sexual behavior in important ways (Grello et al., 2006). Focusing solely on a single dimension of sexual behavior such as “vaginal sex” or high risk events like “unprotected anal sex” in an adolescent population presents problems when some or many of the individual behaviors are rare. In these cases, rare items reflect a data selection process that acts to limit the effective sample size of data sets. For example, behavioral interventions to reduce unprotected sex by adolescents often use counts of condom unprotected vaginal sex as an outcome measure (e.g., DiClemente et al., 2004; Kirby et al., 2004) or may even use sexually transmitted infection (STI) rates if the evaluation design is sufficiently longitudinal (Jemmott et al., 2005). While STI outcomes may be the putative gold standard for evaluating intervention effectiveness (but see Fishbein and Jarvis, 2000), respondents who report no vaginal sex during the retrospective reporting period cannot be used in the analysis, leading to inevitable selection biases (Winship and Mare, 1992). For a recent example of this bias in the adolescent and sexual behavior area, see Martino et al. (2005, page 917) and although simultaneous equation methods exist for the correction of such biases (Stolzenberg and Relles, 1997), we know of no example of these corrections in the adolescent HIV prevention research literature.

An additional reason for a multi-behavioral sex index appropriate for media studies is that “sex in the media” is more often implied or discussed than explicitly shown (Martino et al., 2005.) Thus, a focus on more explicit behaviors like vaginal or oral sex selects a behavioral outcome that is rarely modeled in the media, especially on television. For example, Kunkel and his colleagues have been tracking sexual portrayals on television for over a decade using both random sampling methods to select “average weekly” programming and targeted sampling of nationally popular shows based on Nielsen ratings. In their analysis, they differentiate between “talking about sex,” “precursory” sexual behaviors (e.g., flirting, passionate kissing, intimate touching) and sexual intercourse (either implied or depicted). In general, the former are far more common than the latter:

Nearly twice as many shows include talk about sex as contain sexual behavior. Similarly, the number of scenes per program with talk about sex is almost twice as high as the number of scenes with sexual behavior. Precursory behaviors, such as passionate kissing or intimate touching, are the most common form of sexual activity shown on television. Still, about one of every seven programs on television (excluding news, sports, and children’s programming) presents sexual intercourse either by depicting it directly or by portraying characters who are about to begin or have just finished having sex (Kunkel et al., 2003, page 16).

Finally, there is one intervention context where an omnibus sexual behavior outcome measure is obviously superior to single item dichotomous ones: the case of abstinence interventions (Bruckner and Bearman, 2005; Manlove et al., 2004). In these studies, omnibus measures of sexual behavior are required because the behavioral purpose of the intervention is to prevent vaginal sex and it is of research interest to determine which non-coital behaviors (if any) are behavioral substitutes.

In summary, an index that captures a range of sexual behaviors would provide a measure of sexual behaviors which are common in the media, avoid the measurement weaknesses inherent in mono-operational measures (Cook and Campbell, 1979), have higher construct validity than a single item approach (Bollen and Lennox, 1991), and eliminate selection bias. Moreover, it would allow us to detect changes in sexual behavior that would often be missed if we focused on a single behavioral dichotomy.

This type of scaling is traditionally referred to do “difficulty” ordering because of its origins in item response theory (IRT) and educational testing. As it pertains to sexual behavior in practical terms the word “difficulty” does not necessarily imply differences in behavioral skills or knowledge required to engage in the behaviors, but instead refers to the progression of behaviors from more common, pre-coital activity such as kissing to coital activity such as oral, vaginal, or anal sex. There is an increasing inherent difficulty in the progression of behaviors, whether it be because kissing is more common than anal sex, or that kissing is less physically and emotionally complicated. Thus, we use the term difficulty ordering to be consistent with how this technique is applied to other outcomes as well as how it is referred to as a generic scaling methodology.

Difficulty Ordering versus Correlational Definitions of Unidimensionality

Another reason for developing a multi-behavior sex scale that is also difficulty-ordered is because the construct validity of such a scale is higher than with a simple summative index constructed through an examination of intercorrelations of dichotomous items using the KR20 formula, a version of “alpha” for dichotomous items (Streiner, 2003). To demonstrate the advantages of a difficulty ordered scale, consider a ten item true/false HIV knowledge index (e.g., Carey and Schroder, 2002). Suppose a respondent scores a “4” on the index. With ten items there are **210** (e.g. $10!/((4!(10-4)!))$) different ways to get a “4” on the knowledge test. This situation raises two questions: are all respondents who answered 4 items correctly at the “same level of knowledge” and do we care if some respondents achieved the same total score by answering four different questions correctly? In other words, what does a knowledge level of “4” actually mean if there are 210 different ways to demonstrate this level of HIV knowledge? The brief answer is this: we do not care how any respondent achieved an identical sum *if all the items are of equal difficulty* because in this case, all “4”s are completely interchangeable and reflect the same level of knowledge. However, if the items are not completely interchangeable, then some “4”s reflect different knowledge levels, and this is clearly a problem of construct validity of the summative index.

Taking item difficulty into account increases construct validity in another important way. If the items are difficulty ordered, then the researcher knows precisely what a “4” means: it means that the respondent with a “4” answered the first four items correctly and the last six items incorrectly. That is the interpretive advantages of difficulty ordering: *the value of the summative scale score indicates which items were passed and which were failed*. Thus difficulty ordering is an alternative (and stricter) definition of unidimensionality, a definition that directly connects the summative score with particular items in a battery of questions and as such reflects a stronger form of unidimensionality than correlational association. Scaling sexual behaviors in this way provides researchers with an index reflecting a “sexual behavior hierarchy.”

Past Studies of Sex Behavior Difficulty Scaling

Difficulty-based scaling of dichotomous behavioral items has been commonly used in the drug and alcohol field due to researcher interest in identifying the “gateway sequences” of drug use (e.g., Ellickson et al., 1992; Kandel and Yamaguchi, 1993). There have also been some attempts to use the same logic with sexual behaviors because survey self-report data on individual sex behavior items often imply an underlying behavioral sequence although some studies investigate the ordering of sexual behaviors as a part of a ranking exercise based on perceived prevalence in the population (Derogatis, Melisaratos and Clark, 1976) or normative social appropriateness (Johnson, 1986).

But ordering empirical behaviors, not hypothetical comparative rankings, is the research issue here. For example, Bentler (1968a, 1968b) reported results on 21 behavioral self-reports from separate samples of male and female college students. The easiest behavior was “one minute continuous lip kissing” and the most difficult was “mutual oral manipulation of genitals to mutual orgasm” (the percents affirmatively reporting the behaviors were not provided). Both his 21 item list and a smaller 10 item subscale were internally consistent using the KR20 formula. For males, KR20 was .95 and .89 for the long and short version of the scale and .95 and .89 respectively for females. Both sets of scales also fit the hypothesis of difficulty-ordered unidimensionality using Loevinger’s *H* measure (for males, .76 and .81 respectively, for females, .79 and .93 respectively).

In another early study of college students of both genders, Zuckerman (1973) used 12 sexual experience dichotomies appropriate for each gender and showed that they could be scaled for males and females. The easiest behavior for males was “feeling the covered breast” (92% affirmative) and the most difficult was “coitus, enter the vagina from the rear” (29% affirmative) while the easiest for females was “covered breast felt” (85% affirmative) and the most difficult was “coitus, face to face, side” (19% affirmative). The rank order correlation of the 12 behaviors between genders was .95 (Zuckerman, 1973, page 28). No appropriate statistics of scalability were reported. More recently, Anderson and Cyranowski (1995) reported on the 24 sexual behavior items of the Derogatis Sexual Functioning Inventory (Derogatis and Melisaratos, 1979) from a sample of female college students and showed an ordering of “kissing on the lips” as the easiest behavior (98.8% of the respondents reported affirmatively) and “anal intercourse” as the most difficult (12.3% reported affirmatively). Unfortunately, no statistics of scalability were reported.

The same approach has been used with adolescents. Smith and Udry (1985) applied scaling logic to sexual behavior using a school-based sample of adolescents between 12 and 15 years of age. There were 11 non-coital items originally, but some of these did not scale and are not reported on in the original article. They used the following behaviors in their scale “necked,” “feel breasts clothed,” “feel breasts directly,” “feel sex organ directly,” “feel penis directly,” and “intercourse” and made cross gender and cross ethnicity comparisons. They found that within ethnicity behavior rankings were identical for males and females: for White adolescents, “necked” was the easiest behavior and “intercourse” the most difficult and for African-American adolescents, “necked” was the easiest and “feel penis directly” the most difficult. “Intercourse” fell at position 4 (between “None” at position 1 and “feel penis directly” at position 7) for African-Americans. Jakobsen (1997) reported on a nationally representative Norwegian sample of adolescents from 13–16 years of age. His focus was on the non-coital behaviors “going steady,” “kissing,” “French kissing,” “light petting,” and “heavy petting.” Except for 16 year old girls, “going steady” was always the easiest behavior and “heavy petting” the most difficult (Jakobsen, 1997: page 544). The scalability statistics were good: Loevinger’s *H* ranged from .77 to .85 for males over the age range, from .82 to .83 for females over the age range, and from .79 to .85 with both genders combined over the age range. Cronbach’s alpha

was between .76 to .87 for males, .72 to .86 for females, and from .74 to .87 for the genders combined over the age range (Jakobsen, 1997, page 548).

Sexual Behavior Index Validation

Self-report measures of sexual behavior are by their nature difficult to validate. Here we use three approaches for validating our measure. First, is it consistent with previous research on adolescents of the same age? Using this criterion, our measure should have a positive association with respondent age and show higher average values for African-American adolescents than White adolescents (Centers for Disease Control and Prevention, 2006; Mosher et al., 2005; Silver and Bauman, 2006). Gender differences are more difficult to predict because the scale is tailored – because of its behavioral specificity – to males and females separately, so we leave this issue unaddressed here, but do investigate gender differences in the analysis. Second, does the scale correlate with the theoretical determinants of behavior? At the very least, the sex index should correlate with intentions and other integrated model variables concerning vaginal sex.

Note that this kind of validation test is identical to that of Donovan and Jessor (1983, page 549) who validated their 6 item index of adolescent drug and alcohol use by correlating their measure with theoretical variables such as religiosity, injunctive norms favoring drinking, and role modeling of drinking. It is also the same procedure used by L'Engle, Jackson, and Brown who validate their “cognitive susceptibility” to initiating sexual intercourse by correlating their measure with peer behavior, parental closeness, and sensation seeking, among other variables (L'Engle et al., 2006).

Finally, another aspect of validation requires that the behavioral scale correlates with other psychological and social traits that are related to sexual behavior. Here we investigate the association between our index and a selection of these characteristics that, based on “strong and consistent evidence” (Kirby et al., 2005, page 28) are linked with adolescent sexual behavior: the respondent's age, self-reported physical development, ever having a romantic partner, religious attendance, and sensation-seeking. All of these traits should have non-zero associations with self-reported sexual behavior if the proposed index is valid.

METHODS

Study Design and Participants

Data collection took place via a Web-based survey fielded during the spring and summer of 2005. Adolescent respondents were recruited through print and radio advertisements, direct mail, and word of mouth to complete the survey. Respondent eligibility criteria included age at the time of the survey (14, 15, or 16) and race/ethnicity (White, African-American, or Hispanic). The sampling strategy was quota-driven with a desire for roughly equal sample sizes in all Race*Age*Gender cells (a 3*3*2 design). In practice, Hispanic respondents were extremely difficult to locate and recruit, so their cell frequencies are low. The survey was launched in April 2005 following a test of the technology and a pre-test of the survey instrument measures in February of the same year.

The survey was accessible from any computer with internet access. Participants were given the option of taking the survey at the University or an off-site location (e.g. home, school, or community library). The majority of the participants (84.9%) took the survey at home or another off-site location while 15.1% took the survey at the University. Respondents were assigned a password to access the survey, as well as an identification number and personal password to ensure confidentiality and privacy protection. Respondents were compensated \$25 dollars upon completion of the survey and on average, took one hour to complete the survey.

After submitting respondent assent/parental consent forms, 547 adolescents completed the survey. The survey results reported here reflect a subset of all respondents because 89 Hispanic or mixed race respondents are excluded due to insufficient sample sizes. The analysis data set includes 458 adolescents, ages 14 to 16, resulting in a respondent group that is 60% female and 51% African-American with 33% 14 year olds, 34% 15 year olds, and 33% 16 year olds.

Measures

Sexual Behaviors—The survey collected data for lifetime, more than a year ago, and during the past 12 months on the following *sexual behaviors*: deep kissing (item: *Have you ever participated in deep kissing (some people call this “French kissing”)?*), sexy dancing (item: *Have you ever done really sexy dancing (some people call this “bump and grind” or “dirty dancing”)?*), touching the breasts of a female partner (item: *If you ever have had a female partner, did you touch her breasts?*), the respondent having their breasts touched (item: *Have you ever had your breasts touched by a partner?*), genital touching of the respondent by a partner (item: *Has a partner ever touched your private parts?*), the respondent giving oral sex (item: *Have you ever put your mouth on a partner’s private parts (some people call this “oral sex”)?*), the respondent receiving oral sex (item: *Has a partner ever put their mouth on your private parts (some people call this “oral sex”)?*), the respondent receiving anal sex (item: *If you ever had a male partner, has he ever put his penis in your anus (some people call this “anal sex”)?*), the respondent giving anal sex (asked of males only, item: *Have you ever put your penis in your partner’s anus (some people call this “anal sex”)?*), and having vaginal sexual intercourse (item: *Have you ever had sexual intercourse (i.e., a penis in the vagina) with a partner of the opposite sex?*).

Because of the limited age range of the respondents, we focus on the lifetime items here because many of the behaviors are rare or zero using shorter recall periods. We limit the analyses to heterosexual behaviors, so the breasts touched and the receiving anal sex variables are used only with females, and the touching breasts and giving anal sex variables are used only with males. We also drop from the analysis six males who have received anal sex because their inclusion reduces the hierarchical nature of the index males. Because all of these male respondents did not report any lifetime occurrence of vaginal sex, our heterosexual behavioral index is inappropriate for them.

Measures of the Integrated Model Variables—The survey also collected measures of the precursors to behavior: direct measures of attitudes, two types of normative measures, a global measure of self-efficacy, and behavioral intentions. To validate the sex behavior index from a convergent validity point of view, we examine the correlation between the theoretical determinants of behavior and the self-reported sex score. In particular, the correlation between behavioral intentions and the sex score should be substantial. The items measuring *behavioral intentions* were: “I am willing to have sexual intercourse in the next 12 months”, “I will have sexual intercourse in the next 12 months”, and “I intend to have sexual intercourse in the next 12 months.” All items were coded from 1 to 7 with an “unlikely” to “likely” format (Cronbach’s $\alpha = .93$, $M = 3.15$, $SD = 2.25$).

The measures of *attitudes* were semantic-differential items coded from 1 to 7 prefixed by the stem: “My having sexual intercourse in the next 12 months would be...” The semantic extremes were: bad/good, foolish/wise, unpleasant/pleasant, not enjoyable/enjoyable, and harmful/beneficial (Cronbach’s $\alpha = .92$, $M = 3.99$, $SD = 1.89$).

The measure of *self-efficacy* was a global item scored from 1 (“certain I could not”) to 7 (“certain I could”) in response to: “If I really wanted to I am certain that I could/could not have sexual intercourse in the next 12 months” ($M = 4.67$, $SD = 2.35$).

The normative items were of two kinds: injunctive and descriptive. Injunctive norms measure what important referent others think the respondent should do and descriptive norms measure respondent perceptions of what others are doing (Cialdini et al., 1990). The measure of *injunctive norms* was a global item scaled from 1 to 7 where 1 represented the response “Should not have” and 7 the response “Should have” to: Most people who are important to me think that I should/should not have sexual intercourse in the next 12 months? ($M = 2.55$, $SD = 1.81$). The measures of *descriptive norms* were two ordinal items coded from 1 to 7. The first item were the responses “Will not” to “Will have” to the question: Most people like me will not/will have sexual intercourse in the next 12 months. The second item were the responses “Have not” and “Have had” to the question: Most people like me have not/have had sexual intercourse. Because these two items were highly correlated (polychoric correlation = .73, $p < .05$), they were combined into a single index. (Cronbach’s $\alpha = .85$, $M = 4.03$, $SD = 2.06$).

Measures of the Validating Traits—*Physical development* was measured using a five level ordinal variable. The stem was: Compared to other teens your age, how advanced is your physical development? The responses were coded as “I look younger than most,” “I look younger than some,” “I look about average,” “I look older than some” and “I look older than most.” ($M = 3.31$, $SD = 1.04$). Ever having a *romantic partner* was a dichotomous item in response to: During your lifetime, have you ever had a boyfriend or girlfriend? ($M = .85$, $SD = .36$). *Religiosity* was measured by a combination of two different variables, each coded in three levels. The first was asked about attending religious services (responses were “Never,” “Occasionally” or “Regularly”) and the second about the importance of religion in your life (responses were “Not important,” “Somewhat Important,” and “Very Important”). The seven cells with 98% of the cases were recoded into a 1–7 ordinal measure as follows: “Not Important/ Never Attend,” “Not Important/Occasionally attend,” “Somewhat important/Never attend,” “Somewhat important/Occasionally attend,” “Somewhat important/Regularly attend,” “Very important/Occasionally attend” and “Very important/Regularly attend” ($M = 4.50$, $SD = 1.96$). The *sensation seeking* scale was constructed from a four-item measure based on a reduced form of the Brief Sensation Seeking Scale, the BSSS (Hoyle et al., 2002) that captures the four dimensions of the original, 40-item Sensation Seeking Scale (SSS-V): Thrill and Adventure Seeking, Experience Seeking, Disinhibition, and Boredom Susceptibility (Zuckerman et al., 1978). In a study comparing the four-item scale with more established measures of sensation seeking, Stephenson and colleagues (Stephenson et al., 2003) demonstrated convergent validity as the four-item scale correlates highly ($r = 0.89$, $p < 0.001$) with the BSSS, which itself has been found to have high construct validity (Hoyle et al., 2002). Respondents were asked to rate their agreement on a five-point agree/disagree scale for each of the following statements: I would like to explore strange places; I like to do frightening things; I like new and exciting experiences, even if I have to break the rules; I prefer friends who are exciting and unpredictable (Cronbach’s $\alpha = .73$, $M = 3.37$, $SD = .84$).

Scaling the Lifetime Sexual Behavior Items—We used Mokken scaling to assess the scalability of the dichotomous sexual behavior items. Mokken scaling is a procedure for assessing dichotomous or ordinal response items for unidimensionality. Unidimensionality of the items is defined by the empirical ranking of the items along an unobserved “difficulty” dimension such that all items after the initial failure are also failed and all items before the initial failure are passed (Ringdal et al., 1999). In practice, the marginal proportions of positive responses to the items are used to estimate difficulty (Sijtsma and Verweij, 1992). If the items scale using this definition, then empirical index scores correspond to passing the number of difficulty-ranked items less than or equal to the observed score and failing all difficulty-ranked items greater than the value of the observed score. Item sets are evaluated for unidimensionality using Loevinger’s H coefficient. H for a proposed scale is defined as $1 - (E_{\text{scaleable}}/E_{\text{random}})$ where $E_{\text{scaleable}}$ is the number of observed response errors under the scalability assumptions

and E_{random} is the number of errors expected under statistical independence given the marginals of the proposed scale items (Ringdal et al., 2003, page 169). As Ringdal et al. (1999, page 27) summarize, "... H is interpreted as an index for the degree to which subjects can be accurately ordered by means of k items." Mokken (1971, page 185) recommends a minimum value of .30 for H (the value can be negative) while values between .3 and less than .4 are a weak scale, between .4 and less than .5 a medium scale, and .5 or more as a strong scale.

RESULTS

Table I shows the distribution of the sexual behavior dichotomies used in the ASAMS survey. The behavioral dichotomies are sorted by the prevalence in the sample as a whole, so the easiest behavior that is not gender specific is deep kissing and the most difficult non-gender specific behavior is giving oral sex. Gender comparisons show only one significant difference: sexy dancing, with boys being much more likely to report engaging in this activity than girls. The age-specific comparison show that the prevalence of most behavioral dichotomies increases with age due to the positive correlation between age and sexual activity, and the race differences show higher prevalence for all but one behavior (giving oral sex) for African-American respondents, which is consistent with survey results of representative samples of adolescents.

The scaling analysis for the sexual behavior items were performed in a partially *a priori* and partially exploratory way. Because of the gender specific nature of some of the items, all the scaling runs for females and males were estimated separately. The scaling program (the *MSP* and *LoevH* procedures in STATA) uses internal thresholds based on item specific Loevinger's H values to select scalable items, and thus this component of our approach was exploratory. For both genders, "dirty dancing" was never selected as a scalable item. Table II displays the final scale results by the prevalence of the behavioral dichotomy for each gender. It shows first that the gender common dichotomies (deep kissing, genital touching, receiving oral sex, vaginal sex, and giving oral sex) and gender specific dichotomies (e.g., touching breasts, breasts touched, receiving anal sex, giving anal sex) are rank ordered identically by both males and females and that the prevalences of the non-gender specific behaviors are similar. The fit of the scaling model is excellent: Loevinger's H is .74 or above for both genders, making these "strong scales." Note that our dichotomous sex behaviors also scale using a correlational definition of unidimensionality, the KR20 coefficients are .84 and .85 for male and female respondents respectively. Figure 1 plots the histogram of the cumulative score by gender. The distributions for males and females are similar, both with a right skewness and a modal value of zero. The average value for males is 2.80 (SD = 2.19) and for females it is 2.67 (SD = 2.22), not a statistically discernable difference in terms of central tendency ($t = .548$, $p > .05$) or variance ($F = 1.03$, $df = 259, 161$, $p > .05$). These statistics suggest that the average respondent has progressed up through "genital touching" in the sexual behavior hierarchy but no farther.

Table III shows the validity correlations first with the theoretical variables (intentions, attitudes, self-efficacy, and injunctive and descriptive norms) and then with the trait variables (age, physical development, lifetime romantic partners, religiosity, and sensation seeking). All the statistically significant correlations are in the correct direction, although the sensation seeking correlation is low. However, the relationship between sensation seeking and sexual behavior is complicated, and past reviews of the association between the two concepts shows a heterogeneous pattern of positive and zero associations depending on the context of the relationship and the specific sexual behavior under consideration (Henderson et al., 2005). The religiosity measure is also a bit surprising, but the bivariate associations between the original two variables that were used to construct the composite variable and the sex score were also very small: the polychoric correlation between frequency of attendance of religious services and the sex score was $-.024$ and between importance of religion and the sex score was $-.015$.

In all other cases, the results here suggest meaningfully high correlations between validating theoretical and trait variables.

The final validation task is to model the sexual behavior index simultaneously over the three important demographic features of the respondents to determine differences on the sex score by these variables. To combine these three predictor variables into a single graph, we first defined the appropriate dummy variables and their products to test the saturated model with three main effect, two two-way effects, and one three way interaction using regression. Then these results were compared with the main effect only model using the likelihood ratio test (Kennedy, 2003, pages 66–68). The saturated model was not a significant improvement over the main effect model ($F = 1.03$, $df = 7, 408$, $p > .05$). Thus we present the main effect predicted values from the regression here. Figure 2 shows that predicted sex scale score for each sub-classification of the respondent in the gender*age*race cross classification assuming no interaction. This shows an increase in predicted sex score by age, with African-American adolescents starting at and maintaining higher values than White respondents and within those classifications, males having slightly higher values on the score than females. However, even in the main effects only model, the age and race adjusted difference between the genders was not statistically different.

DISCUSSION

The correlations between the sex index score and the validating theoretical variables were meaningfully high, but not “too high.” Because the sex score is a composite of many behaviors, it should be moderately correlated with theoretical response items that focus on a specific behavior, which in the case of the integrated model variables here is vaginal sex. That is not to say that the correlation between the sex score and specific sexual behaviors (as opposed to theoretical constructs concerning a specific behavior) is low: the polychoric correlation between the sex score and the lifetime vaginal sex dichotomy is .92 ($SE = .014$, $p < .05$). The sex score was also correctly directionally correlated with all but one of the other validating trait variables and also produced average values that were consistent with epidemiological data on the sexual behavior of adolescents when broken down by gender, age, and race.

Of course, difficulty-based scaling of behavioral dichotomies is not appropriate for all research questions. It should be used when the sexual behaviors are relatively rare, as is the case for adolescents in our age range, and it works well when the purpose of the analysis is to differentiate between levels of on-set over a range of sexual behaviors. Thus, this type of measure is appropriate for measuring intervention outcomes for either abstinence-only or more comprehensive sex education/contraception behavioral interventions because it is precisely the prevention or delay of on-set of particular sexual behaviors that are relevant for these types of interventions. The desire to delay on-set of sexual behavior for adolescents is also consistent with federal funding policies relating to sex education in the United States (Bleakley, Hennessy, and Fishbein, 2006). If an intervention were successful at delaying on-set of vaginal sex, for example, the intervention group would ideally have no or minimal (upward) change on the scale. A decrease on the scale would not be valid if the items were lifetime behaviors (as is our scale).

But difficulty scaling is not limited to dichotomous items and can be appropriate for other intervention contexts. When the focus is on “unsafe sex,” clearly on-set measures are irrelevant and count or frequency measures of “unsafe sex” or “safe sex” events over a limited recall period for each type of sexual behavior would be superior (Fishbein et al., 2001; Greenberg et al., 2000). Fortunately these kinds of ordinal variables can also be evaluated using a definition of unidimensionality based on difficulty ordering (Hemker, Sijtsma, and Molenaar, 1995; Ringdal et al., 2003). Even more detailed types of self-reports are possible. Fortenberry and

his colleagues are now collecting longitudinal daily diary data on a relatively small sample of female adolescents to look at predictors of vaginal sex based on psycho-social variables such as the respondent's "mood" and level of sexual arousal, sexual partner variables, parental influence variables, and other contextual features like time of day (Fortenberry et al., 2005, 2006). While their diary method collects an impressive array of micro-data on the respondent (and up to 5 of her sexual partners), one striking result is that for most of the daily diary records, the adolescent respondents do not have sex. In one study 146 respondents reported on 28,376 diary days, but sexual activity was reported only for 8.3% of all the days (Fortenberry et al., 2005, page 254) and in another, 106 respondents reported on 12,168 days, but only for 12% of the diary days was sex reported (Fortenberry et al., 2006, page 278). Diaries may be the "gold standard" for some sexual behavior related research problems, but clearly not for all, especially when the sexual behavior of interest is infrequent.

The earlier work on scaling of sexual behavioral data focused on establishing a hierarchy of behaviors that were ordered in time and the identification of such hierarchies was the purpose of the enterprise – early studies did not emphasize the measurement properties of the scaling approach. The lack of current interest in identifying behavioral hierarchies in the adolescent sex area may be due to the sexual behaviors that were used in earlier research. These behaviors ranged from the bizarrely specific (e.g., types of sexual positions, minutes of lip kissing) to broad behavioral categories (e.g., "petting"). This approach leads to published findings that are difficult to compare in any meaningful way. However, both prospective interventions and observational studies need multi-behavioral unidimensional sexual behavior outcomes to track the effects of behavioral interventions as well as to estimate the temporal correlations between demographic, psychosocial theory-based predictors, and other potential determinants of sexual behavior such as media exposure. Due to the risks of STD and HIV infection in the adolescent population (Mosher et al., 2005; Institute of Medicine, 1997), there is a research need to accurately measure and model sexual behavior cross-sectionally and over time. With the advanced scaling methods now available, constructing meaningful hierarchical measures of sexual behavior is greatly facilitated.

Acknowledgments

This publication was made possible by Grant Number 5R01HD044136 from the National Institute of Child Health and Human Development (NICHD). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NICHD.

References

- Anderson B, Cyranowski J. Women's sexuality: behaviors, responses, and individual differences. *Journal of Consulting and Clinical Psychology* 1995;63:891–906. [PubMed: 8543712]
- Bentler P. Heterosexual behavior assessment – I males. *Behaviour Research and Therapy* 1968a;6:21–25. [PubMed: 5644763]
- Bentler P. Heterosexual behavior assessment – II females. *Behaviour Research and Therapy* 1968b;6:27–30. [PubMed: 5644764]
- Bleakley A, Hennessy M, Fishbein M. Public opinion on Sex Education in US Schools. *Archives of Pediatric and Adolescent Medicine* 2006;160:1151–1156.
- Bollen K, Lennox R. Conventional wisdom of measurement: a structural equation perspective. *Psychological Bulletin* 1991;110:305–314.
- Bruckner H, Bearman P. After the promise: the STD consequences of adolescent virginity pledges. *Journal of Adolescent Health* 2005;36:271–278. [PubMed: 15780782]
- Bushman B, Cantor J. Media ratings for violence and sex. *American Psychologist* 2003;58:130–141. [PubMed: 12747015]
- Carey M, Schroder K. Development and psychometric evaluation of the brief HIV knowledge questionnaire. *AIDS Education and Prevention* 2002;14:172–182. [PubMed: 12000234]

- Centers for Disease Control and Prevention. Youth Risk Behavior Surveillance- United States, 2005. Surveillance Summaries, June 9, 2006. MMWR, 2006 2006;55 (No. SS-5).
- Cialdini R, Reno R, Kallgren C. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology* 1990;58:1015–1026.
- Cook, T.; Campbell, D. Quasi-experimentation: design and analysis issues for field settings. Chicago, IL: Rand McNally; 1979.
- Derogatis L, Melisaratos N. The DSFI: a multidimensional measure of sexual functioning. *Journal of Sex and Marital Therapy* 1979;5:244–281. [PubMed: 513144]
- Derogatis L, Melisaratos N, Clark M. Gender and sexual experience as determinants in a sexual behavior hierarchy. *Journal of Sex and Marital Therapy* 1976;2:85–105. [PubMed: 978760]
- DiClemente R, Wingood G, Harrington K, Lang D, Davies S, Hook E, Oh K, Crosby R, Hertzberg V, Gordon A, Hardin J, Parker S, Robillard A. Efficacy of an HIV prevention intervention for African American adolescent girls. *Journal of the American Medical Association* 2004;292:171–179. [PubMed: 15249566]
- Donovan J, Jessor R. Problem drinking and the dimension of involvement with drugs: a Guttman scalogram analysis of adolescent drug use. *American Journal of Public Health* 1983;73:543–553. [PubMed: 6837819]
- Ellickson P, Hays R, Bell R. Stepping through the drug use sequence: longitudinal scalogram analysis of initiation and regular use. *Journal of Abnormal Psychology* 1992;101:441–451. [PubMed: 1500601]
- Fishbein M. The role of theory in HIV prevention. *AIDS Care* 2000;12:273–278. [PubMed: 10928203]
- Fishbein M, Jarvis B, Peterman et al. failure to find a behavioral surrogate for STD incidence: What does it really mean? *Sexually Transmitted Diseases* 2000;27:452–455. [PubMed: 10987450]
- Fishbein M, Hennessy M, Kamb M, Bolan G, Hoxworth T, Iatesta M, Rhodes F, Zenilman J. Using intervention theory to model factors influencing behavior change: Project RESPECT. *Evaluation and the Health Professions* 2001;24:363–384.
- Fortenberry J, Temkit M, Tu W, Graham C, Katz B, Orr D. Daily mood, partner support, sexual interest, and sexual activity among adolescent women. *Health Psychology* 2005;24:252–257. [PubMed: 15898860]
- Fortenberry J, Katz B, Llythe M, Juliar B, Tu W, Orr D. Factors associated with time of day of sexual activity among adolescent women. *Journal of Adolescent Health* 2006;38:275–281. [PubMed: 16488826]
- Greenberg J, Hennessy M, MacGowan R, Celentano D, Gonzales V, VanDevanter N. Modeling intervention efficacy for high-risk women - the WINGS project. *Evaluation and the Health Professions* 2000;23:123–148.
- Grello C, Welsh D, Harper M. No strings attached: the nature of casual sex in college students. *Journal of Sex Research* 2006;43:255–267. [PubMed: 17599248]
- Hemphill T. Self-regulation, public issue management and marketing practices in the US entertainment industry. *Journal of Public Affairs* 2003;3:338–357.
- Hemker B, Sijtsma K, Molenaar I. Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model. *Applied Psychological Measurement* 1995;19:337–352.
- Henderson V, Hennessy M, Barrett D, Curtis B, McCoy-Roth M, Trentacoste N, Fishbein M. When risky is attractive: sensation seeking and romantic partner selection. *Personality and Individual Differences* 2005;38:311–325.
- Hoyle R, Stephenson M, Palmgreen P, Lorch E, Donohew L. Reliability and validity of a brief measure of sensation seeking. *Personality and Individual Differences* 2002;32:401–414.
- Eng, TR.; Butler, WT., editors. Institute of Medicine. The Hidden Epidemic: Confronting Sexually Transmitted Diseases. Washington, D.C: National Academy Press; 1997.
- Jakobsen R. Stages of progression in noncoital sexual interactions among young adolescents: an application of the Mokken scale analysis. *International Journal of Behavioral Development* 1997;27:537–553.

- Jemmott J, Jemmott L, Braverman P, Fong G. HIV/STD risk reduction interventions for African American and Latino adolescent girls at an adolescent medicine clinic: a randomized controlled trial. *Archives of Pediatric and Adolescent Medicine* 2005;5:440–449.
- Johnson M. The dimensionality of perspectives on premarital sex: a comparison of Guttman and INDSCAL dimensionality. *The Journal of Sex Research* 1986;22:94–107.
- Kandel D, Yamaguchi K. From beer to crack: developmental patterns of drug involvement. *American Journal of Public Health* 1993;83:851–855. [PubMed: 8498623]
- Kennedy, P. *A Guide to Econometrics*. Cambridge: MIT Press; 2003.
- Kirby D, Blumler E, Coyle K, Basen-Engquist K, Parcel G, Harrist R, Banspach S. The "Safer Choices" intervention: its impact on the sexual behaviors of different subgroups of high school students. *Journal of Adolescent Health* 2004;35:442–452. [PubMed: 15581523]
- Kirby, D.; Lepore, G.; Ryan, J. National Campaign to Prevent Teen Pregnancy. Sexual and risk protective factors: Factors affecting teen sexual behavior, pregnancy, childbearing and sexually transmitted disease: Which are important? Which can you change?. ETR Associates; 2005.
- Kunkel, D.; Biely, E.; Eyal, K.; Cope-Ferrar, K.; Donnerstein, E.; Fandrich, R. Sex on TV3: A Biennial Report to the Kaiser Family Foundation. Menlo Park, CA: The Kaiser Family Foundation; 2003.
- Kunkel, D.; Cope-Farrar, K.; Biely, E.; Donnerstein, E. Sex on TV2: A Biennial Report to the Kaiser Family Foundation. Menlo Park, CA: The Kaiser Family Foundation; 2001.
- Kunkel, D.; Cope, K.; Colvin, C. Sexual messages on family hour television: Content and context. Oakland and Menlo Park, CA: Children Now and Kaiser Family Foundation; 1996.
- L'Engle K, Jackson C, Brown J. Early adolescents' cognitive susceptibility to initiating sexual intercourse. *Perspectives on Sexual and Reproductive Health* 2006;38:97–105. [PubMed: 16772191]
- Manlove, J.; Papillio, A.; Ikramullah, E. Not Yet: Programs to Delay First Sex Among Teens. Washington, D.C: The National Campaign to Prevent Teen Pregnancy; 2004.
- Martino S, Collins R, Kanouse D, Elliot M, Berry S. Social cognitive process mediating the relationship between exposure to television's sexual content and adolescents' sexual behavior. *Journal of Personality and Social Psychology* 2005;6:914–924. [PubMed: 16393024]
- Mokken, R. *A Theory and Procedure of Scale Analysis*. The Hague: Mouton; 1971.
- Mosher, W.; Chandra, A.; Jones, J. Sexual behavior and selected health measures: Men and women 15–44 years of age, United States, 2002. Advance data from vital and health statistics; no. 362. Hyattsville, MD: National Center for Health Statistics; 2005. 2005
- Ringdal G, Jordhøy M, Kaasa S. Measuring quality of palliative care: psychometric properties of the FAMCARE scale. *Quality of Life Research* 2003;12:167–176. [PubMed: 12639063]
- Ringdal K, Ringdal G, Kaasa S, Bjordal K, Wisløff F, Sundstrøm S, Hjermstad M. Assessing the consistency of psychometric properties of the HRQoL scales within the EORTC QLQ-C30 across populations by means of the Mokken scaling model. *Quality of Life Research* 1999;8:25–43. [PubMed: 10457736]
- Sijtsma K, Verweij A. Mokken scale analysis: theoretical considerations and an application to transitivity tasks. *Applied Measurement in Education* 1992;5:355–373.
- Silver E, Bauman L. The association of sexual experience with attitudes, beliefs, and risk behaviors of inner-city adolescents. *Journal of Research on Adolescence* 2006;16:29–45.
- Smith E, Udry R. Coital and non-coital sexual behaviors of white and black adolescents. *American Journal of Public Health* 1985;75:1200–1203. [PubMed: 4037163]
- Stephenson M, Hoyle R, Palmgreen P, Slater M. Brief measures of sensation seeking for screening and large-scale surveys. *Drug and Alcohol Dependence* 2003;72:279–286. [PubMed: 14643945]
- Stolzenberg R, Relles D. Tools for intuition about sample selection bias and its correction. *American Sociological Review* 1997;62:494–507.
- Streiner D. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment* 2003;80:99–103. [PubMed: 12584072]
- Ward L. Understanding the role of entertainment media in the sexual socialization of American youth: a review of empirical research. *Developmental Review* 2003;23:347–388.
- Ward L, Friedman K. Using TV as a guide: associations between television viewing and adolescents' sexual attitudes and behavior. *Journal of Research on Adolescence* 2006;16:133–156.

- Winship C, Mare R. Models for sample selection bias. *Annual Review of Sociology* 1992;18:27–350.
- Zuckerman M. Scales for sex experience for males and females. *Journal of Consulting and Clinical Psychology* 1973;41:27–29.
- Zuckerman M, Eysenck S, Eysenck H. Sensation seeking in England and America: Cross-cultural age and sex comparisons. *Journal of Consulting and Clinical Psychology* 1978;46:139–149. [PubMed: 627648]

Graphs by Gender

Figure 1. Sexual Behavior Index
Histogram of Cumulative Sex Score

Notes: Gender/Age/Race Interactions All Not Significant

Figure 2. Predicted Sexual Behavior Index
By Gender, Age, and Race

Table 1

Frequency of Sexual Behaviors for All Respondents and by Gender, Age, and Race

Sexual Behavior	All Respondents		Gender				Age						Race			
			Males		Females		14 years		15 years		16 years		White		African-American	
	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n
Deep kissing	68	451	66	179	68	272	62	147	70	158	71	146	64	222	71	229
Genital touching	54	451	55	179	53	272	44*	147	56	158	61	146	42*	222	66	229
Sexy dancing	48	452	65*	179	37	273	52	148	46	158	47	146	30*	222	66	230
Breasts touched (females only)	62	273	NA	--	62	273	54	82	61	93	68	98	54*	134	69	139
Oral sex, receive	28	451	27	179	29	272	20*	147	30	158	34	146	19*	222	38	229
Vaginal sex	26	451	26	179	26	272	17*	147	28	158	33	146	13*	222	38	229
Touching breasts (males only)	67	167	67	167	NA	--	57	65	72	57	76	45	51*	80	82	87
Oral sex, give	19	451	17	179	19	272	11*	147	20	158	25	146	19	222	18	229
Anal sex, give (males only)	12	175	12	175	NA	--	9	65	19	64	6	46	3*	87	20	88
Anal sex, receive (females only)	9	259	NA	--	9	259	4	77	9	90	13	92	5*	129	13	130

Notes: * t-test or F test value relating the column classification and the behavior has a prob value of <.05. NA: Not applicable.

Table II

Scaling Results for Sex Behavior Dichotomies

Sexual Behavior	Gender	
	Males (N = 161)	Females (N = 259)
Deep kissing	68%	68%
Touching breasts	66%	NA
Breasts touched	NA	61%
Genital touching	58%	54%
Oral sex, receive	29%	30%
Vaginal sex	28%	26%
Oral sex, give	19%	20%
Anal sex, receive	NA	9%
Anal sex, give	12%	NA
Loevinger's <i>H</i> KR20	.75 .84	.81 .85

Notes: NA: Not applicable. Sample sizes are lightly smaller here than in Table I dues to missing values on individual scale items.

Table III
Validating Correlations between Sex Score, Theoretical Variables, and Trait Variables

Theory Variables	Intentions	Attitudes	Self-Efficacy	Injunctive Norms	Descriptive Norms
Correlation with Sex Score	.65	.53	.46	.29	.46
Sig. Level	< .05	< .05	< .05	< .05	< .05
Sample Size	419	415	408	406	413

Trait Variables	Age	Physical Development	Lifetime Romantic Partner	Religiosity	Sensation Seeking
Correlation with Sex Score	.22 ^P	.28 ^P	.81 ^P	-.01 ^P	.16
Sig. Level	< .05	< .05	< .05	> .05	< .05
Sample Size	420	418	411	405	419

Notes: P indicates use of polychoric correlation due to limited variance in validating variable.