Published in final edited form as:

Arch Sex Behav. 2009 December ; 38(6): 922–935. doi:10.1007/s10508-008-9385-2.

Measuring Sexual Risk for HIV: A Rasch Scaling Approach

Michael Fendrich,

Helen Bader School of Social Welfare, Center for Addiction and Behavioral Health Research,
University of Wisconsin-Milwaukee, Enderis Hall, Room 1191, PO Box 786, Milwaukee, WI
53201, USA fendrich@uwm.edu

Everett V. Smith Jr.,

Department of Educational Psychology, University of Illinois at Chicago, Chicago, IL, USA

Lance M. Pollack, and

Center for AIDS Prevention Studies, University of California, San Francisco, San Francisco, CA,
USA

Mary Ellen Mackesy-Amiti

Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at
Chicago, Chicago, IL, USA

Abstract

In this study, we developed an HIV transmission risk scale and examined its psychometric properties using data on sexual behavior obtained from a probability sample of adult men who have sex with men living in Chicago. We used Messick's (Am Psychol 50:741–749, 1995) conceptualization of unified validity theory to organize the psychometric properties of data. Evidence related to scale content was investigated via Rasch item fit statistics, point-measure correlations, and expert evaluation. The substantive aspect of validity was addressed by interpreting the meaningfulness of the item difficulty hierarchy (continuum of risky behaviors) and assessment of person fit. The structural aspect of validity was assessed using Rasch item fit statistics, principal component analysis of standardized residuals, and other residual analyses. The generalizability aspect of validity was investigated via internal consistency reliability estimates for both items and persons, and aspects of external validity were addressed by examining between-group differences with respect to levels of risky behavior. Applications and suggested future studies are discussed.

Keywords

HIV; Sexual risk; Homosexuality; Rasch scaling

Introduction

The incidence of HIV/AIDS among men who have sex with men (MSM) has rebounded in recent years (CDC, 2005, 2007). The number of HIV/AIDS diagnoses among MSM increased 11% from 2001 through 2005 (CDC, 2007). By 2005, the CDC estimated that over 230,000 MSM were living with this disease (CDC, 2007). Concurrent with increasing rates of disease, there is evidence that sex without a condom and other risky sexual behavior continues to be widespread among MSM (Mansergh et al., 2001; Wolitski, 2005). There is

also evidence supporting a shift toward more risky sexual behavior when data obtained in the late 1990s is compared with data from the early part of this decade (Osmond, Pollack, Paul, & Catania, 2007).

Although over two decades of epidemiological and intervention research on HIV risk have elapsed, these statistics underscore the fact that the need for improved research in this area persists. If we want to design effective HIV behavior risk prevention programs—programs that are effective in the current decade—we need to better understand the correlates of risk behavior. Research on the correlates of risk behavior has approached the measurement of risk in a fairly rudimentary way.

Researchers examining the correlates of high-risk sexual behavior have typically used a number of different categorical and “count” measures and, occasionally, continuous risk behavior indices based on a single measure. Among the most common items selected for analysis are total number of partners within a time period (usually one year or less) (e.g., Greenwood et al., 2001), total number of “one night stands” within a time period (“had sex with men with whom sex only happened once”; e.g., Stall et al., 2001a), and episodes of unprotected anal intercourse (e.g., Osmond et al., 2007). The last measure, unprotected anal intercourse, has been operationalized in multiple ways. For example, Greenwood et al. (2001) and Stall et al. (2001a) defined unprotected intercourse as total episodes within a one-year time period. Some researchers have investigated unprotected intercourse as a percentage of all encounters within a time period (McKirnan, Venable, Ostrow, & Hope, 2001). Other researchers further refine the status of partners in the unprotected encounters. For example, Paul, Stall, Crosby, Barrett, and Midanik (1994) and Craib et al. (2000) evaluated whether the partner was “primary” or “non primary.” Koblin et al. (2003) and Osmond et al. (2007) took into account the partner’s HIV status (with the latter researchers categorizing unprotected behavior as “serodiscordant” risk to either the insertive or receptive partner).

While the level of detail in the assessment may be helpful in defining risk, it may place considerable cognitive demands on subjects, possibly increasing the potential for a variety of response errors (Tourangeau, Rips, & Rasinski, 2000). More problematic, however, is the tendency for researchers to rely on constructs based on a single measure. Researchers generally have treated each type of measure as a separate construct rather than combining items into an overall risk scale. The latter approach may be more statistically powerful and thus more informative in the analysis of etiological factors. Conceptualizing risk behavior as a “continuum” could also directly inform intervention strategies, i.e., strategies designed to prevent escalation to the highest level of sexual risk.

In the Rasch model approach, which is gaining more currency with behavioral researchers, a scale is assumed to measure a single ability or trait, as operationalized by a set of items that vary with respect to “difficulty.” A range of item difficulties is desired for norm-referenced interpretations as more precise measurement (lower *SEs*) is obtained when the difficulty level of the items match the range of person abilities. Thus, an effective scale (from a statistical perspective) is obtained when the difficulty level of the items match the range of person abilities. An effective scale for measuring any given trait would consist of a mix of difficult items (those that are only correctly answered or endorsed by the most able respondents), mid range items (those that are correctly answered or endorsed by all respondents except those with the lowest ability levels), and easy items (those that are correctly answered or endorsed by nearly everyone).

We propose that these concepts can be applied to the trait of sexual risk behavior (abbreviated in this article henceforth with “RB”). We operationally define this as the

tendency to engage in behaviors which increase the risk of HIV transmission. We theorize that certain sexual behaviors are characteristic only of those who are highest in RB, but would never be carried out by those who are average or lower than average in RB. Those who are lower than average in RB may, at most, engage in behaviors that carry only minor risk for disease transmission (although those who are average and above average in RB would take part in such behaviors as well). If the data fit the model requirements, the Rasch model will construct a linear scale for RB by using survey response data as realizations of the probabilities of item endorsement given the level of risk represented by that item (i.e., its difficulty) and the behavioral riskiness of the person responding to that item. Using the Rasch approach, our analyses will take an important first step in reconceptualizing the statistical analysis of sexual risk behavior in epidemiological research focused on MSM.

Specifically, in this study, we developed an HIV transmission risk scale and examined its psychometric properties using data on sexual behavior obtained from a probability sample of adult MSM living in Chicago with methods based on Rasch measurement (Smith & Smith, 2004, 2007; Wright & Masters, 1982; Wright & Stone, 1979). Using these methods, we attempt to provide evidence for various aspects of Messick's (1995) conceptualization of unified validity theory as interpreted by Wolfe and Smith (2007a, 2007b). We chose to couch our validity evidence in terms of Messick's unified validity theory because the unified perspective has replaced the traditional view of validity as consisting of content, criterion, and construct-related evidence among psychometricians and other researchers. This is exemplified by the most recent publication of the Standard for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Unfortunately, relatively few published validity studies have used unified validity theory to present their findings (Wolfe & Smith, 2007a). Therefore, this research will not only present evidence of the potential usefulness of the RB scale, but will also introduce Rasch measurement and the terminology of unified validity to a broader audience of researchers.

Method

Sample

A complete description of the sampling frame construction and overall sample demographics are described elsewhere (Catania et al., 2001). To summarize, The Urban Men's Health Study (UMHS 1997) drew a stratified probability sample of adult MSM from the areas demarcated by the geographic city limits of San Francisco, New York, Chicago, and Los Angeles. Data from the Federal Communications Commission and the 1990 U.S. Census indicated that telephone coverage in those cities was on the order of 93–95% (Lavrakas, 1987; U.S. Bureau of the Census, 1997). Furthermore, research found that households with telephones did not differ from households in general on health characteristics (Anderson, Nelson, & Wilson, 1998). Preliminary work identified areas of moderate-to-high MSM residential density within each city through mapping of MSM AIDS caseload data, male–male partnered household data from the 1990 U.S. Census, addresses from a gay commercial mailing list, and areas designated as gay neighborhoods by local informants. Those areas were operationalized as ZIP codes. Telephone exchanges covering the selected ZIP codes were identified and then stratified by *estimated* cost per interview (Binson et al., 1996). In order to limit costs while maximizing coverage, the sampling frame was constrained to ZIP codes in each city with an *estimated* cost per interview below \$1,000. Following the recommendations of Fowler (1989), disproportionate sampling (Kalton, 1993) and adaptive sampling techniques (Blair, 1999) were used to construct a random digit dial (RDD) sample for the designated ZIP codes in each city.

Prior to fielding the survey, community awareness programs (in the form of general publicity as well as organized town meetings) were conducted in each of the four cities to alert respondents as well as local public and community agencies to the study. All contacts were made by telephone using computer-assisted telephone interview (CATI) technology. Using adult informants only, households were first screened to determine geographic (ZIP code) and age-gender (at least one male resident age 18 or older) eligibility. An adult male informant was then used to screen the household for MSM eligibility by asking him a series of questions which allowed him to describe himself (or any other adult male in the household), either behaviorally (any same gender sexual contact since age 14) or through self-definition as gay or bisexual, to be an MSM. For households containing more than one adult MSM, one was randomly selected for interview. Interviews were conducted in English and Spanish and, on average, took 75 min to complete.

Specifically in Chicago, during the period from April 1997 through February 1998, 34,564 telephone numbers were dialed and 15,671 households were identified. A total of 10,133 households were successfully screened, of which 9,402 were geographically eligible, 5,795 were age-gender eligible, and 528 were MSM eligible. One of those households was dropped because the selected respondent was not competent to be interviewed. From the remaining 527 households, 414 interviews were obtained, a participation rate of 78.6%. Of these 414, 407 had sufficient data to be included in these analyses. Although probability of selection weights were constructed for the UMHS 1997, the data presented in the present study were unweighted since our primary goal was to investigate the relationships between items and constructs and not to make inferences about point estimations from the sample to the population it represents.

Table 1 provides background characteristics for the 407 participants. The majority of the men in the sample were sexually active within the past year (86%), and 46% reported that they had a primary partner, i.e., a partner with whom they were in love or felt a special commitment to. Most of the participants had received HIV testing (88%), and 14% had tested positive. The sample was mostly White (79%) and the educational level of the sample was high, with 76% having a college degree or greater. The former appears to be the result of in-migration by white MSM (Catania et al., 2001) and the latter apparently is a correlate of childlessness (Catania, Canchola, Pollack, & Chang, 2006). The mean age of the sample was 37 years.

Measures

The UMHS 1997 cohort is the only multi-city probability sample of adult MSM in the U.S. The survey was designed to yield data that would provide an estimation of the prevalence of HIV and HIV-related risk behavior (Catania et al., 2001; Dolcini, Catania, Stall, & Pollack, 2003; Mills et al., 2001) and information on correlates of risk. Interviews covered a range of social, psychological, and behavioral phenomena related to HIV. Major sections of the survey instrument included questions about the gay community (involvement, services used, attitudes toward) (Barrett & Pollack, 2005), demographic characteristics (Catania et al., 2006), attendance at gay venues ("places where men go to meet and socialize with other men") (Binson et al., 2001; Woods et al., 2003), sexual behavior (both a "global" and partner-by-partner assessment), sexual development, including the "coming out" process (Barrett, Pollack, & Tilden, 2002), experiences with harassment and violence (including adverse familial events, anti-gay victimization, sexual coercion, and partner violence) (Arreola, Neilands, Pollack, Paul, & Catania, 2005; Greenwood et al., 2002; Paul, Catania, Pollack, & Stall, 2001; Relf, Huang, Campbell, & Catania, 2004), sexual problems, depression, and suicidal ideation (Mills et al., 2004; Paul et al., 2002), substance use (Klitzman, Greenberg, Pollack, & Dolezal, 2002; Stall et al., 2001a, 2003), HIV testing and serostatus (Osmond et al., 2000; Stall et al., 2001b), and AIDS care-giving.

In the present study, risk transmission was operationalized as the risk of either transmitting or contracting HIV. UMHS 1997 included a “global” assessment of sexual behavior in the past 12 months by asking the respondent with how many different men did he engage in each of 16 sexual behaviors: 6 oral sex behaviors (performing/receiving oral sex without condoms to ejaculation/without condoms but withdrawal before ejaculation/with condoms), 6 anal sex behaviors (insertive/receptive anal sex without condoms to ejaculation/without condoms but withdrawal before ejaculation/with condoms), rimming, fisting, use of sex toys, and bondage/discipline/S&M practices (the latter four items making no distinction between performing and receiving). For our analyses, the two oral sex and two anal sex items involving condom use were dropped from the scale because they are non-risky behaviors while the bondage item was dropped as being too non-specific. The remaining items were recoded into yes/no dichotomies.¹

Additional questions explored the context in which unprotected anal sex (i.e., without a condom) in the past 12 months occurred. Respondents were asked separately if they had sex in a public place (“like a bookstore, park, club, or bathhouse”) and in group situations (with two or more other people). Affirmative responses were followed by a two-question sequence asking if they engaged in insertive or receptive anal intercourse without condoms when they had sex in that context. In order to avoid extending the instrumental dependency within the two 3-item sequences (see original items 13–15 and 16–18 in the Appendix)² to the current analysis, data from these questions were combined to formulate two yes/no dichotomies for unprotected anal intercourse in a public place (UAIPP) and unprotected anal intercourse in a group situation (UAIGS).

A subsequent partner-by-partner sexual assessment specifically asked about anal sexual behavior with the respondent’s four most recent partners, but always included an assessment of his primary partner if he had one. This extra information allowed for the derivation of a yes/no dichotomy reflecting whether or not the respondent engaged in unprotected anal intercourse with a secondary (i.e., non-primary) partner. If the respondent reported two or more partners for any of the four risky anal sex behaviors (insertive/receptive anal sex without condoms to ejaculation/without condoms but withdrawal before ejaculation—see the Appendix), or one partner for any of the four behaviors that was not their primary partner, then he was designated to have engaged in unprotected anal sex with a secondary partner.

¹Responses to items 1–11 were in the form of counts while responses to items 12–18 were dichotomous. Therefore, the data could be modeled using a combination of the Rasch dichotomous and Poisson counts models. Unfortunately, count data rarely fit the Poisson counts model (see Smith & Kulikowich, 2004) and such was the case with the current analysis. Therefore, all counts data were recoded into dichotomies (0 = did not engage in the behavior; 1 = engaged in the behavior) and the Rasch dichotomous model implemented to model the data.

²For example, consider items 13–15 and 16–18 in the Appendix. A positive response to either 14 or 15 would necessitate a positive response to item 13. Likewise for the relationship among responses to items 17 and 18 with respect to item 16. As Rasch models require local independence (i.e., after controlling for the construct of interest, the responses to items should be statistically independent), the decision was made to recode response to items 13–15 and 16–18 into new indicators of risky behavior. Specifically, if items 13 or 16 were not endorsed, then the respective recoded variables were assigned a value of 0 (i.e., did not have sex in a public place). If these items were endorsed and both the corresponding follow-up questions (items 14–15 for item 13 and items 17–18 for item 16) were not endorsed, the respective recoded variables were also assigned a value of 0 (i.e., did not engage in UAI in a public place). Thus, a value of 0 on the newly recoded items would mean the participant did not have UAI in a public place although he may have engaged in other forms of sex in a public place. If item 14 or 15 (and, correspondingly, item 17 or 18) were endorsed, then the respective recoded variables were assigned a value of 1 (i.e., did engage in UAI in a public place). Therefore, any UAI was considered as risky behavior. Other potential dependencies may exist in the data. For example, items 5–8 address UAI and items 16–18 UAI in the context of group sex. So an endorsement of either item 17 or 18 would require endorsement of at least one of the items 5–8. This relationship, however, is unidirectional as endorsement of behaviors in items 5–8 does not require an endorsement of items 17 or 18. Therefore, we decided to leave these “partially” dependent items alone and let the Rasch residual analyses determine if the degree of local dependence was enough to warrant post-hoc recoding of the data.

These last three “contextual” items (UAIPP, in a group situation, or with a secondary partner) focus on unprotected anal sex with casual or anonymous sex partners. These partners were men for whom the respondent was likely to have little or no information concerning their HIV serostatus or sexual history. Thus, the risk for HIV transmission was magnified. See the Appendix for the original 18 items.

Statistical Analysis

In the sections that follow, the methods used for measurement construction and data analysis are outlined. We briefly outline how each of the methods is couched in terms of Messick’s (1995) conceptualization of validity as interpreted by Wolfe and Smith (2007a, 2007b). For the current investigation, the following aspects in Messick’s framework are addressed: content, substantive, structural, generalizability, and external.

Unified Validity—The content aspect of validity addressed the relevance and representativeness of the content of the items and the technical quality of those items. In our study, the item development process as detailed in the *Instrumentation* section gives support for the content aspect of validity by providing details on item development, selection, and recoding. The technical quality of the items was checked using item–measure correlation (analogous to the traditional item–total correlations in classical test theory) and standardized item mean-square fit indices (Wolfe & Smith, 2007a). The item–measure correlation was simply a Pearson correlation between the item scores and the estimates of the respondents’ level of RB. The correlation should be positive with negative values indicating a likely miscode (i.e., forgetting to reverse recode a negatively phrased item) or an item not related to the remaining items on the scale. The fit statistics are described in detail when the Rasch dichotomous model is introduced. The *structural* aspect of validity is concerned with the degree to which the scoring structure conforms to the dimensional structure of the construct. Evidence can take the form of correlation analyses of response consistencies (e.g., factor analytic models) and indicators that the Rasch dichotomous model requirements are satisfied. In our study, we investigated the Rasch model requirements of local independence and unidimensionality using principal component analyses of the standardized residuals, residual correlations, and Rasch item fit statistics (Wolfe & Smith, 2007b). Substantive validity seeks evidence to explain the observed consistencies among item responses. For this aspect of validity, we evaluated person fit statistics and the item difficulty hierarchies (Wolfe & Smith, 2007b). The *generalizability* aspect of validity evaluates the degree to which item and person measures maintain their meaning across contexts. In this study, the relevant support for generalizability came from estimates of internal consistency reliability for both items and persons (Wolfe & Smith, 2007b). Finally, the *external* aspect of validity covers what has traditionally been referred to as convergent and discriminant validity and also entails the applied utility of the measures, between-group differences, and within-person changes over time. For our exploration of evidence for external validity, we make various between-group comparisons to investigate any potential differences among levels of RB with respect to various background characteristics. For the analyses addressing the external aspect of validity, only those men indicating sex with other men in the last 12 months were included.

Rasch Dichotomous Model—The Rasch model for dichotomies (Rasch, 1960/1980; Wright & Masters, 1982; Wright & Stone, 1979) was the primary method used for the item analyses. The Rasch dichotomous model is appropriate for any scoring structure with two ordered categories (e.g., yes/no, correct/incorrect, and present/absent). The program used for the Rasch analysis is Winsteps (Linacre, 2005), which first uses a normal approximation algorithm to obtain initial estimates of model parameters and uses these initial estimates for

iterative Joint Maximum Likelihood Estimation. The iterative process stops once convergence criteria are reached.

Rasch models may be used to solve a variety of measurement problems (Smith & Smith, 2004, 2007; Wright, 1977). For example, Rasch models can be used for locating a person on the latent continuum, for understanding the structure of items, for standard setting, equating, and differential item functioning.

The rationale for using Rasch models over the more commonly used traditional item statistics such as p -values and corrected point-biserial correlations as the main method of analysis is a consequence of the many deficiencies of classical test theory upon which traditional item analysis is based. The most pertinent limitation of classical test theory is the sample dependency of item and test indices (e.g., bi-serial, point bi-serial, item-total correlations, p -values, and reliability indices) and the item dependency of person's ability (used generically in this case to represent the level on any latent construct). For example, if a test item is given to both a more able and a less able group, the item p -values will be higher (i.e., indicating easier or more frequently endorsed items) for the more able group. Likewise for persons, a person will appear more able if an easy test is administered versus a more difficult test. These limitations do not occur in Rasch measurement. When the data fit model expectations, the item and person parameters are freed from the distributional properties of incidental parameters.

Other limitations include the inability of classical test theory to determine how a person might respond to any given item. The different metrics for persons and items does not allow the possibility of predicting the outcome of the interaction between person ability and item difficulty. For example, under classical test theory, it is not possible to predict how a person who endorsed 50% of the items would perform on another item that was endorsed by 80% of the persons responding. Furthermore, many of the statistical models based on classical test theory assume at least an interval scale of measurement. However, raw scores for person and items (and their linear transformations) obtained from tests and surveys are ordinal. This makes valid mathematical comparisons among individuals, groups, or items difficult as equal raw score differences between pairs of points do not necessarily imply equal amounts of the construct under investigation. Ordinal scales of measurement do not support the mathematical operations needed to calculate means and SD (Mebitz, Morris, & Grip, 1989; Wright & Linacre, 1989).

Other weaknesses of classical test theory include (1) the lack of procedures for determining how measurement error varies across the range of values of the latent construct (usually only one set of measurement is applied to all scores although it is known that extreme scores are less precise), (2) the inability to directly compare scores obtained from the same set of items unless complete data are available, and (3) the lack of techniques for validating response patterns (e.g., if an individual does not endorse the five easiest items and endorses the five most difficult items, is 50% a valid reflection of his/her ability?).

When implemented properly, Rasch models overcome these limitations of classical test theory (Wright, 1977). The specific model used for the current analysis is the dichotomous Rasch model given by:

$$\ln\left(\frac{P_{ni}}{1 - P_{ni}}\right) \equiv B_n - D_i \text{ or equivalently} \quad (1)$$

$$P_{ni} = \frac{\exp(B_n - D_i)}{1 + \exp(B_n - D_i)} \quad (2)$$

Here, P_{ni} is the probability of person n with level of risk behavior B_n endorsing item i which has difficulty level D_i .

This model provides estimates of item locations (item difficulty, which is indexed by frequency of endorsement) and person proficiency (level of RB) along a common logit (log-odds) continuum. A logit is defined as the natural log odds of endorsing an item chosen to represent the center (or “zero” point) of the measurement scale. Being on a common scale means that person RB can be directly related to the item content. Consider the following example,

<i>lower RB</i>	P1	P2	P3	<i>higher RB</i>
<i>more frequently endorsed items</i>	I1	I2	I3	<i>less frequently endorsed items</i>

where the P's represent persons and the I's represent items. Given that P1 is lower than all the items, P1 would not likely endorse any of these items. P2 would likely endorse I1 and not endorse I2 and I3 while P3 would likely endorse all three items.

If the data fit the model, this logit continuum is on an interval scale, making the estimates appropriate for parametric statistical analysis. Greater logit values for items indicate increasing item difficulty (less frequently endorsed risk behaviors); persons with higher logit values have higher levels of RB than persons at lower levels. Standard errors associated with each estimate quantify the precision of the estimate.

Rasch Model Assumptions—The assumptions of the dichotomous Rasch model include unidimensionality and local independence. Unidimensionality means that a single dominant construct is being measured. Local independence means that, after controlling for the latent trait, responses to items should be independent of each other. The models further assume relatively homogeneous discriminating power among the items and minimal guessing. These assumptions may in part be evaluated using fit statistics and principal component analysis (PCA) of standardized residuals provided by Winsteps along with other residual analyses.

Once the parameters of a Rasch model are estimated, they are used to compute expected responses for each person to every item. Fit statistics are then derived from a comparison of the expected and observed responses. Item fit statistics are used to identify items that may not be contributing to a unidimensional construct or items that may be statistically dependent with other items. Those items identified as misfitting the model need to be examined for potential reasons for the misfit and then either revised or eliminated.

Winsteps provides two types of fit statistics for persons and items: Infit, which is less sensitive to surprising responses to items far from a person's RB level, and Outfit, which is sensitive to atypical behavior on items far from a person's RB level. When reported as mean-square statistics, the Infit and Outfit values are simply χ^2 statistics divided by their degrees of freedom. This results in an expected value of 1 and a range from 0 to ∞ (Linacre & Wright, 1994). Values <1 suggest a lack of stochasticity in the data, potentially due to a violation of local independence. Values >1 are indicative of excessive variability, which may signify a departure from unidimensionality. Because the distribution of mean-square statistics is not symmetrical around the mean, the interpretation of misfit has a different Type I error rate for the upper and lower tails (R. Smith, 2004). As a consequence, mean-

square fit statistics are often transformed into standardized fit statistics using a Wilson–Hilferty cube-root transformation (Wilson & Hilferty, 1931). These standardized fit statistics have an approximate unit normal distribution (mean = 0 and $SD = 1$). Values of standardized infit or outfit greater than +3.0 will be used to identify items for further review. Unless obvious content redundancy is apparent, values less than –3.0 will not be further explored as significant amounts of statistical redundancy needs to be present to impact the measurement of persons (Smith, 2005).

A PCA of standardized residuals was performed as a complement to the use of fit statistics when looking at the dimensionality of the data. The PCA of standardized residuals has an advantage over fit statistics in detecting departures from unidimensionality when (1) the level of common variance between components in multidimensional data increases and (2) there are approximately an equal number of items contributing to each component (E. Smith, 2004a). To judge whether a residual component adequately constitutes a separate dimension, we will look at the size of the first eigenvalue (units of the previously unexplained variance that are now attributable to this residual contrast) (E. Smith, 2004a).

Adherence to the local independence assumption is important as violations may lead to inaccurate estimates of item difficulties and person RB levels and may result in an overestimation of reliability and information functions (Sireci, Thissen, & Wainer, 1991; Thissen, Steinberg, & Mooney, 1989). The degree of local dependence is evaluated through the computation of Fisher's Z statistic (Shen, 1996, 1997; Shen & Yen, 1997), which involves three steps:

1. calculate the standardized residuals for each candidate–item interaction,
2. correlate the standardized residuals, and
3. compute Fisher's Z statistic to normalize the Pearson correlation, using the formula:

$$Z_{ij} = \frac{1}{2} \ln \frac{(1+r_{ij})}{(1-r_{ij})} = 1/2 \left[\ln(1+r_{ij}) - \ln(1-r_{ij}) \right] \quad (3)$$

A pair of items is considered to be statistically dependent if their Fisher's Z statistic is more than two standard deviations above or below the mean of Fisher's Z .

Rasch Reliability—Rasch analysis provides internal consistency reliability estimates for both persons and items ranging from 0.0 to 1.00. They are indices of how well the persons or items are spread out along the continuum. Conceptually, Rasch person reliability is analogous to Cronbach α /KR 20 in the classical test theory in terms of interpretation and calculation (see Smith 2001 for the advantages of using Rasch estimates of internal consistency). Rasch item reliability is an important aspect for construct validation as it indicates the spread of items along the continuum of interest. A spread of items is required to form a well-defined variable for interpretation (Smith, 2001).

Results

Our results are presented for each aspect of Messick's validity framework that we examined.

Evidence of Content Validity

Previous sections detailed the survey/item development process. Further evidence for the content aspect of validity was sought using item–measure correlations and Rasch item fit

statistics. As a technical quality indicator of consistency between scores on the items and the average scores across the remaining items, the item–measure correlations were all positive (ranging from .28 to .81). Another indicator of the technical quality of items, the standardized mean-square item fit statistics will be described in the next section as they also contribute to the dimensionality evaluation aspect of structural validity.

Structural Aspect of Validity

The fit of the items to Rasch model expectations revealed items 1 (Received oral sex without a condom, with withdrawal before ejaculation) and 11 (Used sex toys, such as dildoes or vibrators) were associated with fit statistics greater than 3. The misfit of item 1 was due to 6 persons. These six persons had large standardized residuals associated with their observed versus expected responses. For item 11, there were 3 persons responsible for the misfit. At this point three courses of action are available: (1) delete the item if an item flaw can be found or if the item can be deemed to measure another construct separate from the other items, (2) delete the person if they can be reasonably identified as coming from a population other than that targeted for the survey purpose, or (3) edit the response data and change the unexpected observations to missing (Rasch models are robust to missing data—see Linacre, 2005). The third option was selected for two reasons. First, the misfit of items 1 and 11 resulted from only 1.5 and .7%, respectively, of the respondents. Conversely, this means that these items were functioning ‘correctly’ for 98.5% and 99.3%, respectively, of respondents. Second, since the main aim of this research was scale development, retention of as many items as possible for content validity purposes was a priority. Accordingly, we decided to retain all items except those that functioned so poorly that they would inhibit measurement construction.

With the misfitting person 9 item interactions edited, examination of the recalculated item fit statistics revealed that the misfit of item 3 (Performed oral sex without a condom, with withdrawal before ejaculation) was due to seven unexpected responses (five of these seven were also responsible for the misfit in the first calibration—diagnosing these misfitting person is discussed below). After editing of these responses, the third calibration yielded item fit statistics all within the specified criteria.

As discussed previously, item fit statistics are not useful in all situations to identify items departing from model assumptions and a PCA of standard residuals can be used to supplement the item fit statistics in assessing the dimensionality of the data. The PCA of the standardized residuals demonstrated that the Rasch measures accounted for 97.6% of the total variance with the first residual contrast accounting for .3% of the total variance. The strength of the first residual contrast (i.e., eigenvalue) was deemed too small to have any impact on the total measurement structure. This evidence supports the unidimensional interpretation of the collection of items comprising the survey.

Investigation of local independence using Fisher’s Z statistic revealed three residual correlations beyond the specified criteria, meaning responses to these item pairs have something in common over and above the construct being defined by the collection of items. These item pairs are as follows: (1) “Received oral sex without a condom, with withdrawal before ejaculation” and “Performed oral sex without a condom, with withdrawal before ejaculation”; (2) “Received oral sex where he ejaculated in your mouth, without a condom” and “Performed oral sex where you ejaculated in his mouth, without a condom”; and (3) “Unprotected anal intercourse with a secondary partner” and the recoded item 13 dealing with “sex in a public place.”

The large residual correlations between the two sets of oral sex item pairs are expected because sexual reciprocity appears to be the behavioral norm for MSM sexual dyads (Jones

& Candlin, 2003; Middlethorpe & Aggleton, 2001; Vincke, Bolton, & De Vleeschouwer, 2001). A correlation between unprotected anal sex with a secondary partner and unprotected anal sex in a public place was also expected because partners met in a public sex environment are almost always not the primary partner. Although these residual correlations are readily interpretable, we decided to retain the coding of the items and not further combine item responses because the identified residual correlations represent <4% of the all the residual correlations (5% would be expected by chance if the data fit the model), and given the minimal levels of residual correlations, the estimates of item and person parameters and indicators of item and person reliability were unlikely to be adversely impacted (Smith, 2005).

Validity Relevant to Generalizability

Large variability in the item difficulties led to a high item reliability of .99, indicating that the items are well spread out (beyond measurement error) along the variable and afford the possibility of defining a meaningful item hierarchy (see Substantive validity). Person reliability (analogous to KR20) was .79. This is considered quite high for a new scale and “solid” for an extant scale (see, for example, Crano and Brewer, 1973).

Substantive Aspect of Validity

Two aspects were addressed for the substantive aspect of validity—person fit and meaningfulness of the item hierarchy. For person fit we first return to the results presented for the structural aspect of validity; specifically the large standardized residuals associated with the misfitting items identified prior to any editing of response data. The purpose of this discussion is to demonstrate the potential diagnostic utility of further examining the unexpected response patterns of the persons associated with these large standardized residuals. Figure 1 provides two expected score maps (for two different people) produced by Winsteps. For each map the items were ordered from least frequently endorsed (top) to most frequently endorsed. The unexpected responses are inside the parentheses and are associated with the observed response (indicated as a 0 or 1 on the same line). Observed responses that are not associated with large standard residuals are provided between two periods. The logit metric is provided along the horizontal axis and ranges from −5 to 3.

The underlying logic of the HIV transmission risk scale is that MSM “work their way up” from less risky behaviors to more risky behaviors, i.e., that anyone who reports the more risky behaviors will likely also report the less risky behaviors, but not vice versa. The two maps above demonstrated two different violations of that underlying logic. In the first (person 51), misfit was exemplified by a respondent who reported sex toy use without reporting any of the other behaviors listed, particularly the more frequently reported items “below” the sex toys item. This pattern probably arose because, for the subset of MSM represented by this case, use of sex toys provided an alternate sexual script to the usual patterns of oral then anal sex. That is, sex toy use was the sexual objective, and thus the typical sequence of behaviors was bypassed. In the second figure, the misfit of this respondent (person 66) was the result of failing to endorse the most frequently reported behaviors (oral sex with withdrawal) although he did report several other riskier behaviors. Two different contexts might explain this deviation. One is that members of long-standing dyads may no longer see the need for withdrawal during oral sex. The other is that MSM who frequent public sex environments or patronize sex workers may do so in order to obtain access to riskier behaviors. In such contexts, less risky behaviors may be dispensed with.

In the ideal context, the theory behind the unexpected responses provided above could be explored further using follow-up interviews. This was not feasible in the current study. These maps, however, hopefully provide readers with a general idea of the utility of Rasch

measurement to investigate and diagnose the observed (in)consistencies among item responses. It should also be noted that after the editing of response data only two persons were associated with standardized person fit statistics greater than 3. This indicates that the observed responses were consistent with the ordering of the items as defined by the majority of respondents and thus provided support for the substantive aspect of validity.

With respect to the ordering of the items, we have already discussed that the Rasch model placed the RB levels of persons and item difficulties on a common logit scale. This relationship is illustrated by a person/item variable map (see Fig. 2). In the map, the people having higher levels of RB and the more difficult to endorse behaviors were on the higher positions along the logit scale. Conversely, those persons who displayed less RB and behaviors easier to endorse were lower on the logit scale. For our purposes, we were interested in the ordering of the items for support of the substantive aspect of validity. Readers interested in how these maps may be used for both norm and criterion reference interpretations should see E. Smith (2004b), Woodcock (1999), and Wright, Mead, and Ludlow (1980).

Risk of HIV transmissibility during a given unprotected sexual act depends for the most part on the type of behavior engaged in, the HIV serostatus of the insertive partner, and the HIV status of the receptive partner (Moss et al., 1987; Osmond et al., 2007; Winkelstein et al., 1987). As displayed in the item hierarchy, generally speaking, adult MSM appear to function as rational actors vis-à-vis transmission risk for HIV with the item ordering matching expectations. For example, in the person/item variable map, the most frequently reported behaviors were the least risky for HIV transmission (1mrorlwd = Received oral sex without a condom, with withdrawal before ejaculation and 3mporlwd = Performed oral sex without a condom, with withdrawal before ejaculation) while the least frequently reported behaviors were the riskiest for HIV transmission (14uaigs = Unprotected anal intercourse during group sex and 13uaipp = unprotected anal intercourse in a public place). Confirmation of the expected ordering of items from least to most risk provided additional support for substantive validity.

External Aspect of Validity

Evidence to support external validity was investigated by examination of between group differences. Specifically, we compared Rasch measures of RB with respect to a number of attributes among men who were same gender sexually active in the previous 12 months ($n = 352$).

RB measures varied significantly by presence of a primary partner, and HIV status; RB measures did not differ significantly by education level, race/ethnicity, or age (see Table 2 for the descriptive statistics). Men who had a primary partner received RB measures indicative of more risky behavior compared to those without a primary partner ($t[349] = 3.24, p < .001$). This was consistent with findings suggesting that MSM were more likely to engage in unprotected anal sex with primary partners (Hays, Kegeles, & Coates, 1997; McClean et al., 1994; Theodore, Duran, Antoni, & Fernandez, 2004). Men who tested positive for HIV (all of whom self-reported positive in the interview) also received RB measures indicative of more risky behavior, compared to those who tested negative or were not tested ($t[333] = 2.04, p < .05$). This was also consistent with previous research findings suggesting that MSM who were HIV-positive engaged in serosorting, having unprotected sex with other HIV-positive men (Cox, Beauchemin, & Allard, 2004; Golden, Brewer, Kurth, Holmes, & Handsfield, 2004; Parsons et al., 2006).

Although RB levels were not significantly associated with race/ethnicity ($F[2, 328] = 2.56, p < .10$), Hispanic respondents ($N = 27$) had the highest mean RB level and Black/African-

American respondents ($N = 27$) demonstrated the lowest RB level. RB measures also tended to decline with increasing age, although this relationship was not statistically significant. Finally, there were no differences among RB levels with respect to level of education; we note, however, that the skewed educational distribution (most respondents were college-educated) may have limited the power to detect differences on this variable.

The strong relationships between RB levels and other markers of sexual risk were supported by prior research. These findings, when considered along with weak to non-existent relationships between RB levels and demographic characteristics, provided strong support for the external aspect of validity.

Discussion

Epidemiological researchers of HIV risk typically collect rich and detailed survey information about sexual activity that may or may not be related to disease transmission risk. Risk analyses typically focus on discrete behaviors with particular emphasis on identifying trends in relation to those that are objectively perceived as the most risky (i.e., unprotected receptive anal intercourse). Researchers have not taken advantage of up-to-date psychometric procedures, such as Rasch scaling, to develop measures that classify individuals based on the full range of their reported sexual activity. This article suggests that such classification is feasible and that a valid index of behavioral risk, based on the totality of items collected about potentially risky sexual behaviors, can be developed.

Using the validity framework as conceptualized by Messick (1995) and interpreted by Wolfe and Smith (2007a, 2007b), we assessed five aspects of construct validity: content, substantive, structural, generalizability, and external. Evidence related to scale content was obtained by detailing the development and coding process and via evaluation of item quality using Rasch item fit statistics and point-measure correlations. Unidimensionality of the data was supported using PCA of standardized residuals and Rasch item fit statistics. Local independence was also supported by the lack of numerous large correlated standard item residuals. Support for both unidimensionality and local independence (assumptions of the Rasch model employed) contributed to the structural aspect of validity. Support for the generalizability aspect of validity was investigated via internal consistency reliability estimates for both items and persons. Item reliability was high, indicating that the items were spread out (had statistically different levels of item difficulty) on the HIV transmission risk continuum beyond what could be accounted for by measurement error. The person reliability was also reasonably high given only 14 dichotomous scored items were employed. Support for the substantive aspect of validity came from the very few misfitting persons (meaning respondents interpreted the item hierarchy in a similar manner) and the meaningfulness and expected order of the items as displayed in the person/item variable map. Finally, for the external aspect of validity, group comparisons were made to investigate any potential differences in level of RB. Significant difference among levels of RB were found for other markers of sexual risk (having a primary partner and HIV status) while weak to non-existent relationships were found between RB levels and demographic characteristics (age, education level, and race/ethnicity).

With respect to how others can use our results without having to calibrate their data using a Rasch scaling program, Table 3 provides a raw score to Rasch logit conversion table. In this table, each possible raw score (0–14) has a corresponding Rasch measure (extreme raw scores have estimated (E) Rasch measures as extreme scores are not used in the estimation procedure in Winsteps) and standard error. The primary use of this table would be to convert the ordinal raw scores into corresponding interval level measures if parametric statistical analysis of the data is to be conducted. This would remove one threat to internal validity as

it has been demonstrated that the use of raw scores in ANOVA models can lead to spurious effects and underestimation of effect sizes (Embretson, 1996; Romanoski & Douglas, 2002). A second use would be to compare the results of other studies using similar instruments (e.g., the UMHS 1997 data collected from other cities during the same time period or later surveys employing UMHS-like methodology) to the result reported here. This could be as simple as locating an individual in the person/item variable map to see where in the current distribution an individual falls or the comparison may be based on the group means reported in Table 2. The caveat in using this raw score to Rasch measure table is that it requires complete data (each individual must respond to all 14 items) and assumes the new data will fit the Rasch model assumptions (see Fig. 1 for examples of violations). As such, we would endorse using this conversion table in the manner outlined above only if a Rasch calibration program is not available to investigate the degree to which the person response patterns adhere to the Rasch model expectations (i.e., fit the model expectations).

Limitations and Future Research

The present study is the first step to assess the psychometric properties of data from the HIV Transmission Risk Scale. Future studies need to concentrate on gaining additional validity evidence. Specifically, future studies should gain additional support for the generalizability aspect of validity. This study focused on one city at one point in time. It also relied completely on self-report data. Such studies should examine the stability (i.e., generalizability) of the item difficulty estimates across other cities (e.g., San Francisco, New York, and Los Angeles) included in the Urban Men's Health Study (UMHS 1997). Studies of differential item functioning should also be conducted on important demographic groups when sufficient sample sizes exist. These types of studies will help determine whether the item ordering as delineated by the Chicago sample can be applied more generally.

While this article demonstrates the feasibility of Rasch scaling of RB, it is up to future research to demonstrate its utility for the field. In this light, we believe that future researchers employing the Chicago UMHS and studies using similar data can use the individual scores derived from the RB as input into more general epidemiological analyses exploring the interrelationship between HIV transmission risk and other risky behaviors. For example, in future work, we plan to assess the association between RB scores and patterns of drug and alcohol use in the Chicago UMHS sample (using RB as a dependent variable). In these analyses, we hope to explore whether qualitative and quantitative indicators of drug involvement correspond with incremental levels of RB risk. This important epidemiological strategy, which has considerable potential for informing prevention research, will be facilitated by the reconceptualization of risk behavior facilitated by the Rasch analyses demonstrated in this work.

Acknowledgments

We wish to acknowledge support for this work provided by National Institute on Drug Abuse Grant 5R01DA018625.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 1999.
- Anderson J, Nelson D, Wilson R. Telephone coverage and measurement of health risk indicators: Data from the National Health Interview Survey. *American Journal of Public Health*. 1998; 88:1392–1395. [PubMed: 9736886]

- Arreola SG, Neilands TB, Pollack LM, Paul JP, Catania JA. Higher prevalence of childhood sexual abuse among Latino men who have sex with men than non-Latino men who have sex with men: Data from the Urban Men's Health Survey. *Child Abuse and Neglect*. 2005; 29:285–290. [PubMed: 15820544]
- Barrett DC, Pollack LM. Whose gay community? Social class, sexual self-expression, and gay community involvement. *Sociological Quarterly*. 2005; 46:437–456.
- Barrett DC, Pollack LM, Tilden ML. Teenage sexual orientation, adult openness, and status attainment in gay males. *Sociological Perspectives*. 2002; 45:163–182.
- Binson D, Moskowitz J, Mills T, Anderson K, Paul J, Stall R, et al. Proceedings of the Section on Survey Research Methods, American Statistical Association. 1996. Sampling men who have sex with men: Strategies for a telephone survey in urban areas in the United States; p. 68-72.
- Binson D, Woods WJ, Pollack L, Paul J, Stall R, Catania JA. Differential HIV risk in bathhouses and public cruising areas. *American Journal of Public Health*. 2001; 91:1482–1486. [PubMed: 11527785]
- Blair J. A probability sample of gay urban males: The use of two-phase adaptive sampling. *Journal of Sex Research*. 1999; 36:39–44.
- Catania JA, Canchola J, Pollack L, Chang J. Understanding the demographic characteristics of urban men who have sex with men. *Journal of Homosexuality*. 2006; 51(3):33–51. [PubMed: 17135114]
- Catania JA, Osmond D, Stall RD, Pollack L, Paul JP, Blower S, et al. The continuing HIV epidemic among men who have sex with men. *American Journal of Public Health*. 2001; 91:907–914. [PubMed: 11392933]
- CDC. Trends in HIV/AIDS diagnoses—33 states, 2001–2004. *Morbidity and Mortality Weekly Report*. 2005; 54:1149–1153. [PubMed: 16292247]
- CDC. HIV/AIDS surveillance report. Vol. 17. Atlanta: Department of Health and Human Services, CDC; 2007. p. 2007-146. Retrieved February 22, 2008 from <http://www.cdc.gov/hiv/topics/surveillance/resources/reports/>
- Cox J, Beauchemin J, Allard R. HIV status of sexual partners is more important than antiretroviral treatment related perceptions for risk taking by HIV positive MSM in Montreal, Canada. *Sexually Transmitted Infections*. 2004; 80:518–523. [PubMed: 15572627]
- Craib KJP, Weber AC, Cornelisse PGA, Martindale SL, Miller ML, Schecter MT, et al. Comparison of sexual behaviors, unprotected sex, and substance use between two independent cohorts of gay and bisexual men. *AIDS*. 2000; 14:303–311. [PubMed: 10716507]
- Crano, WD.; Brewer, MB. *Principles of research in social psychology*. New York: McGraw-Hill; 1973.
- Dolcini MM, Catania JA, Stall RD, Pollack L. The HIV epidemic among Older men who have sex with men. *Journal of Acquired Immune Deficiency Syndromes*. 2003; 33(Suppl.2):S115–S121. [PubMed: 12853860]
- Embretson SE. Item response theory models and spurious interaction effects in factorial ANOVA designs. *Applied Psychological Measurement*. 1996; 20:201–212.
- Fowler, FJ. *Health survey research methods conference proceedings*. Bethesda, MD: U.S. Department of Health and Human Services; 1989.
- Golden MR, Brewer DD, Kurth A, Holmes KK, Handsfield HH. Importance of sex partner HIV status in HIV risk assessment among men who have sex with men. *Journal of Acquired Immune Deficiency Syndromes*. 2004; 36:734–742. [PubMed: 15167293]
- Greenwood GL, Relf MV, Huang B, Pollack LM, Canchola JA, Catania JA. Battering victimization among a probability-based sample of men who have sex with men. *American Journal of Public Health*. 2002; 92:1964–1969. [PubMed: 12453817]
- Greenwood GL, White EW, Page-Shafer K, Bein E, Osmond DH, Paul J, et al. Correlates of heavy substance use among gay and bisexual men: The San Francisco Young Men's Health Study. *Drug and Alcohol Dependence*. 2001; 61:105–112. [PubMed: 11137274]
- Hays RB, Kegeles SM, Coates TJ. Unprotected sex and HIV risk taking among young gay men within boyfriend relationships. *AIDS Education and Prevention*. 1997; 9:314–329. [PubMed: 9376206]
- Jones R, Candlin C. Constructing risk across timescales and trajectories: gay men's stories of sexual encounters. *Health, Risk & Society*. 2003; 5:199–213.

- Kalton, G. Sampling considerations in research on HIV risk and illness. In: Ostrow, DG.; Kessler, RC., editors. *Methodological issues in AIDS behavioral research*. New York: Plenum Press; 1993. p. 53-74.
- Klitzman RL, Greenberg JD, Pollack LM, Dolezal C. MDMA ('ecstasy') use, and its association with high risk behaviors, mental health, and other factors among gay/bisexual men in New York City. *Drug and Alcohol Dependence*. 2002; 66:115-125. [PubMed: 11906799]
- Koblin BA, Chesney MA, Husnik MJ, Bozeman S, Celum CL, Buchbinder S, et al. High-risk behaviors among men who have sex with men in 6 US cities: Baseline data from the Explore study. *American Journal of Public Health*. 2003; 93:926-932. [PubMed: 12773357]
- Lavrakas, PJ. *Telephone survey methods: Sampling, selection, and supervision*. Newbury Park, CA: Sage Publications; 1987.
- Linacre, JM. *A user's guide to WINSTEPS/MINISTEP Raschmodel computer programs Version 3.55*. Chicago, IL: MESA Press; 2005.
- Linacre JM, Wright BD. Chi-square fit statistics. *Rasch. Measurement Transactions*. 1994; 8:350.
- Mansergh G, Colfax GN, Marks G, Rader M, Guzman R, Buchbinder S. The Circuit party men's health survey: Findings and implications for gay and bisexual men. *American Journal of Public Health*. 2001; 91:953-958. [PubMed: 11392940]
- McKirnan DJ, Venable PA, Ostrow DG, Hope B. Expectancies of sexual "escape" and sexual risk among drug and alcohol-involved gay and bisexual men. *Journal of Substance Abuse*. 2001; 13:137-154. [PubMed: 11547615]
- McLean J, Boulton M, Brookes M, Lakhani D, Fitzpatrick R, Dawson J, et al. Regular partners and risky behavior—why do gay men have unprotected intercourse. *AIDS Care*. 1994; 6:331-341. [PubMed: 7948089]
- Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. *Archives of Physical Medicine and Rehabilitation*. 1989; 70:308-312. [PubMed: 2535599]
- Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*. 1995; 50:741-749.
- Middelthon AL, Aggleton P. Reflection and dialogue for HIV prevention among young gay men. *AIDS Care*. 2001; 3:515-526. [PubMed: 11454272]
- Mills TC, Paul J, Stall R, Pollack L, Canchola J, Chang YJ, et al. Distress and depression in men who have sex with men: The Urban Men's Health Study. *American Journal of Psychiatry*. 2004; 161:278-285. [PubMed: 14754777] *American Journal of Psychiatry*. 2004; 161:776. Erratum in.
- Mills TC, Stall R, Pollack L, Paul JP, Binson D, Canchola J, et al. Health-related characteristics of men who have sex with men: A comparison of those living in "gay ghettos" with those living elsewhere. *American Journal of Public Health*. 2001; 91:980-983. [PubMed: 11392945]
- Moss R, Osmond D, Bacchetti P, Chermann JC, Barre-Sinoussi F, Carlson J. Risk factors for AIDS and HIV seropositivity in homosexual men. *American Journal of Epidemiology*. 1987; 125:1035-1047. [PubMed: 3646828]
- Osmond DH, Catania J, Pollack L, Canchola J, Jaffe D, MacKellar D, et al. Obtaining HIV test results with a home collection test kit in a community telephone sample. *Journal of Acquired Immune Deficiency Syndromes*. 2000; 24:363-368. [PubMed: 11015153]
- Osmond DH, Pollack LM, Paul JP, Catania JA. Changes in prevalence of HIV infection and sexual risk behavior in men who have sex with men in San Francisco: 1997-2002. *American Journal of Public Health*. 2007; 97:1677-1683. [PubMed: 17463390]
- Parsons JT, Severino J, Nanin J, Punzalan JC, Von Sternberg K, Missildine W, et al. Positive, negative, unknown: Assumptions of HIV status among HIV-positive men who have sex with men. *AIDS Education and Prevention*. 2006; 18:139-149. [PubMed: 16649959]
- Paul JP, Catania J, Pollack L, Moskowitz J, Canchola J, Mills T, et al. Suicide attempts among gay and bisexual men: Lifetime prevalence and antecedents. *American Journal of Public Health*. 2002; 92:338-1345.
- Paul JP, Catania J, Pollack L, Stall R. Understanding childhood sexual abuse as a predictor of sexual risk-taking among men who have sex with men: The Urban Men's Health Study. *Child Abuse and Neglect*. 2001; 25:557-584. [PubMed: 11370726]

- Paul JP, Stall RD, Crosby GM, Barrett DC, Midanik LT. Correlates of sexual risk-taking among gay male substance abusers. *Addiction*. 1994; 89:971–983. [PubMed: 7950855]
- Rasch, G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1980. Original work published 1960
- Relf MV, Huang B, Campbell J, Catania J. Gay identity, violence victimization, and HIV risk behaviors: Empirically testing theoretical relationships among a probability based sample of urban MSM. *Journal of the Association of Nurses in AIDS Care*. 2004; 15:14–26. [PubMed: 15090130]
- Romanoski J, Douglas G. Rasch-transformed raw scores and two-way ANOVA: A simulation analysis. *Journal of Applied Measurement*. 2002; 3:421–430. [PubMed: 12486309]
- Shen L. Quantifying item dependency. *Rasch Measurement Transactions*. 1996; 10:485.
- Shen, L. Quantifying item dependency by Fisher's Z; Paper presented at the annual meeting of the American Educational Research Association; Chicago. 1997 Mar.
- Shen L, Yen J. Item dependency in medical licensing examinations. *Academic Medicine*. 1997; 22(10) (Suppl. 1):S19–S21. [PubMed: 9347727]
- Sireci SG, Thissen D, Wainer H. On the reliability of testlet-based tests. *Journal of Educational Measurement*. 1991; 28:237–247.
- Smith EV Jr. Evidence for the reliability of measures and the validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement*. 2001; 2:281–311. [PubMed: 12011511]
- Smith, EV, Jr. Detecting and evaluating the impact of multi-dimensionality using item fit statistics and principal component analysis of residuals. In: Smith, EV., Jr; Smith, RM., editors. *Introduction to Rasch measurement*. Maple Grove, MN: JAM Press; 2004a. p. 575-600.
- Smith, EV, Jr. Metric development and score reporting in Rasch measurement. In: Smith, EV., Jr; Smith, RM., editors. *Introduction to Rasch measurement*. Maple Grove, MN: JAM Press; 2004b. p. 342-365.
- Smith EV Jr. Effect of item redundancy on Rasch item and person estimates. *Journal of Applied Measurement*. 2005; 6:147–163. [PubMed: 15795483]
- Smith EV Jr, Kulikowich JM. An application of generalizability theory and many-facet Rasch measurement using a complex problem solving skills assessment. *Educational and Psychological Measurement*. 2004; 64:617–639.
- Smith, EV., Jr; Smith, RM., editors. *Introduction to Rasch measurement: Theory, models, and applications*. Maple Grove, MN: JAM Press; 2004.
- Smith, EV., Jr; Smith, RM., editors. *Rasch measurement: Advanced and specialized applications*. Maple Grove, MN: JAM Press; 2007.
- Smith, RM. Fit analysis in latent trait measurement models. In: Smith, EV., Jr; Smith, RM., editors. *Introduction to Rasch measurement: Theory, models and applications*. Maple Grove, MN: JAM Press; 2004. p. 73-92.
- Stall R, Mills TC, Williamson J, Hart T, Greenwood G, Paul J, et al. Co-occurring psychosocial health problems and increased vulnerability to HIV/AIDS among urban men who have sex with men. *American Journal of Public Health*. 2003; 93:939–942. [PubMed: 12773359]
- Stall R, Paul JP, Greenwood G, Pollack LM, Bein E, Crosby GM, et al. Alcohol use, drug use, and alcohol-related problems among men who have sex with men: The Urban Men's Health Study. *Addiction*. 2001a; 96:1589–1601. [PubMed: 11784456]
- Stall R, Pollack L, Mills TC, Martin JN, Osmond D, Paul J, et al. Use of antiretroviral therapies among HIV-infected men who have sex with men: A household-based sample of four major American cities. *American Journal of Public Health*. 2001b; 91:767–773. [PubMed: 11344885]
- Theodore PS, Duran REF, Antoni MH, Fernandez MI. Intimacy and sexual behavior among HIV-positive men-who-have-sex-with-men in primary relationships. *AIDS and Behavior*. 2004; 8:321–331. [PubMed: 15475679]
- Thissen D, Steinberg L, Mooney J. Trace lines for testlets: A use of multiple-categorical response models. *Journal of Educational Measurement*. 1989; 26:247–260.
- Tourangeau, R.; Rips, LJ.; Rasinksi, K. *The psychology of survey response*. New York: Cambridge University Press; 2000.

- U.S. Bureau of the Census. Census of population and housing, 1990: Public use microdata samples
U.S. Washington, DC: United States Department of Commerce; 1997.
- Vincke J, Bolton R, De Vleeschouwer P. The cognitive structure of the domain of safe and unsafe gave sexual behavior in Belgium. *AIDS Care*. 2001; 13:57–70. [PubMed: 11177465]
- Wilson EB, Hilferty MM. The distribution of chi-square. *Proceedings of the National Academy of Sciences of the United States of America*. 1931; 17:684–688. [PubMed: 16577411]
- Winkelstein W, Lyman DM, Padian N, Grant R, Samuel M, Wiley JA, et al. Sexual practices and risk of infection by the human immunodeficiency virus. The San Francisco Men's Health Study. *Journal of the American Medical Association*. 1987; 257:321–325. [PubMed: 3540327]
- Wolfe, EW.; Smith, EV, Jr. Instrument development tools and activities for measure validation using Rasch models: Part I-Instrument development tools. In: Smith, EV., Jr; Smith, RM., editors. *Rasch measurement: Advanced and specialized applications*. Maple Grove, MN: JAM Press; 2007a. p. 202-242.
- Wolfe, EW.; Smith, EV, Jr. Instrument development tools and activities for measure validation using Rasch models: Part II-Validation activities. In: Smith, EV., Jr; Smith, R., editors. *Rasch measurement: Advanced and specialized applications*. Maple Grove, MN: JAM Press; 2007b. p. 243-290.
- Wolitski R. The emergence of barebacking among gay men in the United States: A public health perspective. *Journal of Gay and Lesbian Psychotherapy*. 2005; 9:13–38.
- Woodcock, RW. What can Rasch-based score convey about a person's test performance?. In: Embretson, SE.; Hershberger, SL., editors. *The new rules of measurement: What every psychologist and educator should know*. Mahwah, NJ: Lawrence Erlbaum Associates; 1999. p. 105-128.
- Woods WJ, Binson D, Pollack LM, Wohlfeiler D, Stall RD, Catania JA. Public policy regulating private and public space in gay bathhouses. *Journal of Acquired Immune Deficiency Syndromes*. 2003; 32:417–423. [PubMed: 12640200]
- Wright BD. Solving measurement problems with the Rasch model. *Journal of Educational Measurement*. 1977; 14:97–116.
- Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation*. 1989; 70:857–860. [PubMed: 2818162]
- Wright, BD.; Masters, GN. *Rating scale analysis: Rasch measurement*. Chicago: MESA Press; 1982.
- Wright, BD.; Mead, RJ.; Ludlow, LH. *Kidmap: Research memorandum number 29*. Chicago: MESA Press; 1980.
- Wright, BD.; Stone, MH. *Best test design: Rasch measurement*. Chicago: MESA Press; 1979.

Appendix

Original 18 items on the HIV Transmission Risk Scale.

ID	Item content
mrrolwd	1. Received oral sex without a condom, with withdrawal before ejaculation
mrrolnc	2. Received oral sex where he ejaculated in your mouth, without a condom
mporlwd	3. Performed oral sex without a condom, with withdrawal before ejaculation
mporlnc	4. Performed oral sex where you ejaculated in his mouth, without a condom
mranlwd	5. Had receptive anal intercourse without a condom, with withdrawal before ejaculation
mranlnc	6. Had receptive anal intercourse without a condom, with ejaculation inside you
mpanlwd	7. Performed insertive anal intercourse without a condom, with withdrawal before you ejaculated
mpanlnc	8. Performed insertive anal intercourse without a condom, with ejaculation inside him
mrmmng	9. Engaged in rimming or been rimmed
mfistng	10. Engaged in fisting or been fisted

ID	Item content
msxtoys	11. Used sex toys, such as dildos or vibrators
munpsec	12. Unprotected anal intercourse with a secondary partner
sexpblc	13. Sex in a public place in past year
pblcrai	14. Unprotected receptive anal intercourse in a public place
pblciai	15. Unprotected insertive anal intercourse in a public place
sexgrp	16. Had group sex in past year
grprai	17. Unprotected receptive anal intercourse during group sex
grpiaai	18. Unprotected insertive anal intercourse during group sex

Fig. 1.

Expected score maps for misfitting persons. *Note:* Items 13 and 14 are the recoded versions of items 13–15 and 16–18, respectively, unprotected anal intercourse in a public place (uaipp) and unprotected anal intercourse in a group situation (uaigs). Otherwise, item numbers and abbreviations correspond to those in the Appendix

Fig. 2.

Person/item variable map. *Note:* Each '#' is 6 persons and each '.' is one person. M, S, and T represent the mean, 1 *SD*, and 2 *SD*, respectively. Items 13 and 14 are the recoded versions of items 13–15 and 16–18, respectively, unprotected anal intercourse in a public place (uaipp), and unprotected anal sex in a group situation (uaigs). Otherwise, item numbers and abbreviations correspond to those in the Appendix

Table 1

Background characteristics

	<i>N</i>	%
Sex with men past year		
Yes	352	86
No	55	14
Have a primary partner		
Yes	189	46
No	218	54
HIV status		
Positive	56	15
Negative	303	79
Never tested	26	7
Education		
High school or less	23	6
Some college	72	18
College degree	175	43
Postgraduate	135	33
Race/Ethnicity		
White	323	79
Black	31	8
Hispanic	27	7
Other	22	5
Age		
18–29	98	24
30–39	169	42
40–49	94	23
50+	46	11

Table 2Means and *SD* for levels of risk behavior^a

Independent variable	<i>N</i>	<i>M</i>	<i>SD</i>
Have a primary partner **			
Yes	178	-1.22	2.41
No	173	-2.04	2.35
HIV status *			
Tested positive	51	-1.01	2.68
Tested negative/not tested	284	-1.75	2.35
Race/Ethnicity ***			
White	277	-1.57	2.37
Black	27	-2.34	2.24
Hispanic	27	-0.88	2.61
Age			
Less than 29	86	-1.27	2.09
30–39	161	-1.73	2.42
40–49	75	-1.63	2.69
50+	30	-2.17	2.42

*
 $p < .05$ **
 $p < .001$ ***
 $p < .10$

^aMean values were constructed by counting the number of items (out of 14) endorsed by each respondent and converting the total scores via the Rasch scaling program, Winsteps. These analyses were limited to the 352 men who were same gender sexually active during the past year

Table 3

Raw score to logit conversion table

Score	Measure	SE
0	-7.36	1.96
1	-5.69	1.35
2	-3.80	1.41
3	-2.18	1.10
4	-1.22	0.88
5	-0.55	0.77
6	-0.01	0.71
7	0.48	0.68
8	0.94	0.68
9	1.41	0.69
10	1.91	0.72
11	2.47	0.77
12	3.14	0.87
13	4.09	1.11
14	5.46	1.88