Original article

Reliability of the 2005 Middle School Youth Risk Behavior Survey

Keith J. Zullig, M.S.P.H., Ph.D.[a,*], Scott Pun, M.S.[a], Jon M. Patton, Ph.D.[b],
and Valerie A. Ubbes, Ph.D., C.H.E.S.[a]

[a]Program in Health Promotion, Miami University, Oxford, Ohio
[b]Information Technology Services, Miami University, Oxford, Ohio

Abstract

Purpose: To evaluate the reliability of the middle school version of the 2005 Youth Risk Behavior Survey (MSYRBS) questionnaire.

Methods: A convenience sample of 232 Midwestern seventh and eighth grade middle school students completed the MSYRBS questionnaire twice in a 2-week period (14 days apart). The MSYRBS questionnaire, which queries a variety of health risk behaviors, was administered in a manner that preserved anonymity but allowed Time 1 and Time 2 matching. This was accomplished by using two questionnaire scantrons coded with the same unique number, and destroying all used materials to ensure that each participant was matched with their code. Kappa statistics were calculated for individual questions and group characteristics using SAS.

Results: The mean kappa was 62.6% and the median was 66.5%. Kappa statistics for each item ranged from −2.4% (injection drug use) to 83.8% (suicide contemplation). Negative kappa values were found for two items that had extremely small cell sizes. Kappas did not differ by gender, grade, or race. Based on nonoverlapping confidence intervals, there were no items that had significantly different prevalence estimates at Time 1 vs. Time 2. Nine items (24.3%) and one category (alcohol-drugs) had kappas below 61.0%.

Conclusions: This preliminary study suggests that the reliability of the MSYRBS is high over time. A number of items should be further examined to determine whether they should be amended or omitted from future versions of the MSYRBS. Further research with larger and more diverse samples is recommended, potentially involving children as content experts. © 2006 Society for Adolescent Medicine. All rights reserved.

Keywords: Middle school adolescents; Youth Risk Behavior Survey; Psychometric analysis

The Youth Risk Behavior Surveillance System (YRBSS) is a systematic surveillance of youth behaviors that contribute to the leading causes of adolescent and emerging adult morbidity and mortality in the United States. These categories consist of the following: 1) behaviors that contribute to intentional and unintentional injuries, 2) tobacco use, 3) alcohol and other drug use, 4) sexual behaviors that contribute to unintended pregnancy and sexually transmitted diseases including HIV infection, 5) unhealthy dietary be-

*Address correspondence to: Dr. Keith J. Zullig, Assistant Professor, Health Education, Physical Education, Health, & Sport Studies, 106 Phillips Hall, Miami University, Oxford, OH 45056.
E-mail address: zulligkj@muohio.edu

haviors, and 6) physical activity [1]. College, high school, alternative school, and middle school versions of the YRBS are available to states and local school districts to use for needs assessment, program evaluation, and policy decisions. The high school version of the YRBS is an 84-item questionnaire developed and first piloted by the U.S. Centers for Disease Control and Prevention, Division of Adolescent and School Health in 1990 (CDC-DASH) [2].

National- and state-level data of representative samples of youth are collected on a biannual basis with the high school YRBS; these data are used to develop and track the effectiveness of health promoting policies and programs to help reduce the aforementioned youth risk behaviors [3]. Owing to the importance of these data, and to increase the

confidence of these data, the high school version of the YRBS has undergone three separate test-retest reliability studies as items were modified over time among students in grades 9 through 12 [4–6]. These studies consistently demonstrated the reliability of the high school version of the YRBS. Findings were consistent by gender, grade, and race/ethnicity, but the kappa values in the Brener et al study [5] were lower than in the other studies, according to the criteria set forth by Landis and Koch [7]. Overall, these studies determined that questionnaire items related to tobacco use, alcohol and other drug use, and sexual behavior display greater reliability than items related to dietary behaviors, physical activity, and other health-related topics.

The middle school version of the YRBS (MSYRBS) was implemented for the first time in 1995 for use by states and school districts, but has never been tested for its psychometric properties. The MSYRBS is a 49-item version of the YRBS, except items are written at what is believed to be an appropriate reading level for students in middle school. Previous YRBS research [4] suggests that students in seventh grade responded less consistently than those in higher grades. However, those researchers tested a version of the YRBS designed for students in grades 9 through 12, which may have biased their findings. Therefore, the purpose of this study was to test the reliability of an age-appropriate version of the YRBS (MSYRBS) among students in grades seven and eight.

Methods

Sample

During the spring of 2005, a convenience sample of 232 middle school students in grades seven and eight were selected from two public school districts in this Midwestern state. Schools were selected to participate in the study based on previous work with the authors in curriculum deliberation and program evaluation. Each school was given $250 to assist with the recruitment of students for this study. Of the 402 students selected to participate in the study, 249 (62%) completed the questionnaire during the first survey administration. Students were not included in the analysis if they did not complete surveys during both survey administrations (Time 1 and Time 2) (n = 17, 6.8%). Classes that met during second period were eligible for classroom-level sampling selection to maximize student eligibility. As can be seen from Table 1, our demographics were similar to national distributions in gender and age, but not for race or ethnicity or grade [8]. Specifically, seventh grade students were overrepresented, as were students who reported themselves of "white" race or ethnicity.

Questionnaire description

The MSYRBS is self-administered, and the standard questionnaire contains 49 multiple-choice items. However,

Table 1
Demographic characteristics of respondents those researchers tested and of students in grades 7–8

Characteristic	Sample distribution (%)	National distribution (%)
Gender		
Male	45.3	51.2%
Female	54.7	48.8%
Grade		
7	80.9	50.3%
8	19.1	49.7%
Race		
White	93.5	76.7%
Other	6.5	23.3%
Age (years)		
<11	1.7	3.1%
12 to 13	90.9	80.9%
>14	7.4	16.0%

our two participating school districts would not permit data collection until the four items querying about students' sexual behavior were deleted from the questionnaire. In addition, one item asking students about recent cigar use was also omitted. Thus, the remaining 44 items were included in the questionnaire used for this study. Four items request demographic information, two items request students to report their height and weight, and the remaining items query students regarding health risk behaviors. As with the high school version of the YRBS, no skip patterns are included on the MSYRBS, allowing some degree of privacy because equivalent amounts of time are needed to complete the survey despite the number of risk behaviors a student engages in and because individual students cannot be identified based on skip patterns.

Data collection procedures

All data collection procedures were approved by the referent university's review board for the rights of human subjects in research. Following similar methods from previous test-retest reliability studies described elsewhere [4,5,9], two questionnaire scantrons were coded with the same unique number ranging from 1 to 250. Each pair of identically numbered scantrons were then placed in a single large envelope along with the MSYRBS instrument and distributed to each student participant during the first survey administration (Time 1). Each student then removed and used one scantron. The envelope, which contained only the second identically numbered scantron, was then sealed and signed across the seal by the student. When survey administrators returned for the second survey administration (Time 2, 14 days later), each student received the envelope with his or her name signed across the seal, removed the second scantron, and destroyed the envelope.

Parent-notification forms were distributed at least 7 days in advance of survey administration; those parents who wanted their children to participate were required to sign and return the form (active consent). The survey was ad-

ministered in each school's auditorium or cafeteria by trained data collectors, emphasizing anonymity, privacy, and confidentiality. During the first survey administration, data collectors reminded students that they would be returning to the school in 2 weeks to ask students to complete a similar questionnaire.

Data analysis

Collapsing and editing protocols. Scantrons that contained fewer than 20 valid responses or contained the same response option 15 or more times were discarded according to standard CDC procedures [5]. Variables were collapsed into "risk" and "at risk" categories based on their responses, according to standard YRBSS practice [5]. For instance, students who reported smoking 0 days out of the past 30 days were coded as "no risk." Alternatively, students who reported smoking on 1 or more days during the past 30 days were collapsed and coded as "at risk." Items querying about "the last time" and "in the past 7 days" were excluded from these analyses because response options are not consistent for a two-week reference period.

Kappa statistic and prevalence rates. A kappa statistic was calculated for each of the questionnaire items. A kappa statistic provides a measure of agreement that corrects for what would be expected by chance. Prevalence rates of risky behaviors, calculated at Time 1 and Time 2, were also computed and were considered significantly different if their 95% confidence intervals did not overlap. This criteria is consistent with standard CDC practice for assessing statistical significance of subgroup differences in standard YRBSS data reports [10].

Results

Data were analyzed similarly to Brener et al [5] in the examination of the reliability of the 1999 YRBS questionnaire on high school students. The mean kappa for all items was 62.6%, and the median kappa was 66.5%. Kappas ranged from −2.4% to 83.8% (Table 2). Qualitative values were assigned based on groupings described by Landis and Koch [7]. Results suggest that 75.7% of items can be described as having at least "substantial" reliability (kappa ≥ 61%), and 89.2% of items had at least "moderate" reliability (kappa ≥ 41%). Two items (has used steroids and has ever injected illegal drugs) had kappas ≤ 0 and can be described as having "poor" reliability. After examining the raw data and the cross tabulation tables, it was found that all students who reported using steroids or illegal drugs at Time 1 did not report using these at Time 2. Likewise, students who reported using these substances at Time 2 did not report using them at Time 1. In addition, nine items (24.3%) had kappas ≤ 61%.

Examination of reliability for all grouping characteristics suggest no significant differences for any groups (Table 3),

which is similar to what has been observed among the high school version of the YRBS [5]. In comparing respondent characteristics, females exhibited slightly higher mean kappas than males, but this difference was not significant. Similarly, items that used lifetime as a reference period had mean kappas somewhat higher than those items that used past 30 days, but again, were not significantly different.

Reliability among the risk behavior categories revealed one significant difference. All risk behavior categories had consistent mean kappas (60.4%–72.4%) with the exception of alcohol and drugs (mean kappa = 43.1%). Alcohol and drugs items were likely substantially reduced by the results of two items pertaining to steroid and injection drug use, which had kappas of −2.2% and −2.4%, respectively (Table 2).

Discussion

Approximately 90% of the MSYRBS items had at least "moderate" reliability and nearly 76% of the items had "substantial" reliability in this preliminary study. In addition, no items were determined to display significantly different Time 1 vs. Time 2 prevalence estimates. However, items pertaining to cocaine use, steroid use, and injection drug use displayed poor reliability. A likely explanation for the negative kappa values for the steroid and injection drug items is the inconsistency due to all those who reported using steroids or drugs during Time 1, but reported not using them at Time 2 and vice versa. With a larger sample, these data inconsistencies would likely not be problematic. All items displaying poor reliability, but especially these three items, should be further examined with larger and more diverse samples as a result and potentially revised or deleted in future versions of the MSYRBS.

The results from this preliminary study compare favorably to other reliability studies of the YRBS with students in grades 9–12 [4,5]. For example, Brener et al [5] reported the YRBS to display overall mean and median kappa values of 60.7% and 60.0%, respectively, compared with 62.6% and 66.5%, respectively, in this study. Results from this study also suggest that the appropriately worded items lead to greater consistency of responses among middle school students, which is an improvement over prior results among this age group [4]. In addition, the credence of the findings reported here is bolstered by the demographic item findings and their subsequent reliability. That is, no differences were detected in kappa values between race, gender, or grade, which is similar to findings among high school YRBS reliability studies [4–6].

Interestingly, unlike prior YRBS reliability studies with high school students [4,5], items related to physical activity, dietary behaviors, and other health-related topics demonstrated substantial reliability among this sample of middle school students when using the MSYRBS. This may be because several item response options are worded as 'yes' or 'no' on the MSYRBS. For example, students on the

Table 2
Kappa statistics, Time 1, and Time 2 prevalence rates, by questionnaire item

Item	Kappas	Time 1	Time 2
Behaviors related to unintentional injuries and violence			
Rarely or never wear seatbelts in a car	69.2	9.6	10.9
Rarely or never where a helmet when riding a bicycle	65.1	62.3	61.5
Rarely or never wear a helmet when rollerblading or riding a skateboard	75.1	43.3	40.7
Has ridden in a car driven by someone who had been drinking alcohol	77.9	38.6	39.5
Has carried a weapon, such as a gun, knife, or club	72.3	41.1	46.8
Has been in a physical fight	66.2	53.3	51.1
Has been injured in a physical fight that required treatment from a doctor or nurse	39.7	5.6	8.2
Has ever considered suicide	83.8	23.4	21.2
Has ever planned suicide	74.7	12.5	15.1
Has ever attempted suicide	70.6	6.9	7.3
Tobacco use behavior			
Ever used cigarettes	75.4	25.0	23.7
Age of first cigarette < 14 years old	70.3	13.2	16.2
Smoked > 1 day in past 30 days	81.0	6.9	7.8
Smoke > 20 cigarettes per day on the days smoke during the past 30 days	66.5	.4	.9
Received cigarettes in past 30 days	79.4	7.8	8.2
Has smoked at least one cigarette every day for 30 days/regularly	50.1	8.6	7.3
Has used smokeless tobacco	45.9	9.1	6.5
Alcohol and other drug use behaviors			
Has used alcohol	70.0	36.5	36.5
Age of first alcohol use < 14 years	68.5	34.1	34.1
Has used marijuana	65.2	9.7	8.3
Age of first marijuana use < 14 years	56.4	10.0	8.7
Has used cocaine	23.4	4.8	4.8
Has used inhalants	66.0	13.8	14.2
Has used steroids	−2.2	2.2	2.2
Ever injected illegal drugs	−2.4	1.7	3.5
Dietary behaviors			
Perceive self as overweight	76.6	30.6	30.2
Trying to lose weight	77.3	37.5	39.7
Exercised to lose or keep from gaining weight	70.3	63.6	57.1
Ate less food, calories, or fat to lose or keep from gaining weight	66.1	39.4	37.2
Fasted to lose or keep from gaining weight	51.6	12.9	13.8
Took diet pills, powders, or liquids to lose or keep from gaining weight	62.0	5.6	6.5
Vomited or took laxatives to lose or keep from gaining weight	64.3	6.1	7
Physical activity			
Watch ≤2 hours of television on an average school day	65.6	67.2	67.7
Attends physical education class > 1 day a week	75.9	55.3	54.0
Plays on a sports team	75.7	67.3	66.4
Other health-related topics			
Has ever been taught about AIDS or HIV in school	59.0	84.0	84.6
Describes health as fair or poor	61.8	8.6	7.3

MSYRBS are asked "Do you play on any sports teams?" with responses as either 'yes' or 'no'. The YRBS asks students, "During the past 12 months, on how many sports teams did you play?" with responses ranging from '0 teams' to '3 or more'. The increased kappa values observed for similar items on the MSYRBS may be a result of these wording differences. However, an examination of YRBS and MSYRBS items for other categories does not lend itself to simplistic interpretations. For instance, items for other health-related topics are nearly identical between the YRBS and MSYRBS, yet demonstrated significantly higher reliability among this sample when compared with prior YRBS reliability studies with high school students [4,5]. Future

research is needed before robust conclusions can be drawn from these preliminary findings.

Limitations

Inconsistent responses from Time 1 to Time 2 in this preliminary study were considered to be response errors, as is the case in other reviewed test-retest reliability studies [4–6]. As previously posited by Brener et al [5], these inconsistencies may reflect the possibility of an actual behavior change from Time 1 to Time 2 in this study, making these values of kappa conservative in nature. For instance, a student who did not report being in a physical fight that needed to be treated by a doctor or nurse at Time 1 may

Table 3
Mean kappa statistics and 95% confidence intervals by demographic and question characteristics

Characteristics	Mean kappa %	95% CI
Gender		
Male	55.3	47.8, 62.7
Female	65.9	58.9, 73.0
Grade		
7	62.4	55.9, 69.0
8	63.3	54.2, 72.3
Race or ethnicity		
White	62.0	55.3, 68.6
Other	59.4	49.3, 69.5
Reference period		
Past 30 days	68.2	42.4, 94.0
Lifetime	61.9	54.7, 69.1
Risk behavior categories		
Unintentional injuries and violence	69.5	61.1, 78.0
Tobacco use	66.9	54.1, 79.8
Alcohol – drugs	43.1	16.6, 69.7
Dietary behavior	66.9	58.6, 75.1
Physical activity	72.4	57.8, 87.0
Other health-related topics	60.4	42.5, 78.4

actually have been in a fight at Time 2. Table 2 indicates such an increase for this item from Time 1 to Time 2, and is subsequently low in reliability. However, this inconsistent response may actually be accurate if a student was in a fight that required treatment during the 2-week test-retest time-frame from Time 1 to Time 2.

Second, reliability is a necessary prerequisite to validity, but reliability alone cannot be construed as validity for the MSYRBS instrument. Although some research has established the validity of adolescent self-reports across a variety of behaviors measured by the MSYRBS [11–15], additional research should be conducted to establish the validity of self-reports among middle school adolescents. Third, because several items (four sexual behavior and one tobacco) on the standard MSYRBS were deleted from these analyses, the reliability of these items is unknown for middle school students. Fourth, additional research needs to be conducted to replicate the findings reported in this study before full confidence in the reliability of the MSYRBS can be determined, especially given that our sample is not consistent to the national distribution in race or grade. Future research should consider the use of a larger and more diverse sample, including the use of computer technology to further this preliminary research, and potentially involving children as content experts [16].

References

[1] Kolbe LJ. An epidemiological surveillance system to monitor the prevalence of youth behaviors that most affect health. J Health Educ 1990;21:44–8.

[2] Kolbe LJ, Kann K, Collins JL. Overview of the Youth Risk Behavior Surveillance System. Public Health Rep 1993;108:2–10 (suppl 1).

[3] Sussman MP, Everett SA, Wilson TW, Kann L. The Youth Risk Behavior Surveillance System: updating policy and program applications. J Sch Health 2002;72:13–7.

[4] Brener ND, Collins JL, Kann L, et al. Reliability of the Youth Risk Behavior Survey questionnaire. Am J Epidemiol 1995;41:575–80.

[5] Brener ND, Kann L, McManus T, et al. Reliability of the Youth Risk Behavior Survey questionnaire. J Adolesc Health 2002;31:336–42.

[6] Klein JD, Graff CA, Sanitelli JS, et al. Improving adolescent health care surveillance. In: Cynamon MA, Kulka RA, eds. Seventh Conference on Health Survey Research Methods (DHHS Pub No. 01-1013). Hyattsville, MD: National Center for Health Statistics, 2001: 11–8.

[7] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.

[8] U.S. Census Bureau. Scholars of all ages: school enrollment. In: Current Population Reports, Series P23-205. Population Profile of the United States: 1999. Washington, DC: U.S. Government Printing Office, 2001:35.

[9] Popham WJ. Appraising two techniques for increasing the honesty of students' answers to self-report assessment devices. J Pers Eval Educ 1993;7:33–41.

[10] Eaton DK, Kann L, Kinchen S, et al. Youth Risk Behavior Surveillance—United States, 2005. MMWR Surveill Summ 2006; 55:1–108.

[11] Treuth MS, Hou N, Young DR, et al. Validity and reliability of the Fels physical activity questionnaire for children. Med Sci Sports Exerc 2005;37(3):488–95.

[12] Winters KC, Stinchfield RD, Henly GA, et al. Validity of adolescent self-report of alcohol and other drug involvement. Int J Addict 1990–1991;25:1379–95.

[13] Wills TA, Cleary SD. The validity of self-reports of smoking: analyses by race/ethnicity in a school sample of urban adolescents. Am J Pub Health 1997;87:56–61.

[14] De Man AF, Leduc CP. Validity and reliability of a self-report suicide ideation scale for use with adolescents. Soc Behav Pers 1994;22: 261–6.

[15] Rockett HR, Breitenbach M, Frazier AL, et al. Validation of a youth/adolescent food frequency questionnaire. Prev Med 1997;26: 808–16.

[16] Stewart JL, Lynn MR, Mishel MH. Evaluating content validity for children's self-report instruments using children as content experts. Nurs Res 2005;54(6):414–8.