HHS Public Access

Author manuscript

Am J Health Promot. Author manuscript; available in PMC 2024 August 28.

Published in final edited form as:

Am J Health Promot. 2024 July ; 38(6): 843–851. doi:10.1177/08901171241239735.

Reliability of the 2021 National Youth Risk Behavior Survey Questionnaire

Sherry Everett Jones, PhD, MPH, JD[1], **Nancy D. Brener, PhD**[1], **Barbara Queen, MS**[2], **Molly Hershey-Arista, MA**[2], **William A. Harris, MM**[1], **Jonetta J. Mpofu, PhD, MPH**[1], **J. Michael Underwood, PhD**[1]

[1]Division of Adolescent and School Health, National Center for HIV, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA, USA

[2]Westat, Rockville, MD, USA

Abstract

Purpose: The Youth Risk Behavior Survey (YRBS) monitors behaviors, experiences, and conditions affecting the health of high school students nationwide. This study examined the test-retest reliability of the 2021 national YRBS questionnaire.

Design: Respondents completed a Time 1 and Time 2 paper-and-pencil questionnaire approximately 2 weeks apart during February to May 2022. Data were linked in such a way as to preserve anonymity.

Setting: Convenience sample of high schools.

Subjects: High school students (N = 588).

Measures: Health risk behaviors and experiences assessed on the 2021 national YRBS questionnaire.

Analysis: Time 1 and Time 2 responses were compared for each questionnaire item using the McNemar's test. Then, Cohen's kappa coefficients tested the agreement between Time 1 and Time 2 responses overall, and by sex, grade, and Black, White, and Hispanic race and ethnicity.

Results: Among the 74 items analyzed, 96% had at least moderate reliability, and 73% had substantial or almost perfect reliability. The mean Cohen's kappa was .68. McNemar's test findings showed Time 1 and Time 2 data significantly differed ($P < .01$) for 9 items (12%).

Conclusion: Reliable health behavior measures are important in the development of youth-focused public health programs and policies. Findings suggest the national YRBS questionnaire is

Corresponding Author: Sherry Everett Jones, PhD, MPH, JD, Division of Adolescent and School Health, National Center for HIV, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, 1600 Clifton Rd, NE, MS US8-1, Atlanta, GA 30329-4027, USA. sce2@cdc.gov.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Disclaimer

The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or Westat.

a reliable instrument. Such findings lend support to relying on adolescent self-reported data when monitoring health behaviors using the YRBS.

Keywords

youth risk behavior survey; reliability; test-retest study

Purpose

The Youth Risk Behavior Surveillance System (YRBSS) is a system of school-based surveys comprised of the national Youth Risk Behavior Survey (YRBS) conducted by the Centers for Disease Control and Prevention (CDC), and site YRBSs conducted by state, tribal, territorial, and local health and education agencies in those jurisdictions.[1,2] The system captures student demographic information (sex, sexual identity, race and ethnicity, and grade) and is designed to monitor behaviors, experiences, and conditions affecting the health of high school students, including health behaviors (sexual, injury and violence, bullying, diet and physical activity, and mental health, including suicide) and substance use behaviors (alcohol, tobacco and electronic vapor product use, and other drug use).[1,2]

The national YRBS has been conducted biennially, during odd-numbered years, among a cross-sectional, nationally representative sample of students in grades 9 through 12 since 1991. The YRBS data are used by health and education officials for a variety of purposes, such as describing risk behaviors, identifying key topics for professional development, planning and monitoring school-based programs, supporting health-related policies and legislation at the state and local level, and seeking funding.[3] Further, YRBS data are used to demonstrate relationships among demographic characteristics, social determinants of health (eg, housing instability, witnessing violence), and indicators of physical and mental health.[4] These types of analyses provide evidence to schools and communities regarding which groups of students may be at highest risk. Such evidence can then facilitate relevant policies and practices that address the needs of students as well as the environments in which they live, play, and go to school.[4] As such, it is important that the instrument generating the data is of high quality.

The last test-retest reliability study conducted by CDC examined the 1999 YRBS questionnaire.[5] A survey test-retest reliability study examines the extent to which respondents answer consistently across the administration of a survey given at 2 time periods.[6] This stability of survey results is important because it means that data are reproducible.[6] The last YRBS reliability study found that nearly all items on the questionnaire had at least moderate reliability and almost half had substantial reliability.[5] Other reliability studies of the high school YRBS, or subsets of YRBS questions, also have been conducted previously[7–9]; however, the YRBS questions change over time to reflect emerging health concerns among youth, and no up-to-date reliability study is currently available. The purpose of this study, therefore, is to update previous reliability studies by examining the test-retest reliability of the 2021 version of the questionnaire, which was the most current at the time of the study, with a particular interest in questions that had been added since the previous reliability study.

Methods

This YRBS test-retest reliability study employed the same survey administration procedures used in the 2021 national YRBS[1] to ensure that the data collected for this study are comparable to data collected during actual YRBS data collection. This study was approved by the Institutional Review Board at Westat, the contractor used for this study. Active parental permission, or opt-in approval for student participation, was used in 1 school as requested by that school's district. All other schools used passive parental permission, allowing parents to opt out of student participation. Students were not required to participate. They could decline participation or skip any question included on the questionnaire.

Sample

The study used a convenience sample aimed at creating a diverse and robust sample of student participants in grades 9–12, typical of communities across the U.S. Importantly, the focus of the sampling was to ensure a diverse sample of schools, not to ensure the sample of students and schools was nationally representative.

The sample included students attending schools from various U.S. census regions and metropolitan status areas (urban, suburban, rural), and with various enrollment sizes, socio-economic status levels, and minority enrollment levels.

Design

Although this study used a convenience sample, a four-stage sample design, involving the selection of (1) states, (2) districts within sampled states, (3) schools within sampled districts, and (4) classes within selected schools, was used to create a diverse sample. Such a strategy was an efficient means of identifying schools with varying characteristics and locations and built on experiences with Youth Risk Behavior Surveys conducted biennially in states and school districts nationwide.

The sampling frame was constructed from the National Center for Education Statistics' Common Core of Data (CCD, https://nces.ed.gov/ccd/) 2018–2019 school year and for 1 measure (the median household income at the zip code level) the Census Bureau's American Community Survey was used (https://www.census.gov/programs-surveys/acs). To select schools, the sampling frame was sorted by the following variables prior to sampling: census region (based on state), type of locale (urban, suburban, rural), grade level (created from grade span with 2 levels: junior/senior high school, high school), enrollment in grades 9 through 12, percentage of students eligible for free and reduced-price lunch, median household income, and percentage of minority students. Systematic equal probability sampling with a random start was used to select schools. The list of schools was reviewed and any schools that overlapped with the 2021 state, local, or national YRBS samples were removed to reduce burden on those schools. The final sample contained 84 schools from 31 districts and 21 states.

After school district-level notification and approval, school principals at each sampled school were sent a study packet, including a sample questionnaire, sample parent permission

forms, and a fact sheet. Each was contacted to obtain their agreement, identify a liaison, and obtain teacher information for 4 selected classes in the school (1 class from each of grades 9 through 12), and arrange dates for a Time 1 and Time 2 administration. Schools and teachers were given a token of appreciation for participating in both Time 1 and Time 2 survey administrations. Specifically, after a class completed the Time 1 survey, a $25 prepaid cash card was provided to and signed for by the teacher at the end of the Time 1 visit. An additional $25 cash card was distributed and signed for at the end of the Time 2 visit. After completion of the Time 2 survey administration, a $300 check made out to the school was provided to and signed for by the principal or other school administrator.

Data collection by trained survey administrators began in February 2022 and was completed in May 2022. Special care was taken to ensure student privacy. The data collector provided an envelope to each participating student, which included 2 computer-scannable questionnaire booklets with the same identification number later used to link the completed Time 1 and Time 2 questionnaires. Following the same procedures used in previous YRBS test-retest reliability studies,[5,7] students were instructed to remove 1 questionnaire booklet from the envelope to complete. They sealed the second, identical questionnaire booklet in the envelope, and printed their name across the envelope seal. During the Time 2 data collection, approximately 2 weeks later, the data collector provided each student with their sealed, signed envelope containing the blank Time 2 questionnaire booklet. Only students who participated in the Time 1 data collection were permitted to participate in the Time 2 data collection. After participating students removed the Time 2 questionnaire booklet from their sealed envelope, they destroyed the envelope. Of the 1059 students enrolled in the selected classes in the participating schools, 738 completed the YRBS questionnaire during the first survey administration (of which 735 had usable questionnaires) and 594 students completed the questionnaire during the second survey administration (of which 591 had usable questionnaires). Thus, 80% of students who completed a questionnaire at Time 1 also completed a questionnaire at Time 2 (N = 591). In 3 cases, a Time 1 or Time 2 questionnaire was not usable after the standard YRBS editing protocol was applied,[10] so the final number of usable matched pairs for Time 1 and Time 2 was 588.

Analysis

All variables were coded as dichotomous to match typical reporting of YRBS data. For example, students who responded that they rode with a driver who had been drinking alcohol on zero of the past 30 days were classified as “no risk,” whereas students who did so on 1 or more days during the past 30 days were classified as “at risk.” Questions that used a “last time” or “past 7 days” reference period were not included in this analysis because the test-retest time frame was 2 weeks.

As noted by Yu, a variety of methods can be used to examine test-retest reliability.[6] This study employed 2 analysis strategies on 76 questionnaire items, although for 2 items, a statistic could not be computed due to zero respondents answering in the affirmative (ie, at risk) for Time 1 or Time 2. The 2 statistical tests used herein were the McNemar’s test and Cohen’s kappa; however, Time 1 and Time 2 prevalence also are shown to give context to the statistics provided.

The McNemar's test is a nonparametric test for dependent samples. All *P*-values are provided, but because of the large number of paired tests (74 questions), *P*-values <.01 were considered statistically significant. Cohen's kappa coefficients were used to test the Time 1 and Time 2 agreement for each question, including 32 items that are new since the 1999 reliability study[5] (new items are noted in Table 2), and by sex, grade, and non-Hispanic Black, non-Hispanic White, and Hispanic race and ethnicity. The number of students in other racial groups was too small for stable estimates.

Cohen's kappa is often used in an assessment of interrater reliability[6]; however, Yu notes that this test can also be used to assess the agreement in a test-retest situation. The kappa coefficient can be used to assess the observed agreement between measures provided by the same respondents on 2 occasions, and when positive, indicates that the agreement is more than one would expect by chance.[6,11,12] The size of the coefficient reflects the strength of agreement.[6,11,12]

Use of Cohen's kappa in the current study is useful in that it allows for a comparison of reliability findings of the 2021 YRBS questionnaire and the 1999 YRBS questionnaire assessed in a previous study.[5] In addition, use of Cohen's kappa allows the findings to include mean kappa scores computed for groups of variables. In this study, findings are presented by topic and self-reported demographic characteristics. This study relied on the following labels to describe the degree of agreement: poor (<.00), slight (.00–.20), fair (.21–.40), moderate (.41–.60), substantial (.61–.80), and almost perfect (.81–1.00).[12] Landis and Koch note these "divisions are clearly arbitrary," but they "provide useful 'benchmarks'" for the level of agreement.[12, p165] In addition, these labels were used in the aforementioned 1999 reliability study,[5] and therefore, provide context in interpreting reliability in an updated YRBS questionnaire.

Results

The sample comprised 50.4% male students, 33.5% Hispanic students, and the following non-Hispanic racial groups: .4% American Indian or Alaska Native, 3.7% Asian, 12.4% Black, .4% Native Hawaiian or other Pacific Islander, 44.8% White, and 4.9% students of 2 or more races (Table 1). A plurality of responding students were in ninth grade (30.7%), with the remaining students in 10th (23.5%), 11th (24.2%), and 12th (21.4%) grade. The sex, race and ethnicity, and grade distribution of the sample was reasonably close to the national distribution[13,14]; however, not all census regions were represented. The 12 participating schools were in the Northeast (50%), Midwest (42%), and West (8%). One school in the South originally agreed to participate; however, the school did not follow the active permission protocol required by that school district making it impossible to survey students in that school. Most participating schools were in a city (67%); the remaining schools were in a suburb (8%), town (8%), or rural area (17%). The median household income was $54,274 (range was $26,708 to $85,385), the median percentage of the school enrollment who were minority students was 65.5% (range was 4.1% to 97.7%), and the median percentage of students eligible for free and reduced-price lunch was 65.4% (range was 33.0% to 94.8%).

For 12.2%, or 9 of the 74 variables included in the analysis, the aggregate Time 1 and Time 2 prevalence estimates were significantly different. Five of those 9 variables with significant prevalence differences between the aggregate Time 1 and Time 2 were among the new items.

Kappa coefficients ranged from .20 to .94 (Table 2), with a mean of .68 (Table 3), for the 74 variables. More specifically, zero variables exhibited poor reliability, 1 (1.4%) exhibited slight reliability, 2 (2.7%) exhibited fair reliability, 17 (23.0%) exhibited moderate reliability, 38 (51.4%) exhibited substantial reliability, and 16 (21.6%) exhibited almost perfect reliability. Among the 32 items that were new since the 1999 reliability study,[5] the kappa coefficients ranged from .45 to .88, with a mean of .67. Twenty-one of the 32 items exhibited substantial reliability.

An assessment of reliability across demographic characteristics of the respondents found the following mean kappa coefficients: female (.69) and male (.64) students; black (.62), Hispanic (.68), and white (.70) students; and ninth- (.67), 10^{th}- (.68), 11^{th}- (.70), and 12^{th}- (.65) grade students. Mean kappa coefficients varied by risk behavior categories. The mean kappa for each risk behavior category, from highest to lowest, were dietary behaviors (.85), sexual behaviors (.76), physical activity (.71), tobacco use behaviors (.68), indicators of mental health (.67), behaviors related to unintentional injuries and violence (.65), alcohol and other drug use behaviors (.64), and other health-related topics (.62).

Discussion

This study's findings suggest YRBS items are generally reliable, with findings comparable to those found in other studies.[5,7–9] For example, the reliability study of the 1999 YRBS found kappas ranging from .24 to .91, with an overall mean of .61; 93% of items had at least moderate reliability.[5] Likewise, this study of the national 2021 YRBS found that among the 74 items analyzed for test-retest reliability, 96% had at least moderate reliability, and 73% had substantial or almost perfect reliability based on labels proposed by Landis and Koch.[12] In a second approach to measuring test-retest reliability, using the McNemar's test, only 9 variables had prevalence estimates that varied significantly between Time 1 and Time 2. It is unclear why the prevalence for those questions changed. Perhaps behaviors or experiences actually changed, or perhaps after Time 1, students thought about the questions and upon reflection, remembered something about the timing or other element of the behavior or experience addressed in the questions. Also, the findings of this study suggest that YRBS items are generally reliable across sex, racial and ethnic, and grade groups, as well as across risk behavior categories, all of which demonstrated mean kappas above .61; that is, having at least substantial reliability. Likewise, of the 32 items that were added to the 2021 YRBS questionnaire since the previous study,[5] two-thirds demonstrated at least substantial reliability. Five had Time 1 and Time 2 significant differences based on the McNemar's test.

Very low or very high prevalence (ie, close to 0% or 100%) is known to negatively affect kappa coefficients.[15,16] In those cases, there is inadequate distribution in the concordant pairs of Time 1 and Time 2 responses. For example, in a low prevalence situation, nearly all responses can be found in a Time 1 "no" and Time 2 "no" response cell, whereas very few responses can be found in a Time 1 "yes" and Time 2 "yes" cell. In this study, 6 of

the 7 items with kappa coefficients below .50 had this very low or very high prevalence issue: (1) threatened or injured with a weapon on school property ≥1 time during the past 12 months (9.2% and 5.5%, respectively), (2) used prescription pain medicine without a doctor's prescription or differently than how a doctor told them to use it ≥ 1 time during the past 30 days (5.7% and 5.9%, respectively), (3) ever used methamphetamines (.2% and 1.2%, respectively), (4) ever injected any illegal drugs (.3% and 1.4%), (5) usually did not sleep in their parent's/guardian's home during the past 30 days (1.6% and 2.3%, respectively), and (6) speak English well or very well (98.8% and 98.1%, respectively). The only other variable with a kappa coefficient below .50 was tried to quit using all tobacco products during the past 12 months. This variable requires review to understand why the reliability was low.

Limitations

This study had several limitations. First, this study used a convenience sample and not a nationally representative sample as found in the biennial national YRBS.[1] For example, this study included an overrepresentation of urban schools and no schools from the South US census region. The findings generally show good reliability for YRBS items, but potential bias may exist if non-urban students or students in the South differ in their responses to a test-retest design. Second, the divisions and subsequent labels used to categorize the Cohen's kappa coefficients are not an exact science as noted by Landis and Koch.[12] Nevertheless, the labels provide a benchmark for summarizing the findings and allow for the comparison of this study to the test-retest study of the 1999 questionnaire.[5] Furthermore, readers are provided with the findings of the McNemar's test as a second means of interpreting the reliability of each of the YRBS items.

Third, although the kappa coefficients and McNemar's test findings suggest generally good reliability for most YRBS items, non-exact agreement between Time 1 and Time 2 could be the result of response error, skipped questions by respondents, or actual change in behavior. To minimize the likelihood that differences between Time 1 and Time 2 reporting were the result of behavior change, the Time 1 and Time 2 questionnaires were administered approximately 14 days apart.[6] Also, items with a "past 7 day" or "last time" reference period were dropped from the analysis because those reference periods could not be expected to be consistent across a 2-week timeframe. Fourth, this study exclusively used paper-and-pencil questionnaires; however, the national YRBS transitioned to electronic data collection (via tablets) in January 2023. Raghupathy found that the mean kappa reliability score using electronic data collection was similar to findings obtained in prior reliability studies of the YRBS administered as paper surveys.[17]

Another limitation of this study is that it did not ask students who missed a Time 1 questionnaire administration to complete a Time 2 questionnaire. Future reliability studies could do so and compare demographic characteristics of students who completed only a Time 1, only a Time 2, or both a Time 1 and Time 2 questionnaire. Further, although the majority (80%) of students who completed a Time 1 questionnaire also completed a Time 2 questionnaire, it is possible that the data were biased by the lack of information from the 20% of students missing a Time 2 administration. The Brener et al. 1999 reliability study

was conducted in spring 2000.[5] In that study, 77% of eligible students completed a Time 1 questionnaire, and of students who completed a Time 1 questionnaire, 89% completed a Time 2 questionnaire.

The YRBSS as a whole has faced increasing challenges with school districts' and schools' willingness to participate in health surveys.[4] But also, the current study was conducted during spring 2022, when COVID-19 was still very much in the minds of school staff and families. We speculate that many students missed a Time 1 or Time 2 administration because of absences related to experiencing COVID-19 symptoms or risks, and that could help explain slight differences in the Time 1 response rate (77% for Brener et al vs 70% in the current study) and the nearly 10 percentage point difference in loss of participants at Time 2 compared to that found in the Brener et al.[5] study.

Lastly, although this study measured test-retest reliability, it did not assess validity. For many behaviors examined on the YRBS, a validity study is impractical, although for some behaviors, objective measures of validity exist, such as tobacco, drug, and alcohol use as studied by Winters et al.[18] and Martin and Newman.[19]

CDC's YRBSS is the largest public health surveillance system in the United States dedicated to monitoring a broad range of health-related behaviors among high school students. Since 1991, the YRBSS has collected data from more than 4.9 million high school students in more than 2100 separate surveys.[2] Data are used by communities across the country to assess youth health needs, guide school health programs, and develop public health policies. A recent survey of state and local health and education agencies found a 2-fold jump in youth health data utilization in recent years, helping decision makers implement evidence-based interventions.[4] Research has provided evidence that anonymous, self-report, school-based surveys provide the best estimates of health risk behaviors.[20,21] The current study found that among items analyzed from the 2021 national YRBS questionnaire for test-retest reliability, 96% had at least moderate reliability, and 73% had substantial or almost perfect reliability, indicating the data are of high quality based on one means of summarizing the findings. In a second means of summarizing reliability, only 9 of 74 items significantly differed between the test and retest of the questionnaire. Thus, overall, the findings suggest the national YRBS questionnaire is a reliable instrument for monitoring health risk behaviors and experiences among high school students. Such findings lend support to relying on adolescent self-reported data when using the YRBS to help guide national, state, tribal, and local public health and school health programs.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Abbreviations

YRBS	Youth Risk Behavior Survey
CDC	Centers for Disease Control and Prevention

References

1. Centers for Disease Control and Prevention CDCBrener ND, Kann L, et al. Methodology of the youth risk behavior surveillance system — 2013. MMWR (Morb Mortal Wkly Rep). 2013;62(RR-01):1–20. [PubMed: 23302815]
2. Mpofu JJ, Underwood JM, Thornton J, et al. Overview and methods for the Youth Risk Behavior Surveillance System — United States, 2021. MMWR (Morb Mortal Wkly Rep) 2023; 72(1):1–12. [PubMed: 36602930]
3. Foti K, Balaji A, Shanklin S. Uses of Youth Risk Behavior Survey and School Health Profiles data: applications for improving adolescent and school health. J Sch Health 2011;81: 345–354. [PubMed: 21592130]
4. Smith Grant J, Pierre K, Stinson J, et al. The increasing utility of school health data to guide evidence-based interventions. J Sch Health. 2022;92(12):1214–1216. [PubMed: 36320177]
5. Brener ND, Kann L, McManus T, Kinchen SA, Sundberg EC, Ross JG. Reliability of the 1999 Youth Risk Behavior Survey questionnaire. J Adolesc Health 2002;31: 336–342. [PubMed: 12359379]
6. Yu CH. Test-retest reliability. Encyclopedia of Social Measurement. 2005;3:777–785.
7. Brener ND, Collins JL, Kann L, Warren CW, Williams BI. Reliability of the Youth Risk Behavior Survey questionnaire. Am J Epidemiol 1995;141:575–580. [PubMed: 7900725]
8. Klein JD, Graff CA, Santelli 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; 2001:11–18.
9. Rosenbaum JE. Truth or consequences: the intertemporal consistency of adolescent self-report on the Youth Risk Behavior Survey. Am J Epidemiol 2009;169(11):1388–1397. [PubMed: 19363096]
10. Centers for Disease Control and Prevention. Youth Risk Behavior Surveillance System (YRBSS) 2021 YRBS data user's guide. https://www.cdc.gov/healthyyouth/data/yrbs/pdf/2021/2021_YRBS_Data_Users_Guide_508.pdf. Accessed August 30, 2023.
11. SAS Institute. SAS/STAT 15.2 user's guide. The FREQ procedure: tests and measures of agreement. https://documentation.sas.com/doc/en/statug/15.2/statug_freq_details78.htm. Accessed September 1, 2022.
12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–174. [PubMed: 843571]
13. United States Census Bureau. American Communities Survey. B14003 sex by school enrollment by type of school by age for the population 3 years and over. https://data.census.gov/table?q=school+enrollment&tid=ACSDT1Y2021.B14003. Accessed June 9, 2023.
14. United States Department of Education, National Center for Education Statistics, Digest of Education Statistics. Table 203.65. Enrollment in public elementary and secondary schools, by level, grade, and race/ethnicity: selected years, fall 1999 through fall 2021. https://nces.ed.gov/programs/digest/d22/tables/dt22_203.65.asp. Accessed June 9, 2023.
15. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther. 2005; 85(3):257–268. [PubMed: 15733050]
16. Maclure M, Willett WC. Reviews and commentary: misinterpretation and misuse of the kappa statistic. J Epidemiol. 1987; 126(2):161169.
17. Raghupathy S, Hahn-Smith S. Reliability of the high school Youth Risk Behavior Survey when administered online. Int Q Community Health Educ 2012;32(2):135–148.
18. Winters KC, Stinchfield RD, Henly GA, Schwartz RH. Validity of adolescent self-report of alcohol and other drug involvement. Int J Addict. 1990;25:1379–1395. [PubMed: 2132719]
19. Martin GL, Newman IM. Assessing the validity of self-reported adolescent cigarette smoking. J Drug Educ. 1988; 18:275–284. [PubMed: 3265435]
20. Wechsler H. Guest Viewpoint: Do Students Tell the Truth about Risky Behavior. Alexandria, VA: School Board News. National School Boards Association. 2004.
21. Brener ND, Eaton DK, Kann L, et al. The association of survey setting and mode with self-reported health risk behaviors among high school students. Publ Opin Q. 2006;70(3):354–374.

Table 1.

Demographic Characteristics of Respondents.

Characteristic	Sample distribution N (%)	National distribution[13,14] %
Sex		
Female	284 (49.6)	49.3
Male	289 (50.4)	50.7
Race/Ethnicity		
American Indian or Alaska Native[a]	2 (.4)	0.9
Asian[a]	21 (3.7)	5.4
Black or African American[a]	71 (12.4)	14.7
Native Hawaiian/Other Pacific Islander[a]	2 (.4)	0.4
White[a]	256 (44.8)	46.3
Hispanic/Latino	191 (33.5)	28.3
Multiple races[a]	28 (4.9)	4.0
Grade		
9	178 (30.7)	27.0
10	136 (23.5)	25.0
11	140 (24.2)	23.9
12	124 (21.4)	23.8

N = 588.

[a]Non-Hispanic.

Table 2.

Kappa Coefficients and Time 1 and Time 2 Prevalence, by Questionnaire Item.

Item	Time 1% (95% CI[a])	Time 2% (95% CI)	Kappa coefficient	McNemar's test *P*-value
Behaviors related to unintentional injuries and violence				
Rarely or never wore a seat belt when riding in a car driven by someone else	45.5 (41.4–49.6)	47.6 (43.6–51.8)	.75	.151
Rode with a drinking driver during the past 30 days	15.2 (12.5–18.4)	15.6 (12.9–18.8)	.58	.900
Drove after drinking during the past 30 days	2.2 (1.0–4.8)	3.8 (2.0–7.2)	.60	.180
Texted or e-mailed while driving on ≥ 1 day during the past 30 days[b]	33.9 (28.6–39.8)	34.9 (29.1–41.2)	.67	.289
Carried a weapon on school property ≥1 day during the past 30 days	1.5 (.8–2.9)	1.7 (.9–3.2)	.66	.414
Carried a gun ≥1 day during the past 30 days (not counting carried a gun only for hunting or for a sport)	4.1 (2.8–6.1)	3.4 (2.2–5.3)	.62	.317
Felt too unsafe to go to school ≥1 day during the past 30 days	17.0 (14.2–20.3)	9.9 (7.8–12.7)	.53	**.000**
Threatened or injured with weapon on school property ≥1 time during the past 12 months	9.2 (7.1–11.8)	5.5 (3.9–7.7)	.38	**.003**
In a physical fight ≥1 time during the past 12 months	17.0 (14.2–20.3)	14.2 (11.5–17.3)	.65	.027
In a physical fight on school property ≥1 time during the past 12 months	5.7 (4.0–7.9)	5.3 (3.8–7.5)	.64	.827
Ever saw someone get physically attacked, beaten, stabbed, or shot in their neighborhood[b]	27.7 (24.2–31.5)	23.6 (20.3–27.2)	.58	.013
Ever forced to have sexual intercourse	6.8 (5.0–9.2)	8.3 (6.3–10.9)	.71	.201
Experienced sexual violence ≥1 time during the past 12 months[b]	10.9 (8.6–13.7)	9.3 (7.1–11.9)	.74	.072
Experienced sexual dating violence ≥1 time during the past 12 months[b]	9.3 (6.6–13.0)	10.5 (7.5–14.5)	.88	1.000
Experienced physical dating violence ≥1 time during the past 12 months[b]	7.0 (4.7–10.2)	6.7 (4.4–10.1)	.64	.782
Were bullied on school property during the past 12 months[b]	16.2 (13.4–19.4)	15.4 (12.7–18.6)	.72	.546
Were electronically bullied during the past 12 months[b]	13.7 (11.1–16.7)	15.6 (12.9–18.8)	.61	.109
Indicators of mental health				
Felt sad or hopeless during the past 12 months	46.9 (42.9–51.0)	40.6 (36.7–44.7)	.73	**.000**
Seriously considered suicide during the past 12 months	21.9 (18.7–25.4)	23.4 (20.1–27.0)	.78	.180
Planned suicide during the past 12 months	19.8 (16.7–23.2)	19.9 (16.8–23.3)	.81	.866
Had ≥1 suicide attempt during the past 12 months	8.5 (6.4–11.2)	10.1 (7.8–12.9)	.83	.033
Had injurious suicide attempt during the past 12 months	3.4 (2.1–5.3)	4.4 (2.9–6.5)	.67	.083
Mental health was most of the time or always not good during the past 30 days[b]	32.2 (28.5–36.2)	26.8 (23.3–30.6)	.55	**.002**
Mental health was most of the time or always not good during the COVID-19 pandemic[b]	39.7 (35.8–43.8)	36.1 (32.2–40.1)	.63	.054

Item	Time 1% (95% CI[a])	Time 2% (95% CI)	Kappa coefficient	McNemar's test *P*-value
Strongly agree or agree that they feel close to people at their school[b]	60.1 (56.1–64.1)	57.0 (52.9–61.0)	.67	.075
Have serious difficulty concentrating, remembering, or making decisions (because of a physical, mental, or emotional problem)[b]	49.6 (45.5–53.7)	45.2 (41.1–49.3)	.62	**.007**
Tobacco use behaviors				
Ever tried cigarette smoking	21.8 (18.6–25.4)	20.9 (17.7–24.4)	.83	.289
Age first smoked whole cigarette <13 years	7.1 (5.3–9.5)	10.8 (8.5–13.6)	.69	**.000**
Smoked cigarettes ≥1 day during the past 30 days	5.4 (3.8–7.5)	4.3 (3.0–6.4)	.58	.275
Smoked ≥10 cigarettes per day on the days smoked during the past 30 days	.0 (. -.)	.0 (. -.)	NA[c]	NA
Ever used an electronic vapor product[b]	40.8 (36.9–44.9)	41.1 (37.1–45.2)	.86	.631
Used electronic vapor products ≥1 day during the past 30 days[b]	19.2 (16.1–22.7)	18.9 (15.8–22.4)	.81	.450
Usually got their electronic vapor products by buying them themselves in a store or gas station during the past 30 days[b]	6.1 (2.7–12.9)	5.2 (2.2–11.8)	.65	.564
Used smokeless tobacco during ≥1 day during the past 30 days	1.7 (.9–3.2)	1.6 (.8–3.0)	.63	.705
Smoked cigars ≥1 day during the past 30 days	2.4 (1.4–4.0)	3.3 (2.1–5.1)	.59	.248
Tried to quit using all tobacco products during the past 12 months[b]	61.6 (53.6–69.0)	49.7 (41.7–57.6)	.45	.465
Alcohol and other drug use behaviors				
Age first drank alcohol <13 years	16.3 (13.5–19.6)	17.8 (14.9–21.2)	.65	.181
Drank alcohol ≥1 day during the past 30 days	23.3 (20.0–27.0)	20.2 (17.0–23.7)	.75	.228
Binge drank on ≥ 1 day during the past 30 days	8.9 (6.8–11.6)	10.1 (7.8–12.9)	.65	.353
Had 10 or more drinks in a row during the past 30 days[b]	2.1 (1.1–3.7)	2.0 (1.1–3.6)	.66	.705
Usually got the alcohol they drank by someone giving it to them during the past 30 days[b]	39.8 (31.6–48.7)	42.2 (33.3 – 51.7)	.51	.491
Ever used marijuana	34.3 (30.5–38.3)	33.4 (29.7–37.4)	.91	1.000
Age first used marijuana <13 years	4.7 (3.2–6.8)	6.1 (4.4–8.3)	.68	.059
Used marijuana during the past 30 days	19.3 (16.3–22.8)	20.2 (17.1–23.7)	.80	.612
Ever used synthetic marijuana[b]	6.2 (4.5–8.4)	6.6 (4.8–8.9)	.62	.695
Ever used prescription pain medicine without a doctor's prescription or differently than how a doctor told them to use it[b]	12.0 (9.6–14.9)	11.7 (9.3–14.6)	.74	.590
Used prescription pain medicine without a doctor's prescription or differently than how a doctor told them to use it ≥ 1 time during the past 30 days[b]	5.7 (4.0–7.9)	5.9 (4.2–8.1)	.45	1.000
Ever used cocaine	2.1 (1.2–3.6)	2.9 (1.8–4.7)	.75	.059

Item	Time 1% (95% CI[a])	Time 2% (95% CI)	Kappa coefficient	McNemar's test *P*-value
Ever used inhalants	6.7 (4.9–9.0)	5.3 (3.8–7.5)	.70	.074
Ever used heroin	.0 (. -.)	1.2 (.6–2.5)	NA	NA
Ever used methamphetamines	.2 (.0–1.2)	1.2 (.6–2.5)	.25	.014
Ever used ecstasy[b]	3.2 (2.1–5.0)	4.0 (2.6–5.9)	.72	.132
Ever injected any illegal drugs	.3 (.1–1.4)	1.4 (.7–2.7)	.20	.034
Offered, sold, or given illegal drugs on school property during the past 12 months	18.7 (15.7–22.1)	15.8 (13.0–19.0)	.59	.038
Ever used hallucinogenic drugs[b]	6.2 (4.5–8.5)	6.7 (4.9–9.1)	.82	.564
Sexual behaviors				
Ever had sexual intercourse	33.7 (29.9–37.8)	33.5 (29.6–37.5)	.94	.439
Age first had sexual intercourse <13 years	3.0 (1.9–4.8)	2.5 (1.5–4.2)	.68	.317
Had ≥4 lifetime sex partners	6.6 (4.8–9.0)	6.1 (4.4–8.5)	.86	.739
Had ≥1 sex partner during the past 3 months	23.0 (19.7–26.6)	22.7 (19.4–26.4)	.87	.841
Ever tested for human immunodeficiency virus (HIV)[b]	4.9 (3.4–7.0)	5.6 (4.0–7.8)	.55	.317
Were tested for a sexually transmitted disease (STD), other than HIV, during the past 12 months[b]	4.2 (2.8–6.1)	5.4 (3.9–7.7)	.68	.090
Dietary behaviors				
Perceive self as overweight	38.8 (34.9–42.8)	39.9 (35.9–43.9)	.85	.435
Trying to lose weight	49.0 (44.9–53.1)	50.2 (46.1–54.3)	.83	.386
Were overweight	18.0 (14.9–21.5)	17.6 (14.5–21.2)	.81	.564
Had obesity	22.6 (19.2–26.4)	21.9 (18.5–25.8)	.92	1.000
Physical activity behaviors				
Spent ≥3 hours per day on screen time, not counting time spent doing schoolwork, on an average school day[b]	81.3 (77.9–84.3)	76.2 (72.5–79.5)	.52	**.004**
Attend physical education class ≥1 day a week	78.8 (75.3–82.0)	77.6 (74.0–80.8)	.87	.162
Played on ≥ 1 sports team during the past 12 months	45.0 (41.0–49.1)	43.6 (39.6–47.8)	.86	.150
Had a concussion from playing a sport or being physically active ≥1 time during the past 12 months[b]	11.6 (9.2–14.5)	11.5 (9.1–14.4)	.59	.662
Other health-related topics				
Saw a dentist during the past 12 months	74.7 (71.0–78.1)	70.6 (66.7–74.2)	.65	.018
Got 8 or more hours of sleep on an average school night[b]	21.9 (18.7–25.5)	20.0 (16.9–23.5)	.68	.121
Usually did not sleep in their parent's/guardian's home during the past 30 days[b]	1.6 (.8–3.0)	2.3 (1.3–3.9)	.47	.366
Had a sunburn ≥1 time during the past 12 months[b]	53.2 (49.1–57.3)	48.2 (44.2–52.4)	.72	**.004**

Item	Time 1% (95% CI[a])	Time 2% (95% CI)	Kappa coefficient	McNemar's test *P*-value
Parent/other adult in the family most of the time or always know where they are going or with whom they will be[b]	90.0 (87.3–92.2)	85.3 (82.2–88.0)	.58	**.000**
Parent/other adult in the home lost their job during the COVID-19 pandemic[b]	30.6 (26.9–34.7)	32.0 (28.2–36.1)	.82	.343
Speak English well or very well[b]	98.8 (97.5–99.4)	98.1 (96.6–98.9)	.46	.096

N = 588 matched pairs for Time 1 and Time 2. Bold indicates Time 1 prevalence significantly different than Time 2 prevalence based on the McNemar's test ($P < .01$).

[a]Confidence interval.

[b]New item since the last YRBS reliability study of the 1999 questionnaire.[5]

[c]Statistic could not be computed due to zero respondents answering in the affirmative for Time 1 or Time 2.

Bolded p-values are < 0.01. This is indicated in the footnote.

Table 3.

Mean Kappa Statistic, by Demographic and Question Characteristics.

	Mean kappa
Total	.68
Sex	
Female	.69
Male	.64
Race and ethnicity[a]	
Black[b]	.62
White[b]	.70
Hispanic/Latino	.68
Grade	
9	.67
10	.68
11	.70
12	.65
Risk behavior categories	
Unintentional injuries and violence	.65
Indicators of mental health	.67
Tobacco use behaviors	.68
Alcohol and other drug use behaviors	.64
Sexual behaviors	.76
Dietary behaviors	.85
Physical activity	.71
Other health-related topics	.62

[a]Mean kappas for other racial groups are not presented because there is not enough data to calculate stable estimates.

[b]Non-Hispanic.