0145-6008/96/2006-1125$03.00/0
Alcoholism: Clinical and Experimental Research

Vol. 20, No. 6
September 1996

Evaluating Three Reading Tests for Use with Alcohol and Other Drug-Abusing Populations

Mark E. Johnson and Dennis G. Fisher

This study compared three reading tests commonly used in research for screening, descriptive, and educational purposes with alcohol and other drug-abusing individuals. To that end, 82 male and 41 female substance abusers were administered the Slosson Oral Reading Test-Revised, Woodcock Reading Mastery Test-Revised, and the Wide Range Achievement Test-Revised in random order. Results revealed that the tests have high concurrent validity, provide approximately the same grade-equivalent level scores, and yield raw scores that, when standardized, do not differ significantly from one another. However, if used for screening purposes, the three tests result in different proportions of subjects meeting specified criteria, particularly at lower grade levels. Specific test selection depends on the purpose of testing. For example, when the entire range of possible scores is of interest, the Woodcock Reading Mastery Test-Revised has a distinct advantage, because it has the widest range of grade-equivalent levels. Other considerations for test selection are discussed.

A VARIETY of reading tests has been used across several research projects investigating alcohol and other drug-abusing populations. A major purpose of such testing has been to screen potential participants to ensure that they have adequate reading ability to benefit from a given project. As an example of how reading tests have been used in research, Project MATCH (Matching Alcoholism Treatment to Client Heterogeneity), a multisite research project funded by the National Institute on Alcohol Abuse and Alcoholism, used ability of a potential subject to read at a minimum level of 6th grade as a screening criterion (Mattson, personal communication, 1995). This reading level was assumed if the potential participant had graduated from high school. For those individuals who did not have a high school diploma or equivalent, screening for 6th-grade reading level was accomplished through the use of the Slosson Oral Reading Test-Revised (SORT-R).[1] The SORT-R is only one of several tests that have been used to assess reading levels of alcohol and other drug-abusing participants. For example, Johnson et al.[2] used the Woodcock Reading Mastery Test-Revised (WRMT-R)[3] and the Wide Range Achievement Test-Revised (WRAT-R)[4] to evaluate the reading levels of injecting drug users and crack cocaine smokers.

Given the use of a variety of different reading tests, it is difficult to establish the comparability of research results across studies. For instance, one study cited herein[2] used two different tests and found significant differences between the tests, with the WRAT-R yielding significantly higher scores than the WRMT-R for the same *individual*. Such a situation can prove problematic if a particular reading level is specified as a minimum requirement for participation. Such documented differences across reading tests suggest that any specified grade level cut-off criterion will have a different meaning, depending on which test was used to assess reading level. That is, a 6th-grade reading level on one test may not necessarily be comparable with a 6th-grade reading level on another test.

Similarly, if results of a reading test are used to determine reading level of participants as a gauge to accessibility of written materials presented to participants, different conclusions may be drawn depending on the reading test selected. It has been found in numerous health care areas that required reading levels to understand written prevention literature typically exceed actual reading levels of the intended audience (e.g., Refs. 4–8). However, if there is variability across different reading tests, results obtained by comparing reading levels of individuals, and readability of written materials may well vary depending on the reading test selected.

Another problem lies in the fact that different reading tests yield different ranges of scores. For example, the SORT-R provides grade-equivalent scores that range from 0.1 to 12.5 grades, the WRAT-R from 3 to 12, and the WRMT-R from 0.5 to 16.9. Thus, selection of tests for use with alcohol and other drug-abusing individuals may depend not only on a test's validity, but also on the purpose of test administration. That is, if the purpose is to assess the widest range of possible grade-equivalent scores, then it would be advantageous to select an instrument that has a high ceiling and low floor. If, on the other hand, the purpose is to obtain a quick screening of an individual's reading level to ensure that she or he can read at a minimum specified reading level, then having a test with a wide range of scores is not necessary.

The purpose of this study was to examine the comparability of three reading tests (SORT-R, WRMT-R, and WRAT-R) that have been used frequently with alcohol and

From the Department of Psychology, University of Alaska Anchorage, Anchorage, Alaska.

Received for publication February 1, 1996; accepted April 15, 1996.

This study was funded in part by Grant U01-DA-07290 from the NIDA.

Reprint requests: Mark E. Johnson, Ph.D., Department of Psychology, 3211 Providence Drive, University of Alaska Anchorage, Anchorage, AK 99508.

other drug-abusing populations. To do so, we compared obtained scores, grade-equivalent scores, and distribution of grade-equivalent scores.

METHODS

Participants

Participants were involved in a National Institute for Drug Abuse (NIDA) funded project designed to provide human immunodeficiency virus/acquired immune deficiency syndrome educational counseling to injecting and other drug users not currently in treatment. Eligibility criteria for participation in the larger NIDA project are: (1) age 18 or older; (2) not having been in substance abuse treatment within the last 30 days; and (3) either positive urinalysis for morphine, cocaine metabolites, or amphetamine, or visible signs of injection. There were 82 men and 41 women, of whom 37.7% were Black, 34.4% White, 2.5% Hispanic/Latino, 0.8% Asian/Pacific Islander, 22.1% Native American/Alaskan Native, and 2.5% other. Ages ranged from 19 to 64 years (mean = 35.28, SD = 7.90, median = 35). Educational levels were as follows: 8th grade or less, 6.6%; less than high school, 14.9%; GED (high school equivalent), 14.9%; high school graduate, 34.7%; trade/technical school, 1.7%; some college, 24.7%; and college graduate, 2.5%.

All participants in the larger NIDA project respond to the Risk Behavior Assessment (RBA),[9] a structured interview that assesses high-risk behaviors, including substance use, needle-sharing, and sexual behaviors. The RBA has been demonstrated to have good reliability, and the questions regarding drug use to have good reliability and validity.[10–13] Based on responses to this instrument, alcohol was the substance most commonly used in the last 48 hr and 30 days by participants in the current study, followed by crack cocaine, marijuana, and cocaine. Of the 123 subjects, 82.1% were eligible because of use of crack or other forms of cocaine, 6.3% for use of opiates, 12% for use of both cocaine and opiates, and 48.2% for visible needle tracks (total exceeds 100% because eligibility can be documented through more than one means).

Instrumentation

SORT-R. This quick measure of a respondent's reading level is individually administered. It contains 10 lists of 20 words arranged in ascending order of difficulty. Participants read the words until reaching a ceiling, defined as the inability to read any of the words on one entire list. A total raw score is obtained by summing the number of items read correctly; grade-equivalent scores can be calculated and range from 0.1 to 12.5. Reliability coefficients for the standardization sample were reported at 0.98.[1] Concurrent validity was also reported as good, with high correlations between the SORT and the Peabody Individual Achievement Test and Woodcock-Johnson Tests of Achievement.[1]

WRMT-R. This individually administered instrument was designed to provide a thorough assessment of a respondent's reading abilities. It consists of six subtests and is appropriate for ages kindergarten to 75 years and older. Two subtests [Word Identification (WID) and Passage Comprehension (PC)] comprise a short form of the WRMT-R and yield an estimate of global reading ability and were used for the purposes of this study. The short form correlates 0.98 with the full instrument and has median split-half reliability across ages of 0.97.[3] Validity for the WRMT-R is well-established through high correlations with other established instruments.[3]

W scores (transformation of raw scores into standardized ability scores) are obtained for the subtests that can range from 338 to 608, with a mean of 500 (the average reading ability of a 5th-grader). A total reading score is obtained by summing the two subtest W scores and dividing by two. Age-equivalent and grade-equivalent scores (ranging from 0.5 to 16.9) are obtained for the subtest and total scores.

WRAT-R. This individually administered instrument was designed as a screening measure for achievement. Although it consists of three subtests (reading, spelling, and arithmetic), only the reading subtest was administered in the current study. The reading subtest has a split-half reliability of 0.94 and median coefficient α of 0.94. As one of the most commonly used achievement tests, this instrument has well-established validity.[4] Based on a respondent's raw score, a grade-equivalent score that can range from 3 to 12 is obtained.

Procedure

Participation in the overall NIDA project involves two separate sessions: the first session is devoted to the RBA, pretest human immunodeficiency virus counseling, and blood drawing; and the second session is devoted to feedback of blood test results, posttest counseling, and educational counseling. Participants were administered the three reading tests at the conclusion of the second session. Administration order of the three tests was randomly determined for each participant. Before testing, examiners were trained in proper test administration and scoring by the first author, a licensed psychologist.

RESULTS

Preliminary analyses were conducted to compare reading levels across ethnic groups and gender. A two-way MANOVA was calculated, with gender and ethnicity as independent variables, and raw scores on the WRAT-R, SORT-R, and the two WRMT-R scales as dependent variables. Results revealed no significant main or interaction effects.

Figure 1 provides a distribution of participants across the possible grade-equivalent levels for each test. When examining the ceilings, for the SORT-R, 39.8% of the participants reached the ceiling; for the WRAT-R, 31.7%; and for the WRMT-R, 24.2%. If the distribution of grade-equivalent levels for the three tests were collapsed in such a way that the scaling for all tests is the same, the tests yield roughly equivalent proportions of participants who reach the ceiling. For example, to force the WRMT-R onto a scale ranging from 2nd to 12th grade, grade-equivalents between 12 and 16.9 would be collapsed into one category. When all three tests are thus collapsed, the figures for participants reaching the ceiling for the SORT-R, WRAT-R, and WRMT-R are 39.8%, 38.3%, and 42.5%, respectively.

Table 1 provides the percentages of participants whose reading levels fall below each of seven potential screening criteria for each of the three tests. Using an example of a screening criterion of at least a 6th-grade reading level (as was used in Project MATCH), when using the SORT-R, 17.1% of the current participants would have been ineligible had this reading test been the sole criterion for eligibility. Comparable percentages for the WRAT-R and WRMT-R would have been 22% and 27.6%, respectively. Across all other possible screening criteria listed in Table 1, disparity between the three tests also was evident. A series of Z-tests comparing proportions of participants who would be excluded at different grade levels using the three different tests revealed significant differences only at the 6th and 7th grades and only between the SORT-R and WRAT-R, $Z = -1.99, p < 0.05$, $Z = 1.68, p < 0.05$, respectively.

Table 2 provides the raw score means and the median

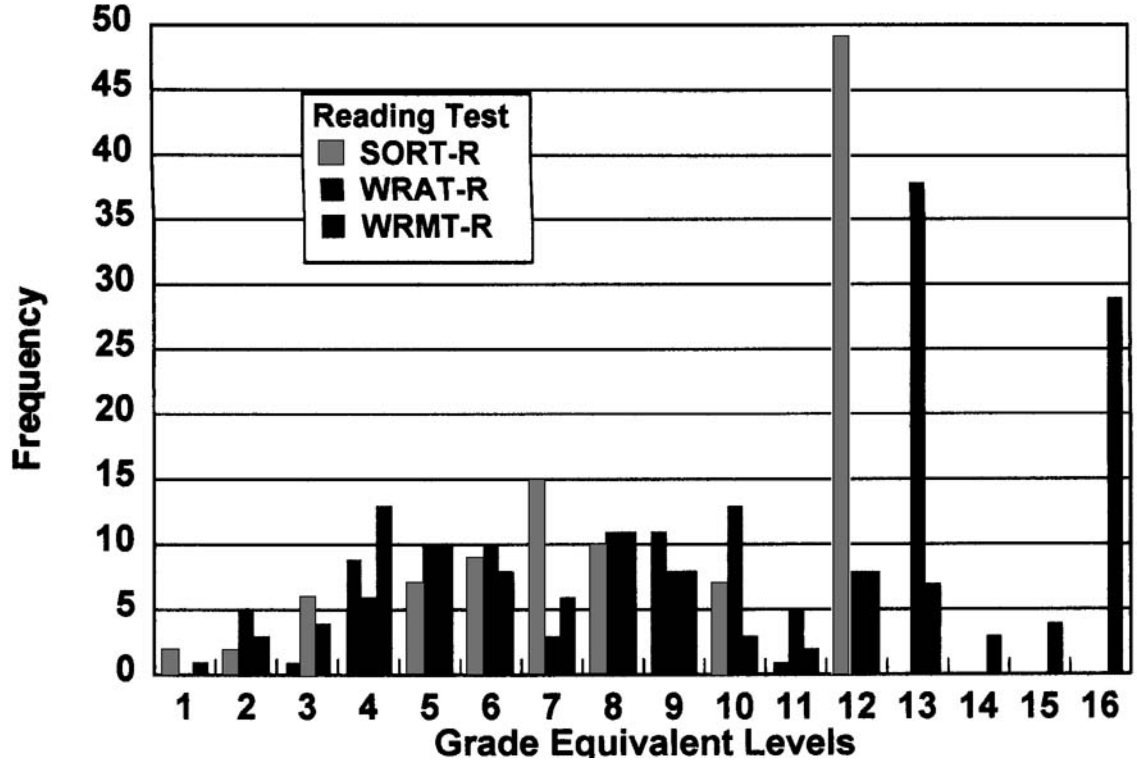

Fig. 1. Frequency of grade-equivalent levels for SORT-R, WRAT-R, and WRMT-R.

Table 1. Percentages of Participants Who Fall Below a Given Grade-Equivalent Level for the Three Reading Tests

	Grade-equivalent level						
Reading test	6	7	8	9	10	11	12
SORT-R	17.1	24.4	36.6	44.7	53.7	59.3	60.2
WRAT-R	27.6	34.1	39.0	48.0	54.5	56.9	58.5
WRMT-R	22.0	30.1	32.5	42.3	48.0	58.5	62.6

Table 2. Mean Raw Scores, Standard Deviations, and Grade-Equivalent Levels

	Mean raw score	SD	Median grade-equivalent
WRAT-R	55.47	14.87	10.1
WRMT-R			
PC	47.21	10.67	9.0
WID	85.54	13.56	9.4
Total	519.29*	19.78	9.3
SORT-R	169.36	33.64	9.5

* W score.

grade-equivalent level on the three reading tests. To compare results from the three reading tests, data were analyzed in two different ways: one analysis compared obtained means, and the other analysis compared medians. In the first set of analyses, all raw scores were standardized to a mean of 0, and a standard deviation of 1 and dependent *t* tests for all possible pair-wise mean comparisons were computed. Results of these analyses indicated no significant differences between the standardized raw scores. In the second set of analyses, a series of Wilcoxon Signed-Rank Tests were calculated, comparing the median grade-equivalent levels obtained on the three tests. This analysis first converted all scores to ranks and then compared the relative ranks of participants across tests. Results revealed no significant differences in any pair-wise mean comparison of the three tests.

The relationship between the highest level of school completed and reading test results was explored through a series of Spearman rank-order correlations. Results revealed the grade-equivalent levels obtained through WID and PC subtests, and overall WRMT-R, SORT-R, and WRAT-R scores were all significantly related to reported level of school completed: r = 0.31, r = 0.34, r = 0.37, r = 0.27, and r = 0.33, respectively (all $p < 0.005$). These correlations indicate level of schooling only accounts for 7.3 to 13.7% of the variance in reading level measures.

Correlation coefficients calculated among the three reading tests indicated high concurrent validity. The highest correlations were revealed between those tests that have similar demand characteristics, SORT, WID, and WRAT-R. All three of these tests presented the examinee with a list of words and asked for the words to be read aloud. The correlation between the SORT-R and WID was 0.94; SORT-R and WRAT-R, 0.86; and WID and WRAT-R, 0.91. The lowest correlations were obtained between the PC subtest and the previously described word tasks, with all coefficients being 0.77. The total WRMT-R correlated 0.92 with the SORT-R and 0.89 with the WRAT-R.

DISCUSSION

This study compared three reading tests commonly used with alcohol and other drug-abusing individuals. Results indicated that the three tests had high concurrent validity, provided approximately the same grade-equivalent scores, and yielded raw scores that, when standardized, were not statistically different from one another. One major difference revealed between the three tests was the ceiling effect. Specifically, the proportion of participants who reached the ceiling for the SORT-R was 39.8%; for the WRAT-R, 31.7%; and for the WRMT-R, 24.2%. These differences were caused largely by the different scaling of the grade-equivalent levels for the three tests, with the SORT-R ranging from 0.1 to 12.5, the WRAT-R from 2 to 13, and the WRMT-R from 0.5 to 16.9. When the scaling of the three tests was collapsed to yield a 2 to 12 scale, the proportion of participants reaching the ceiling was roughly the same, ranging from 38.3% to 42.5%.

Findings demonstrate that, when using reading tests as a screening tool to determine inclusion or exclusion based on a specific grade-equivalent level, the three tests could yield considerably different participant pools, particularly if the cut-off point is the 6th- or 7th-grade level. For example, if a criterion of a 6th-grade reading level was used (as in Project MATCH), the percentage of ineligible participants would vary by as much as 10.5%, depending on the reading test used, with the SORT-R being the most liberal in its determination of 6th-grade reading level.

Further, given the variability of grade-equivalent levels yielded by the three tests, with the WRAT-R providing a higher score than the other two tests, it is difficult to identify the actual reading level of an individual with the use of only one test. With regard to the WRAT-R, other researchers (e.g., Ref. 14) have reported that this test tends to overestimate by 1 to 2 years, corroborating the results of the current study. Thus, it may be that more credence should be lent to grade-equivalent levels provided by the WRMT-R and SORT-R, which were nearly identical to each other.

In terms of test selection, the difference in ceilings should play a major role in the decision-making process, with the purpose of the testing dictating the choice of test. If the purpose is to provide a quick screening to determine whether an examinee can read at a specific reading level no greater than 12th grade, any of the three tests could be used, bearing in mind that the WRAT-R tends to overestimate and that the three tests yield different percentages of eligible participants. However, if information is needed about a wide range of grade levels, then the WRMT-R would be the instrument of choice as it extends up to a grade-equivalent level of 16.9. It is also probable that this test provides the most accurate indicator of an individual's reading abilities, because it assesses not only word recognition but also reading comprehension. The use of the WRMT-R, with its wider range of grade levels, also allows for greater secondary analysis of data. If administration time is a factor, the WRAT-R with the fewest items (89) would be the instrument of choice, followed by the SORT-R (with 200 items), and the WRMT-R (with 106 items in WID and 68 in PC). Training requirements are roughly the same for all three tests as are, with the exception of PC, the demand characteristics of the test.

It should be noted that, as the data for the current study were being collected, a new revision of the WRAT was released. However, the demand characteristics of the reading subtest are the same in both versions, and the correlation between the two versions is very high ($r = 0.94$).[15] Given these two facts, the results of the current study using the WRAT-R should not differ significantly had the WRAT-3 been used.

Last, although we objectively assessed the presence of cocaine metabolites, amphetamines, and opium through urinalysis and consumption of alcohol and other drugs through self-report, we did not collect information on the quantity of substances used immediately before testing. The vast majority of participants reported using some substance within the last 48 hr, as verified by urinalysis results. Although between-subject variations may have effected test performance, we do not have adequate data to assess this question. The only conclusion that can be drawn is that most, if not all, participants were impaired at the time of testing and this may have effected performance. However, because the specific results were not as crucial in this study as comparisons across tests, acute substance use is not as much as an issue as it might be under other circumstances.

REFERENCES

1. Slosson RL, Nicholson CL: Slosson Oral Reading Test, SORT-R. East Aurora, NY, Slosson Educational Publications, 1990

2. Johnson ME, Fisher DG, Davis DC, Cagle HH, Rhodes F, Booth R, Siegal H, Jones A: Assessing reading level of drug users for HIV/AIDS prevention purposes. AIDS Educ Prevent (in press)

3. Woodcock RW: Woodcock Reading Mastery Tests-Revised. Circle Pines, MN, American Guidance Services, 1987

4. Jastak S, Wilkinson GS: Wide Range Achievement Test-Revised. Wilmington, DE, Jastak Associates, 1984

5. Davis TC, Crouch MA, Wills G, Miller S, Abdehou DM: The gap between patient reading comprehension and the readability of patient education materials. J Fam Pract 31:533–538, 1990

6. Malousff J, Gabrilowitz D, Schutte N: Readability of health warnings on alcohol and tobacco products. Am J Public Health 82:464, 1992

7. Sorenson JL, Leder D: Measuring the readability of written information for clients, in Landsberg G, Neigher WD, Hammer RJ, Windle C, Woy JR (eds): Evaluation in Practice: A Sourcebook of Program Evaluation Studies from Mental Health Care Systems in the United States. Washington, D.C., U.S. Government Printing Office, 1979, pp 113–114

8. Streiff LD: Can clients understand our instructions? IMAGE: J Nurs Scholarship, 18:48–52, 1986

9. National Institute on Drug Abuse: Risk Behavior Assessment.

Rockville, MD, National Institute on Drug Abuse (Community Research Branch), 1991

10. Dowling-Guyer S, Johnson ME, Fisher DG, Needle R, Watters J, Andersen M, Williams M, Kotranski L, Booth R, Rhodes F, Weatherby N, Estrada AL, Fleming D, Deren S, Tortu S: Reliability of drug users' self-reported HIV risk behaviors and validity of self-reported recent drug use. Assessment 1:383–392, 1994

11. Fisher DG, Needle R, Weatherby N, et al: Reliability of drug user self-report. IXth International Conference on AIDS (PO-C35-3355). Berlin, 1993 (abstr)

12. Needle R, Fisher DG, Weatherby N, Chitwood D, Brown B, Cesari H, Booth R, Williams ML, Watters J, Anderson M, Braunstein M: The reliability of self-reported HIV risk behaviors of drug users. Psychol Addict Behav 9:242–250, 1995

13. Weatherby NL, Needle R, Cesari H, Booth R, McCoy CB, Watters JK, Williams M, Chitwood DD: Validity of self-reported drug use among injection drug users and crack cocaine users recruited through street outreach. Eval Progr Plan 17:347–355, 1994

14. Snart F, Dennis S, Brailsford A: Concerns regarding the Wide Range Achievement Test. Can Psychol 24:99–103, 1983

15. Wilkinson GS: Wide Range Achievement Test-3. Wilmington, DE, Jastak Associates, 1993