

Passageless Comprehension on the Nelson-Denny Reading Test: Well Above Chance for University Students

Journal of Learning Disabilities 43(3) 244-249 © Hammill Institute on Disabilities 2010 Reprints and permission: http://www.sagepub.com/journalsPermissions.nav DOI: 10.1177/0022219409345017 http://journaloflearningdisabilities.sagepub.com



Chris Coleman, Jennifer Lindstrom, Jason Nelson, William Lindstrom, And K. Noël Gregg

Abstract

The comprehension section of the Nelson-Denny Reading Test (NDRT) is widely used to assess the reading comprehension skills of adolescents and adults in the United States. In this study, the authors explored the content validity of the NDRT Comprehension Test (Forms G and H) by asking university students (with and without at-risk status for learning disorders) to answer the multiple-choice comprehension questions without reading the passages. Overall accuracy rates were well above chance for both NDRT forms and both groups of students. These results raise serious questions about the validity of the NDRT and its use in the identification of reading disabilities.

Keywords

Nelson-Denny Reading Test, passageless comprehension, test validity

The assessment of possible learning disabilities typically includes the administration of standardized, norm-referenced tests designed to measure proficiency in specific aspects of academic achievement. The scores from such tests are important because psychologists rely on them to determine the presence (or absence) of significant deficits (i.e., underachievement) in academic skill areas. At the risk of stating the obvious, the diagnostic results of a psychoeducational evaluation can be only as valid as the scores from the tests that were administered.

In the area of reading comprehension, many standardized tests use a multiple-choice format in which the examinee attempts to answer questions based on a passage that he or she has just read. One of the concerns raised by researchers about tests with this format involves passage independence, that is, the likelihood that on some items, an examinee could respond correctly (based on prior knowledge or educated guessing) without having read the accompanying passage (Keenan & Betjemann, 2006). Regardless of the fact that prior knowledge and inferential reasoning are important factors in reading comprehension, passage-independent items are recognized as major threats to content validity (or, according to some researchers, construct validity). In other words, the utility and validity of a reading comprehension test (the degree to which it measures the target construct) are compromised to the extent that its questions can be answered correctly in a passageless administration. Despite decades of ongoing

attempts by researchers to alert test developers to passage independence and its consequences (e.g., Preston, 1964; Pyrczak & Axelrod, 1976), studies have consistently uncovered passage-independent items on standardized reading comprehension measures including the Minnesota Scholastic Aptitude Test (Fowler & Kroll, 1978), Stanford Achievement Test (Lifson, Scruggs, & Bennion, 1984), Scholastic Achievement Test (SAT; e.g., Daneman & Hannon, 2001; Katz, Lautenschlager, Blackburn, & Harris, 1990), Test of English as a Foreign Language (Tian, 2006), and Gray Oral Reading Test (Keenan & Betjemann, 2006). This trend suggests that the authors of many well-known reading comprehension measures have neglected to address the question of passage dependence in the test development process. For an in-depth review of previous studies of passage dependence, the reader is directed to Keenan and Betjemann (2006).

Purpose of the Study

Our study was designed to augment prior studies examining the content validity of multiple-choice reading comprehension

University of Georgia, Athens, GA

Corresponding Author:

Chris Coleman, University of Georgia, 335 Milledge Hall, Athens, GA 30602 Email: ccoleman@uga.edu Coleman et al. 245

measures. As noted above, in a similar study. Keenan and Betjemann (2006) helped to illustrate the gravity of the problem of passage independence in their examination of the comprehension component of the Gray Oral Reading Test, Third Edition (GORT-3; Wiederholt & Bryant, 1992), a measure with a multiple-choice format similar to that of the Nelson-Denny Reading Test (NDRT; Brown, Fishco, & Hanna, 1993a). The GORT-3 and GORT-4 (which are identical with the exception of an additional low-end passage on the newer version; Wiederholt & Bryant, 2001) differ from the NDRT in that they require examinees to read passages aloud and do not permit examinees to look back at the passages when answering comprehension questions; also, the GORT norms are age-based and range from ages 6 years 0 months to 18 years 11 months. Keenan and Betjemann found that their undergraduate participants (N = 77) were able to answer 86% of GORT-4 questions with above-chance accuracy without having read the passages. In the same passageless format, a smaller group of children (N = 10; age range = 7-15) achieved a mean accuracy rate of 47% (chance would have been 1 in 4 or 25%). Furthermore, in a separate sample of children with and without reading disabilities (N = 82 and 256, respectively), analyses comparing passage-dependent and passage-independent items on the standard administration of the GORT showed that passage-independent items were not sensitive to reading disability, thus confirming that only the most passage-dependent items are useful in detecting differences in comprehension performance as a function of reading disability (Keenan & Betjemann, 2006, p. 373).

The purpose of this study was to investigate passage dependence on the comprehension section of the NDRT among university students with and without at-risk status for learning disorders. Despite the fact that its norms are dated, the NDRT continues to be widely used and accepted in the United States as a valid measure of reading comprehension for high school and college students. For example, the Law School Admission Council (LSAC) identifies the NDRT (Form G or Form H) as the "preferred measure" of reading comprehension for psychologists documenting functional impairments (e.g., dyslexia) in students hoping to take the Law School Admission Test (LSAT) with accommodations (www.lsac.org). To date, no published studies have explored passage dependence on the NDRT (G/H) with any population.

Development and Technical Characteristics of the NDRT

The development and norming of the NDRT are described in both the *Manual for Scoring and Interpretation* (Brown, Fishco, & Hanna, 1993b) and the *Technical Report* (Brown, Fishco, & Hanna, 1993c). Following informal and formal passage/item tryouts (at several grade levels), Forms G and H were finalized and nationally standardized in 1991 and

1992. The 2-year college sample (N = "nearly 5,000") and the 4-year college sample (N = "over 5,000") were drawn from randomly sampled cells stratified by region, size, and institution type (public/private).

With regard to content validity, the development of Forms G and H reportedly (Brown et al., 1993b)

began [in 1989] with a mailing of 1,650 questionnaires to test users in all parts of the country at high school and college levels.... In general, respondents expressed satisfaction with the content and form of the test. Therefore, it was decided that the new edition of the test would have content and form substantially like those of previous editions. (p. 2)

According to the *Technical Report*, the questionnaire response rate was 27% and the rate of satisfied respondents was 74% (Brown et al., 1993c, p. 2). It should be noted that the pool of items for G and H was subsequently developed; thus, the questionnaire responses described as evidence of content validity for G and H were from individuals who had not actually seen any potential G or H test items. From a psychometric standpoint, this procedure surely falls short of reasonable content validity standards (e.g., the 1985 or 1999 AERA/APA/NCME joint Standards for Educational and Psychological Testing).

The development of the NDRT (G/H) took place in three stages: informal item tryouts, national item tryouts, and the national standardization. With regard to informal tryouts of comprehension items, "571 questions were developed for 32 passages. Passages were drawn from latest editions of humanities, social science, and science texts widely used at the high school and college levels" (Brown et al., 1993b, p. 2). In other words, the potential content for the test was drawn from sources/topics that would not be unfamiliar to most college students. This approach might be appropriate for a test of general knowledge, but for a test of reading comprehension, it is likely to introduce a great deal of constructirrelevant variance-in short, some examinees will inevitably know the answers to some questions without reading the passages. The NDRT manuals provide no further information about the source, selection, and adaptation of comprehension passages. There is no mention of passage dependence.

Formal (national) item tryouts included 7,300 students across four grades (9, 11, 13, and 15). The same 32 passages described above, with fewer questions (272), were distributed across eight tryout forms of the test; approximately 900 students took each of the eight forms (Brown et al., 1993c). All questions were subsequently analyzed using the Mantel-Haenszel test to detect possible bias with regard to race (Black vs. White; Hispanic vs. White) and gender. According to the *Manual for Scoring and Interpretation* (Brown et al., 1993b),

This analysis led to the identification of 74 items that were flagged as "possibly biased." Of these items, 49 were eliminated. The remaining 25 items had such high discrimination value that they were retained, so as not to diminish the reliability of the test. However, even these remaining items were chosen so as to obtain the best possible balance between referent and focal groups, thus minimizing favoritism in any one direction. (p. 2)

It is unclear exactly how this logic justifies the retention of items known to be "possibly biased."

Following national item tryouts, an unspecified number of adults diverse in race and gender were asked "to react subjectively to all items and to the reading passages used in the test" (Brown et al., 1993b, p. 2). Based partly on the panelists' comments, 18 of the 32 passages were discarded. Thus, "the best 14 [passages] were selected for Forms G and H" (Brown et al., 1993c, p. 4). With regard to the structure of each form of the test, "There is an opening passage of sufficient length to measure reading rate, followed by six shorter, 200-word passages. Each longer passage has eight questions, each shorter passage five questions" (p. 4). Of the 38 total questions on each form, half were designed to be "literal, i.e., testing student ability to note explicit details. The other half were interpretive, i.e., requiring students to infer or deduce answers, draw conclusions, grasp central idea or purpose, and make judgments" (p. 4).

The NDRT Manual for Scoring and Interpretation cites several studies said to provide evidence of predictive validity (Brown et al., 1993b). Essentially, the theme of the provided study summaries is that NDRT scores function as valid and reliable predictors of reading aptitude, reading attitude, and general academic achievement. However, for all but one described study, results were apparently based on previous versions of the NDRT (E/F). The one study mentioned that presumably used Forms G and H (the year was 1993; the forms and authors were not specified) involved three California community colleges that investigated the relationship between NDRT scores and end-of-course grades in a series of developmental literacy classes. Correlations ranged from .21 to .72, the highest of those figures reflecting the relationship of the NDRT comprehension and total scores to student grades in an advanced reading course (Brown et al., 1993b).

The 1993 version of the NDRT contains a vocabulary section and a reading comprehension section; on the latter section, a reading rate score (based on the first minute of the test) can also be obtained. On both test forms (G and H), the reading comprehension section consists of seven reading passages and a total of 38 multiple-choice questions (each with five possible answers). Examinees are instructed to read completely through a passage before answering its questions; they are also told that they may look back at the passages but should not puzzle too long over any one question. The passages are

Table 1. Descriptive Statistics for Typical and At-Risk Students

		Gender	Age		Years of Education	
Group (Form)	N	F/M	М	SD	М	SD
Typical (G) ^a Typical (H) ^b At risk (G) ^a	125 128 28	79/46 101/27 10/16	19.0 18.9 20.5	1.1 1.9 2.4	12.7 12.6 12.9	0.93 0.94 1.3

^aThese participants were administered Form G in a passageless format.

divided into three types: humanities (2), social sciences (2), and science (3). Of the final six passages on each form, three were reportedly drawn from high school texts and three from college texts (Brown et al., 1993b); the sources of the initial passages (which are longer and used to determine the reading rate score) are not specified. The 38 comprehension questions are divided equally between "primarily literal" items and "largely interpretive" ones; the authors note that these categorizations were somewhat arbitrary (Brown et al., 1993b, p. 11). Finally, the alternate-form reliability for the comprehension sections of Forms G and H is reported to be .81 (Brown et al., 1993c).

Method

Participants

This quasi-experimental study was based on data obtained from two groups of university undergraduates: typical students and students at risk for learning and/or attention disorders. All participants were native speakers of English. The typical students were arbitrarily divided into two subsamples, one to be administered Form G in a passageless format and the other to be administered Form H in a passageless format. Descriptive information about the participants can be found in Table 1.

The typical students (total N=253), undergraduates at the flagship public university in their state, denied any history of special education services. Their voluntary research assistance fulfilled a requirement associated with introductory psychology courses. These participants completed the research tasks during group sessions lasting 1 hour or less.

The at-risk students (N=26) sought comprehensive neuropsychological evaluations (8–10 hours) at a university-based clinic for adults with possible learning disorders. Their presenting complaints were mixed, but most were concerned about possible reading and/or attention deficits. A majority (18) of the 26 students had been previously diagnosed with a learning and/or attention-deficit disorder and were seeking updated documentation of their need for academic modifications. The at-risk participants were asked to complete the research task during their individual evaluations.

^bThese participants were administered Form H in a passageless format.

Coleman et al. 247

Table 2. Overall Accuracy on Passageless Administration of Nelson-Denny Reading Test (NDRT) by Form and Group

		# of Items Correct ^a			
Form (Group)	N	М	SD	% Correct ^b	
H (typical)	128	17.7	4.1	46.6	
G (typical)	125	16.7	4.2	43.8	
G (at risk)	28	15.4	4.7	40.6	

^aOut of a possible 38.

As noted above, the typical sample was divided into two subsamples; thus, the study included a total of three groups. Gender distribution was significantly different across the groups, $\chi^2 = 15.05$, p = .001, as the Form H typical group had a higher proportion of females (78.9%) than the Form G typical group (63.2%), which in turn had a higher proportion of females than the at-risk group (44.0%). t tests indicated that the at-risk group was significantly older (p < .01) and had significantly more education (p < .01) than either of the typical groups, which were comparable with one another in terms of these demographics.

Passageless NDRT Administration

All participants were administered a passageless version of the NDRT. In other words, they were presented with a version of the Form G or Form H booklet from which the passages had been removed. The participants were instructed to attempt to answer the multiple-choice questions correctly even though they could not see the passages on which the questions had been based. There was no time limit, and participants were encouraged to attempt every item.

Results

Overall accuracy rates can be found in Table 2. For all subgroups and both test forms, questions were answered much more successfully than chance would have dictated (chance = 1 in 5 or 20%). The typical students, who answered (on average) nearly half the NDRT questions correctly, were slightly more successful (by about one item) on Form H (46.6%) than on Form G (43.8%). The at-risk students were, on average, less successful (by about one item) than the typical students completing the same test form (G). However, the accuracy rate for the at-risk students (40.6%) was still twice what would have been expected based on chance. Differences in overall accuracy were statistically significant only between the Form H typical group (46.6%) and the at-risk group (40.6%), F(1, 151) = 4.70, p < .05. There was no main

Table 3. Humanities, Science, and Social Science Passages: Accuracy on Passageless Administration of Nelson-Denny Reading Test (NDRT) by Form and Group

		% Correct ^a			
Form (Group)	Ν	Humanities	Science	Social Science	
H (typical)	128	30.7	56.6	50.6	
G (typical)	125	37. 4	52.3	39.3	
G (at risk)	28	30.8	46.7	36.2	

^aChance accuracy would be 20% (each multiple-choice question had five possible answers).

effect for gender, F(1, 272) = .003, p > .05, or interaction between group and gender, F(2, 272) = .217, p > .05.

The next set of analyses was conducted by passage type. Each form of the NDRT includes seven passages: two based on humanities topics, two based on social science topics, and three based on hard science topics. Table 3 shows that in the passageless format, all groups were most successful on questions from the hard science passages (46.7% to 56.6% correct) and least successful on questions from the humanities passages (30.7% to 37.4% correct). As with most other Form G trends, accuracy rates were slightly higher for typical students than for at-risk students. Results from the typical subsamples suggest that the social science questions on Form G are somewhat more passage dependent (i.e., harder to answer without reading the passage) than those on Form H, whereas the humanities questions on Form H appear to be somewhat more passage dependent than those on Form G. Of the 14 passages across the two forms of the NDRT, only one (Passage 1, Form H), a humanities passage, had an accuracy rate that was not well above chance (19.5%). Among typical students, mean accuracy rates for the remaining 13 passages ranged from 35.2% to 60.3%.

The next set of analyses was conducted at the item (i.e., individual question) level. As noted earlier, each NDRT item has five response choices, which would suggest a passageless accuracy rate of about 20%. However, our data indicated that among both typical and at-risk undergraduates, 70% to 80% of test items were answered with above-chance accuracy (see Figure 1). For each group, at least half of the items were answered correctly more than 40% of the time. Again, the atrisk students were slightly less successful than their typically achieving peers, but their accuracy rates were much closer to those of their typical counterparts than to chance.

The authors of the NDRT classified each comprehension question as either literal (i.e., the answer can be found, word for word, in the passage) or interpretive (requiring more indepth analysis/inferencing). Each form is equally divided between literal and interpretive items (19 each). As can be seen in Table 4, both types of questions were answered by

^bChance accuracy would be 20% (each multiple-choice question had five possible answers).

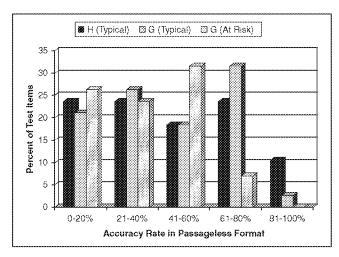


Figure 1. Percentage of test items in each quintile of passageless accuracy (by form and group)

Table 4. Literal and Interpretive Items: Accuracy on Passageless Administration of *Nelson-Denny Reading Test* (NDRT) by Form and Group

		% Correct²		
Form (Group)	N	Literal	Interpretive	
H (typical)	65	46.0	46.3	
G (typical)	65	33.4	54.1	
G (at risk)	28	29.8	47.2	

^aChance accuracy would be 20% (each multiple-choice question had five possible answers).

all groups with above-chance accuracy. The accuracy trends differed across forms, however: Participants taking Form H answered literal and interpretive questions with almost equal accuracy (~46%), whereas participants taking Form G had a much easier time with interpretive items (47% to 54%) than with literal ones (30% to 33%).

Discussion

The aim of this study was to examine the validity of the reading comprehension portion of the NDRT (Forms G and H) by determining whether its multiple-choice questions could be answered accurately in a passageless administration format. Results showed that passageless accuracy rates were consistently above chance for university students with and without at-risk status for learning disorders. Accuracy rates were particularly high (near or above 50%) for science questions, possibly because the content was reportedly adapted from high school textbooks in subject areas that most college-bound students are required to take (e.g., chemistry). Given that the NDRT is intended for use with the college population (in addition to high school students), our findings raise

fundamental concerns about its validity. Consider that *test* validity is generally defined as the degree to which an assessment instrument measures the construct it is intended to measure. If the target population of a reading comprehension test can answer its questions effectively without having access to the text passages on which the questions were based, can that test be deemed a valid measure of reading comprehension?

The importance of the above question becomes magnified when one considers our finding that college students at risk for learning disorders (many of whom reported a documented history of services for learning disorders and/or attentiondeficit/hyperactivity disorder) were able to perform nearly as well as typically achieving peers when answering NDRT Form G questions in a passageless format (mean accuracy rate = 38.5%). As noted earlier, many of these students presented for evaluation specifically due to concerns about their ability to comprehend and keep up with college-level reading assignments. The most obvious and dangerous consequence of using a reading comprehension test with passage independence problems is the probability that it will overestimate students' functional reading comprehension skills and possibly lead to false-negative diagnoses and/or a lack of justification for reading-related modifications.

This problem may be compounded for certain types of high-ability students with disabilities—for example, university students with reading disorders who have an interest in studying law and hope to take the LSAT with appropriate accommodations. The LSAC (the council that oversees LSAT testing, including decisions about eligibility for accommodations) "prefer[s]" evaluators documenting cognitive disabilities to administer the NDRT. The LSAC (n.d.) provides the following scoring instructions in the achievement section of their Web-based *Guidelines for Documentation of Cognitive Impairments*:

A timed reading comprehension measure, which has been normed on adults and which allows for both extended and regular administration, is required. At the present time, the reading skills assessed, and the format utilized, by the Nelson-Denny Reading Test (NDRT) form G or H most closely match those measured on the LSAT; therefore, it is the preferred measure. However, the NDRT provides only education-based norms that result in inaccuracies when compared to age-adjusted aptitude measures such as the WAIS-IV. Thus, the NDRT should be scored twice for standard time conditions, using the actual grade level and, for comparison purposes, using the first-semester college norms. (p. 2, Section c)

Potential LSAT takers are likely to be 3rd- or 4th-year university students. Thus, scores based on first-semester college norms—regardless of how they are used—can only

Coleman et al. 249

exacerbate the skill overestimation trend that is inherent in the NDRT.

It is unfortunate that the NDRT is only the latest multiplechoice reading comprehension test that has been shown to have extensive problems with passage independence. Its flaws are particularly disturbing because it is likely (despite its outdated 1993 norms) the most widely used and accepted reading comprehension measure for the adult population with possible learning disorders. Contrary to what the extant literature might suggest, designing passage-dependent test items is not an impossible task (e.g., note our results for Passage 1 of NDRT Form H). However, it requires a more conscientious approach than has traditionally been adopted by test developers. For example, the use of fictionalized passages (as suggested in Lifson et al., 1984) would certainly minimize the extent to which examinees could answer questions based entirely on background knowledge. It would also seem logical to use passageless administration techniques during the test development process to identify (and subsequently eliminate or rework) items lacking passage dependence. Such techniques are neither complicated nor resource intensive. In closing, we echo the sentiments of Keenan and Betjemann (2006), who hoped that their explication of validity concerns about the GORT would "heighten the field's awareness of the gravity of the passage independence problem and lead to the development of better instruments for testing comprehension" (p. 378).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

Financial Disclosure/Funding

The authors received no financial support for the research and/or authorship of this article.

References

- Brown, J. I., Fishco, V. V., & Hanna, G. (1993a). *Nelson-Denny Reading Test (Forms G and H)*. Chicago, IL: Riverside.
- Brown, J. I., Fishco, V. V., & Hanna, G. (1993b). Nelson-Denny Reading Test: Manual for scoring and interpretation, Forms G and H. Chicago, IL: Riverside.
- Brown, J. I., Fishco, V. V., & Hanna, G. (1993c). Nelson-Denny Reading Test: Technical report, Forms G and H. Chicago, IL: Riverside.
- Daneman, M., & Hannon, B. (2001). Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT. *Journal of Experimental Psychology: General*, 130, 208–223.
- Fowler, B., & Kroll, B. M. (1978). Verbal skills as factors in the passageless validation of reading comprehension tests. *Perceptual and Motor Skills*, 47, 335–338.
- Katz, S., Lautenschlager, G. J., Blackburn, A. B., & Harris, F. H. (1990). Answering reading comprehension questions without passages on the SAT. *Psychological Science*, 1, 122–127.

- Keenan, J. M., & Betjemann, R. S. (2006). Comprehending the Gray Oral Reading Test without reading it: Why comprehension tests should not include passage- independent items. *Scientific Studies* of Reading, 10, 363–380.
- Law School Admission Council. (n.d.). Guidelines for documentation of cognitive impairments. Retrieved October 24, 2008, from http://www.lsac.org/pdfs/2008-2009/GuidelinesCognitive-2008.pdf
- Lifson, S., Scruggs, T. E., & Bennion, K. (1984). Passage independence in reading achievement tests: A follow-up. *Perceptual and Motor Skills*, 58, 945–946.
- Preston, R. C. (1964). Ability of students to identify correct responses before reading. *Journal of Educational Research*, 58, 181–183.
- Pyrczak, F., & Axelrod, J. (1976). Determining the passage dependence of reading comprehension exercises: A call for replications. *Journal of Reading*, 19, 279–283.
- Tian, S. (2006). Passage dependency of reading comprehension items in the GEPT and the TOEFL. *The Reading Matrix*, 6, 66–84.
- Wiederholt, J. L., & Bryant, B. R. (1992). Gray Oral Reading Test, Third Edition. Austin, TX: PRO-ED.
- Wiederholt, J. L., & Bryant, B. R. (2001). Gray Oral Reading Test, Fourth Edition. Austin, TX: PRO-ED.

About the Authors

Chris Coleman, MA, is a clinician and researcher at the Regents' Center for Learning Disorders at the University of Georgia. His current interests include psychometric aspects of assessing dyslexia and other learning disabilities in adults.

Jennifer Lindstrom, PhD, is a member of the faculty in the Department of Communication Sciences & Special Education at the University of Georgia. She has conducted research on issues related to high-stakes tests, accommodations, and students with high-incidence disabilities.

Jason Nelson, PhD, is a psychologist and researcher at the Regents' Center for Learning Disorders at the University of Georgia. His current research interests include the affective and motivational characteristics of students with learning disabilities and measurement issues related to the diagnosis of reading disabilities.

William Lindstrom, PhD, is a psychologist and researcher at the Regents' Center for Learning Disorders at the University of Georgia. His research interests include high-incidence disabilities, peer relations, test validity, and effort assessment.

K. Noël Gregg, PhD, is a University of Georgia Distinguished Research Professor and Associate Dean of Research in the College of Education. She has published widely on issues related to the assessment and accommodation of adolescents/adults with learning disabilities and ADHD.