

NELSON-DENNY

READING TEST



Technical Report Forms G&H

James I. Brown, Ph.D.
Vivian Vick Fishco, M.S.
Gerald Hanna, Ed.D.



THE RIVERSIDE PUBLISHING COMPANY
a Houghton Mifflin company

Copyright © 1993 by The Riverside Publishing Company. All rights reserved. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage or retrieval system without the prior written permission of The Riverside Publishing Company unless such copying is expressly permitted by federal copyright law. Address inquiries to Permissions, The Riverside Publishing Company, 8420 Bryn Mawr Avenue, Chicago, IL 60631.

Nelson-Denny Reading Test

Technical Report

Table of Contents

Part 1:	
Purpose and History of the Nelson-Denny Reading Test	1
Part 2:	
Development of Forms G and H	2
Part 3:	
Standardization of Forms G and H	8
Part 4:	
Technical Characteristics	11
References	25
Appendix A	
Schools in the Standardization Sample	(27)
Appendix B	
Standardization Questionnaire Summaries	(35)
Appendix C	
Item Difficulties (<i>p</i>-values)	(43)

Part 1

Purpose and History of the Nelson-Denny Reading Test

The ability to read well is so important in our culture that it is often the principal cause of success or failure from the first grade of elementary school through college and into professional life. Skillful reading must be mastered and, consequently, measured carefully and fairly.

The primary purpose of the *Nelson-Denny Reading Test*, Forms G and H, is to provide a trustworthy assessment of student ability in three areas of academic achievement: vocabulary, reading comprehension, and reading rate. These important skills are related and interdependent.

From 1929 to 1959, the original forms of the test (Forms A and B), developed by M. S. Nelson and E. C. Denny of Iowa State Teacher's College, served students and teachers as a measure of reading ability and as a guide for instruction. In 1960 Forms A and B were revised by James I. Brown of the University of Minnesota. The original format was retained, except for the addition of a measure of reading rate. The test still could be completed in its entirety in one class period. Forms C and D were completed in 1973 and Forms E and F in 1981.

Forms G and H of the *Nelson-Denny Reading Test* are the latest in a series of revisions of the test that was first administered more than 60 years ago. Content of the test and the statistical data have been periodically updated since the first test in 1929, but the format of this widely used reading survey test remains basically unchanged due to its widespread acceptance.

General Description

The *Nelson-Denny Reading Test (NDRT)* contains three measures: two subtests, Vocabulary and Comprehension, and a measure of reading rate. The Vocabulary test consists of 80 items, each with five answer choices, and has a time limit of 15 minutes. (Previous test forms consisted of 100 items.) The Comprehension test contains seven reading passages and a total of 38 questions, each with five an-

swer choices. (Previous test forms contained eight passages and 36 items.) The time limit for this section is 20 minutes, the first minute being used to determine reading rate. Total administration time for the entire test, including answer sheet preparation, is about 45 minutes.

For Forms G and H of the test, an extended-time administration is available for the first time. This mode of administration allows 24 minutes for the Vocabulary test and 32 minutes for the Comprehension test. Users who exercise the extended-time option should administer the test in two sessions, the first for completing the information required on the answer sheet and taking the Vocabulary test and the second for taking the Comprehension test. (The Reading Rate measure is omitted.)

Test Authors

The senior author of Forms G and H of the *Nelson-Denny Reading Test* is Dr. James I. Brown, Professor Emeritus in the Department of Rhetoric at the University of Minnesota. Dr. Brown has long been recognized as a principal figure in both test development and college reading circles. He is the author of previous forms of the *Nelson-Denny* and of several college texts for the teaching of reading. Multiple editions of his *Efficient Reading*, *Reading Power*, and *Guide to Effective Reading* attest to the wide use of these texts in colleges throughout the country.

Vivian Vick Fishco is English instructor and Director of the Learning Enhancement Center at Coconino County Community College in Flagstaff, Arizona. Ms. Fishco is a first-time author of the *Nelson-Denny Reading Test* and has collaborated with Dr. Brown on recent editions of *Efficient Reading* and *Reading Power*.

Dr. Gerald S. Hanna is Professor of Education at Kansas State University, Manhattan, Kansas. A notable scholar in the field of tests and measurement, Dr. Hanna was a coauthor of Forms E and F of the *Nelson-Denny Reading Test*.

Part 2

Development of Forms G and H

In 1989, prior to the preparation of Forms G and H of the *Nelson-Denny Reading Test*, 1,650 questionnaires were mailed to test users in all parts of the country and at all grade levels to solicit comments that could lead to improvements in the new forms. There was a response rate of 27 percent. Of this group, 74 percent of the respondents expressed satisfaction with the content and form of the test; therefore no major changes seemed desirable. Some minor adjustments were made, however, to bring the new forms into even greater accord with user comments.

To reduce working-time pressures, the Vocabulary subtest was shortened from 100 to 80 items, administration time remaining the same. To complete the 100-item Vocabulary subtest in the allotted fifteen minutes, it was necessary for test takers to complete 6.7 items per minute. The new, slightly modified format reduces the completion rate to only 5.3 items per minute. This change also brings a somewhat closer balance in weighting between the Vocabulary and Comprehension subtests.

In light of the questionnaire results, the Comprehension subtest was also slightly changed. Instead of eight selections, Forms G and H have only seven, but with five questions over each of the shorter selections instead of four as in the earlier forms. A total of 38 comprehension questions, instead of 36 as in the earlier forms, provides more items but with slightly reduced time pressure.

The questionnaire indicated rigid timing to be a particular problem for select populations, such as students with English as a second language or as a foreign language, or returning adults. It was clear that survey respondents felt that some students who take standardized tests do not perform commensurate with their abilities because time limits prohibit them from completing the test, or knowing they are rigidly timed causes them anxiety. In addition, it is believed that some students' cultural backgrounds dictate a more relaxed approach to tasks than the standard administration of the *Nelson-Denny* permits, while others are handicapped by rigid time constraints because of first language interference. As a result, the usual administration of the test may yield distorted results for such students.

A number of articles attested to the problem. "The Effect of Extra Time on Reading Comprehension Scores for University Students with and without

Learning Disabilities" (Runyan, 1991) reported research that found a significant difference between *Nelson-Denny* scores of students with learning disabilities and normally achieving students under timed conditions but no significant difference in test performance between these groups when students with learning disabilities are provided extra time to take the test. "Effects of Extending Administration Time on Standardized Reading Achievement Tests" (Baldwin *et al.*, 1989) recommended that *Nelson-Denny* test authors develop extended-time norms for the test. A decision was made for Forms G and H to develop an extended-time administration of the test with a special set of norms.

Test Development

In developing the items for Forms G and H, three stages of test development were employed: informal item tryouts, national item tryouts, and the national standardization. The following discussion will highlight these crucial steps in the development of the new *Nelson-Denny*.

Informal item tryouts in Spring of 1990 preceded national item tryouts in the fall. In the national item tryouts, eight forms of the test were administered to samples of students in grades 9, 11, 13, and 15. The samples included 7,300 students across these grade levels with approximately 900 students taking each of the eight forms. Males and females were almost equally represented with 54 percent females and 46 percent males. Minority representation consisted of 16 percent black, 4 percent Hispanic, 3 percent Asian/Pacific Islander, and 4 percent Native American. The analyzed data provided the foundation of Forms G and H.

Vocabulary Subtest Development. In developing the Vocabulary subtests for Forms G and H a two-fold procedure was used to ensure a minimum of 320 items from which to draw the 160 high-quality items needed for the final forms.

First, an initial pool of 766 words was created, drawn from current, widely used high school and college texts to ensure maximum relevance—a focus on words that must be known by students in order to cope successfully with school assignments. These items were then administered to high school, college, and adult subjects (the informal item tryouts) and item-analyzed using the Flanagan technique to permit pertinent comparisons with comparable data from the earlier forms.

The best 320 items were then selected for the nationwide formal trials that followed. For the first time in preparing new forms, it was possible to move the selection cutoff point to a new high—from 0.40 or higher to 0.45 or higher with all 320 items, according to the Flanagan statistics.

The national tryouts of these highly discriminating items were made at four grade levels—9th, 11th, 13th, and 15th. At this stage, different discrimination indices were calculated—a P-BIS r and BIS r for each test item at each of the four grade levels. From the 320 best items selected on the basis of the informal tryouts, a final selection was then made of the best 160 items for inclusion in final Forms G and H.

Tables 1, 2, and 3 summarize the Vocabulary test data for the items selected for final Forms G and H.

Table 1 shows the average difficulty (p -values) of vocabulary items in Forms G and H. The Vocabulary tests in Forms G and H are slightly less difficult than those in Forms E and F. But despite the eight-year range covered—grades 9 through 16—there are sufficient difficult items to challenge the superior student as well as easier items to discriminate among the ninth graders. Analysis reveals that 31 percent of the test items are at the 0.50 level or above, to minimize initial emotional tension resulting from overly difficult first items.

Table 2 summarizes data from the formal tryout on the range of difficulty of the vocabulary items at all four grade levels. Table 3 presents the P-BIS r and BIS r averages. Forms E and F averaged 0.42 on P-BIS r and 0.58 on BIS r , as compared with 0.41 (P-BIS r) and 0.58 (BIS r) on Forms G and H.

Table 1
Average Difficulty of Vocabulary Items from the Item Tryouts
for Forms G and H*

Form G	Form H
9th grade – 0.42	9th grade – 0.42
11th grade – 0.52	11th grade – 0.51
13th grade – 0.59	13th grade – 0.58
15th grade – 0.81	15th grade – 0.80

* Three items, revised after the national tryout, were not included in the averages.

Table 2
Range of Difficulty of Vocabulary Items from the Item Tryouts
for Forms G and H**

Form G	Form H
9th grade 0.11 – 0.90	9th grade 0.11 – 0.88
11th grade 0.15 – 0.93	11th grade 0.10 – 0.95
13th grade 0.21 – 0.95	13th grade 0.13 – 0.96
15th grade 0.37 – 1.00	15th grade 0.32 – 1.00

** Three items, revised after the national tryout, were not included in the ranges.

Table 3
Average Discrimination Indices (P-BIS *r* and BIS *r*) for Vocabulary Items
from the Item Tryouts for Forms G and H*

Form G			Form H		
Grade	P-BIS <i>r</i>	BIS <i>r</i>	Grade	P-BIS <i>r</i>	BIS <i>r</i>
9th	0.39	0.51	9th	0.41	0.55
11th	0.44	0.59	11th	0.44	0.58
13th	0.43	0.57	13th	0.45	0.59
15th	0.35	0.59	15th	0.40	0.63

* Three items, revised after the national tryout, were not included in the averages.

Comprehension Subtest Development. With the likelihood of a substantial spread of reading abilities at any one grade level, passages were chosen from widely different grade levels. Typically, incoming freshmen at most state universities range in reading ability from about the fifth grade to beyond the university senior level. Assessing such a range of abilities requires a test covering material that varies in difficulty.

To ensure maximum relevance, all short passages were drawn from texts currently in use at the high school and college levels. In addition, the texts from which passages were taken were the latest editions and among the most widely used in the country, as indicated by their number of editions. The six short test passages in each final form were drawn from six different texts: two social science texts, three science texts, and one humanities text. The first, longer selection is from a humanities text; its length brings a closer balance with the social science and science areas. Most science passages are from secondary-level texts; most social science passages are from college-level texts—again to achieve the best possible balance of difficulty.

The format of the two forms is similar to that used in previous forms. 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.

A critical requirement in building any test is a sufficiently large initial pool of items from which to make a final selection. If the pool is too small, items with less than desired discrimination must sometimes be selected to ensure adequate coverage. Even greater choice than was possible for Forms E and F was

arranged. Initially 32 passages were selected and used in the two tryouts—informal and formal. Of those, the best 14 were selected for Forms G and H.

For the informal item tryouts, many more items than needed were constructed to make sure items of sufficient quality could be selected for the second, more formal tryout. Table 4 displays the exact number of items prepared for each of the 32 passages for the informal tryout, for the national (formal) tryout, and for the final Forms G and H. Each tryout provided data on which to base the selection of items.

Eight test forms were used for the national tryouts. For the longer passages, the twelve best items were selected from the previous administration. For all the short passages, the eight items with the best psychometric properties were chosen for the national tryouts.

As a result of these formal tryouts, 14 reading passages were selected for final Forms G and H, seven for each form. Table 4 summarizes all stages in the selection process.

Approximately half the comprehension items devised for the initial, informal tryouts were 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. This two-fold classification of items was followed from informal tryouts to the final, published forms, each final form containing nineteen literal and nineteen interpretive items.

Tables 5 and 6 summarize data from the formal tryout on the average difficulty and range of difficulty of the comprehension items at all four grade levels.

Table 4
Number of Passages and Items Developed for
Comprehension Subtests

Passages	Informal Tryout				National Tryout				Final Forms	
	Passages				Passages					
	1	2	3	4	1	2	3	4	G	H
	Items				Items				Items	
I	20	25	23	21	12	12	12	12	8	8
II	16	16	16	16	8	8	8	8	5	5
III	17	16	16	16	8	8	8	8	5	5
IV	19	17	19	18	8	8	8	8	5	5
V	19	19	20	20	8	8	8	8	5	5
VI	16	16	16	16	8	8	8	8	5	5
VII	18	18	18	19	8	8	8	8	5	5
VIII	16	17	16	16	8	8	8	8		
Totals:	141	144	144	142	68	68	68	68	38	38
	32 passages 571 questions				32 passages 272 questions				14 passages 76 questions	

Table 5
Average Difficulty of Comprehension Items from
the Item Tryouts for Forms G and H

Form G	Form H
9th grade – 0.50	9th grade – 0.52
11th grade – 0.61	11th grade – 0.63
13th grade – 0.66	13th grade – 0.68
15th grade – 0.82	15th grade – 0.85

Table 6
Range of Difficulty of Comprehension Items from
the Item Tryouts for Forms G and H

Form G	Form H
9th grade 0.19 – 0.89	9th grade 0.15 – 0.85
11th grade 0.25 – 0.93	11th grade 0.23 – 0.91
13th grade 0.28 – 0.94	13th grade 0.31 – 0.94
15th grade 0.35 – 0.97	15th grade 0.38 – 0.98

Table 7 presents the P-BIS r and BIS r averages. Experience and a somewhat larger pool of items made it possible to improve still further on the high aver-

ages of preceding forms. Forms E and F averaged 0.43 on P-BIS r and 0.61 on BIS r , as compared with 0.45 (P-BIS r) and 0.64 (BIS r) on Forms G and H.

Table 7
Average Discrimination Indices (P-BIS r and BIS r) for Comprehension
Items from the Item Tryouts for Forms G and H

Form G			Form H		
Grade	P-BIS r	BIS r	Grade	P-BIS r	BIS r
9th	0.45	0.61	9th	0.48	0.63
11th	0.48	0.64	11th	0.51	0.68
13th	0.44	0.60	13th	0.46	0.64
15th	0.43	0.68	15th	0.38	0.63

The seven passages in each of the new forms were checked for readability by means of three measures: the corrected Dale-Chall Grade Level, the Fry formula, and the Flesch Reading Ease Score. While these measures of readability lack complete agreement, taken together they do provide an improved estimate of text difficulty.

Readability formulas do not always reflect the difficulty of the concepts expressed in a passage. Therefore, balance was sought in construction of the Comprehension test to offer passages with concepts as well as readability appropriate to the varying reading abilities of *Nelson-Denny* test takers. These passages were arranged in order from simpler to more complex.

According to the readability checks, Forms G and H are balanced in terms of difficulty. For each form, readabilities of passages range across the reading levels found among *Nelson-Denny* test takers, averaging at the upper high school levels.

Reading Rate. Devising a valid method of measuring rate within the scope of testing time available is a problem to which the test's developers have devoted special attention. Two methods of measuring rate are the amount-limit and the time-limit. In the amount-limit, the subject is given a passage to read in entirety, the rate computed on the time taken to read it. In the time-limit, the subject is given a passage and directed to read for a specified period of time. Results from either method are usually reported in words per minute. K. H. Humphry (1957) compared these methods and provided detailed information on the results of many different patterns of measurement.

Humphry's criterion test was 7,000 words long, and each reader was permitted to read until finished, a minute-by-minute measure of words read being obtained for each reader. Materials at the upper senior high school level were used with 166 students enrolled in the University of Iowa college reading program. Table 8 shows the results that were obtained.

Table 8
Means and Standard Deviations for
Subscores of Rate of Reading Test IB
(in words per minute)

	Mean	Standard Deviation
Total Test	366	107
1st minute	366	102
2nd minute	350	113
4th minute	364	108
8th minute	365	123
1st 4 minutes	363	101
2nd 4 minutes	362	110

Equivalent forms reliability coefficients of the subscores indicate a reliability of 0.73 for the first minute, 0.67 for the second, 0.72 for the fourth, and 0.72 for the eighth. Such experimental results seem to indicate that a one-minute time period does provide a suitably valid measure of reading rate. However, a further check was made with 123 students in the extension reading classes at the University of Minnesota to see how a reading rate determined by the *Nelson-Denny* one-minute timing would compare with the rate for the same students in the same class period reading a 1,300-word excerpt from *The Mature Mind* by Overstreet. Comparison of the two sets of figures showed an average of 275.9 words per

minute for the *Nelson-Denny* passage, using the time-limit method, and 274.4 words per minute for the longer Overstreet selection of comparable difficulty, using the amount-limit method. As with previous forms of the *Nelson-Denny*, the time-limit method was selected for measuring reading rate on Forms G and H. For additional details, consult the article by K. H. Humphry (1957) "An Investigation of Amount-Limit and Time-Limit Methods of Measuring Rate of Reading."

Measures to Ensure Balanced Treatment

In developing Forms G and H, a two-fold attempt was made to address the growing concern about the balanced treatment of minorities and genders in preparation of tests.

To eliminate or minimize bias, the first step was to analyze all 592 vocabulary and comprehension items used in the nationwide fall tryouts. The Mantel-Haenszel Measure of Differential Item Functioning was used; this is perhaps the most widely used procedure for identifying possible bias. Three analyses were performed on each item: black vs. white, Hispanic vs. white, and male vs. female.

These analyses led to the identification of 74 items that were flagged due to differential item functioning. Of these items, 49 were eliminated. The remaining 25 items had such high discrimination values that they were retained, so as not to diminish the reliability of the test. However, these remaining items were chosen so as to obtain the best possible balance between groups, thus minimizing favoritism in any one direction.

The second step to ensure fair treatment of minorities and genders was to select a panel of men and women representative of different ethnic groups and ask them to react subjectively to all items and to the reading passages used in the test. While, as might be expected, there was not perfect unanimity, comments of the panelists were helpful in selecting specific passages for use in the final test forms. Out of the 32 passages used in the national tryout, 18 were discarded. The remaining 14 finally selected for Forms G and H were those giving the best pos-

sible representation of ethnic and gender groups consistent with acceptable psychometric properties.

The members of the balanced treatment review panel:

Dr. Rebecca Bode, Chief Psychologist/Special Services, Cleveland Heights-University Heights City Schools, Ohio

Dr. Frank Ciriza, Researcher/Evaluator, San Diego Unified School District, San Diego, California

Dr. Valeria A. Ford, Director of Student Assessment, District of Columbia Public Schools, Washington, D.C.

Judy A. Hanks, Director of Indian Education, Cass Lake-Bena School District, Cass Lake, Minnesota

Carolyn Lee Hart, Language Arts Teacher, Atlanta Public Schools, Atlanta, Georgia

Pamela Hashimoto, Guidance Counselor, Kenwood Academy, Chicago, Illinois

Fred Hunter and Jeffery Bertrand, English Teachers, Whitney M. Young Magnet High School, Chicago, Illinois

Phat Mekdara, Adviser, Los Angeles Unified School District Division of Adult and Occupational Education Instructional Service Unit, Los Angeles, California; Chairman of Los Angeles County Refugee Communities; Advisory Vice Chair of Association of Korean Education

Dr. Steve Schrankel, Director of Research/Evaluation, Sunnyside Unified School District No. 12, Tucson, Arizona

Dr. Michael W. Strozeski, Director of Research, Garland Independent School District, Garland, Texas

Dr. Joseph H. Suina, Assistant Professor, University of New Mexico, Albuquerque, New Mexico

Dr. Veronica H. Zepeda, Director, Language Development, Roosevelt School District, Phoenix, Arizona

Following the implementation of these procedures to ensure balanced treatment in test development, standardization data were analyzed for differential item functioning. These analyses are summarized in the discussion of the evidence addressing test fairness in Part 4.