



Test and Answer Document Design and Layout

Betsy J. Case, Ph.D.

February 2003



Test and Answer Document Design and Layout

Introduction

Pearson Inc. (Pearson) drew from a variety of sources to create and validate its innovative test and answer document design for the *Stanford 10 Assessment Series (Stanford 10)*. First, Pearson reviewed the literature and empirical research conducted around the world. Second, Pearson applied lessons learned from assessment programs in key client states, including Arizona, Florida, Massachusetts, Texas, and Virginia. Third, within these state programs, Pearson conducted cognitive labs to try out modified designs. Fourth, Pearson utilized expert reviewers to review the content and graphic design of the test and answer document prototypes.

Literature Review

Pearson continually conducts research to provide evidence to support assessment validity. The literature review for this report was limited to the concept of *Maximum Legibility*. Legibility refers to the capability of being deciphered with ease.

Legible Text. Pearson adapted many of the constructs researched and reviewed by Gaster and Clark (1995) and Worden (1991). The major constructs included: 1) Black type on off-white paper provides the best contrast; 2) Typeface choice influences readability; and 3) Fixed-space fonts as well as increased leading (white space between lines) are recommended. Smith and McCombs (1971) noted that the use of blank space increases legibility.

Legible Response Formats. Pearson drew from researchers such as Grise, Beattie, and Algozzine (1982), who reported that placing answer options in a vertical format with good use of white space was useful for a variety of students, especially those with learning disabilities. Willingham, Ragosta, Bennett, Braun, Rock, and Powers (1988) substantiated the use of blank space on the answer document. Scruggs and Mastropieri (1992) also presented similar findings upon reviewing test and answer document formats.

Focus Groups, Cognitive Labs, and Research in Other Programs

Pearson utilized focus groups and cognitive labs from 1996 to 2000 to review prototype test and answer document designs. The layout of the tests and answer documents, use of white space and borders, and larger print appeared to assist students in responding to test items. These design features were incorporated into a pilot test conducted in Texas during May 2000 for the State-Developed Alternative Assessment (SDAA)—a test designed for students with learning disabilities.

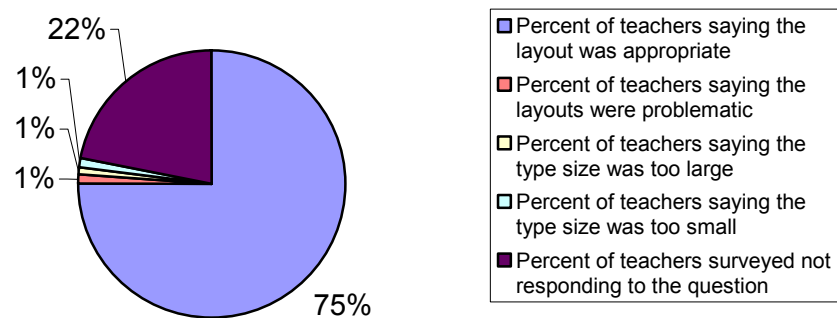
Test and Answer Document Design and Layout

Focus Groups. Pearson convened a series of focus groups during 1998 through 2000 to review and discuss the research results and identify the most inclusive designs for test booklets and answer documents. Researchers, practitioners, and state representatives were included in these focus groups. This technique proved to be quite influential in shaping the ongoing research activities.

Cognitive Labs. During the period of 1996 through 2000, Pearson conducted cognitive labs with over 1,300 students. During that process, students were asked to give verbal reports on test and answer document layouts using protocols such as those described in Ericsson and Simon (1993). The layout of the test booklets and the answer documents' use of white space and borders were substantiated by a majority of those students participating in grades K through 12.

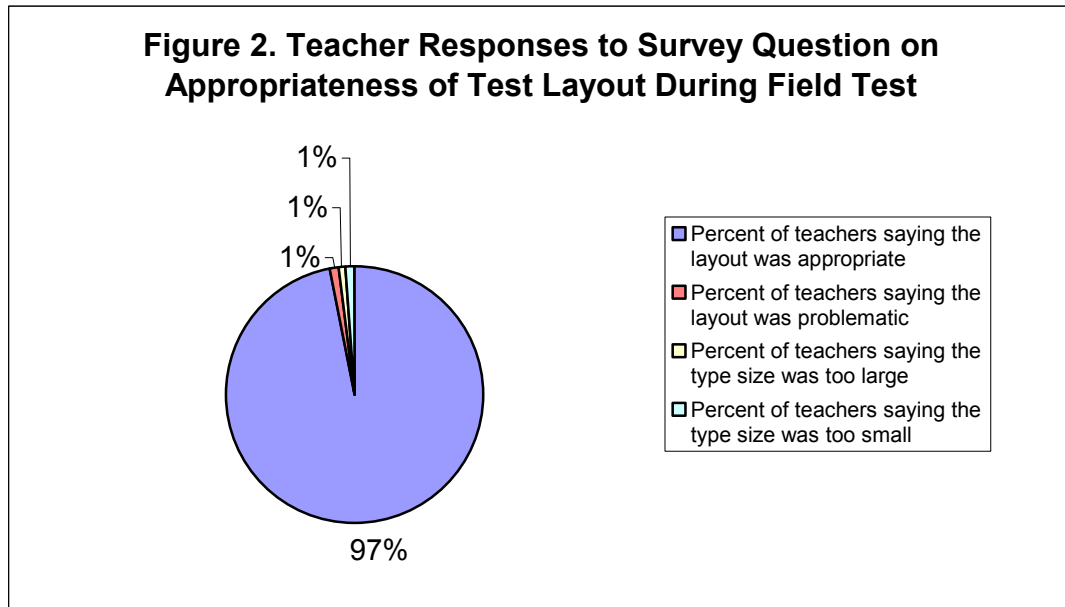
Pilot Study Results. Several key client states also explored the viability of test and answer document design, use of white space and borders, and larger print. These features were incorporated into a pilot study conducted by the Texas Education Agency during May 2000 for the SDAA. Over 1,000 students participated and a survey for teachers was conducted concurrently with the pilot study. Five hundred sixty-four of 732 teachers participating responded to the survey question about the appropriateness of the test layout. The majority of teachers participating—549 or 75%—said that the layout was appropriate. Other responses included 1% of teachers participating who thought the layout was problematic, 1% who indicated the type size was too large, and 1% who indicated that the type size was too small. The remaining 22% of teachers participating in the pilot study did not respond to the survey question. (See Figure 1 below.)

Figure 1. Teacher Responses to Survey Question on Appropriateness of Test Layout During Pilot Study



Test and Answer Document Design and Layout

Field Test Results. In the fall of 2000, the revised test and answer document designs were field tested statewide. In response to the accompanying survey and the question about whether the layout of the tests and answer documents was appropriate, 12,386 or 97% of teachers participating said the layout was appropriate. Of the 3% of teachers participating who dissented, 1% (186) said the layout was problematic, 1% (88) reported the type size was too large, and 1% (113) indicated that the type size was too small. (See Figure 2 below.)



Expert Reviewers. Pearson had a panel of experts review the design and layout of the *Stanford 10* tests and answer documents. Collaboration with experts as well as in-house research and analysis also guided the design process. Each test (grade) level was reviewed for style, layout, and appropriateness by the panel and in-house experts. Care was taken to ensure that the design for *Stanford 10* is appropriate for all students in content, format, and administration.

Follow-Up Data Analysis

Previous data analysis conducted by Pearson indicated that some students had difficulty maintaining the correct position on the answer sheet. That is, there was a frequently observed pattern on passage-related tests (such as the Reading Comprehension tests of both the Stanford and the Metropolitan series) in which omit rates would “spike” on the first item appearing for successive passages. Presumably, these students were losing their place either in the test booklet or on the answer sheet, and appeared to be omitting an answer in order to catch up. Generally, item omit rates are not strongly related to the difficulty of items, but get increasingly large over the length of the test. The “spiking” of omits on Reading Comprehension tests is manifested in omit rates that are relatively high on the first item associated with a given passage, dropping to a lower rate on the second item.

Test and Answer Document Design and Layout

Follow-up data analysis conducted with the publication of the METROPOLITAN8 and its newly formatted answer sheet indicated a much lower incidence of this spiking than on previous editions. Since the spiking of omit rates was most common with lower scoring students (i.e., those scoring in the first three stanines), this lower incidence is viewed as an outcome that should allow these students to provide a more accurate portrayal of their achievement.

References

- Center for Universal Design. (1997). *What is universal design?* Raleigh, NC: North Carolina State University.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: Massachusetts Institute of Technology.
- Gaster, L., & Clark, C. (1995). *A guide to providing alternate formats*. West Columbia, SC: Center for Rehabilitation Technology Services.
- Grise, P., Beattie, S., & Algozzine, B. (1982). Assessment of minimum competency in fifth grade learning disabled students: Test modifications make a difference. *Journal of Educational Research*, 76, 35–40.
- Scruggs, T. E., & Mastropieri, M. A. (1992). *Teaching test-taking skills: Helping students show what they know*. Brookline, MA: Brookline Books.
- Smith, J. M., & McCombs, M. E. (1971). Research in brief: The graphics of prose. *Visible Language*, 5(4), 365–369.
- Tufte, E. R. (1982). *The visible display of quantitative information*. Cheshire, CT: Graphics Press.
- Tufte, E. R. (1990). *Envisioning information*. Cheshire, CT: Graphics Press.
- U. S. Department of Education. (1997). National Institute on Disability and Rehabilitation Research. Washington, D.C.: Author.
- Willingham, W. W., Ragosta, M., Bennett, R. E., Braun, H., Rock, D. A., & Powers, D. E. (1988). *Testing handicapped people*. Boston: Allyn & Bacon.
- Worden, E. (1991). *Ergonomics and literacy: More in common than you think*. Indiana. (ERIC Document Reproduction Service No. ED 329 901).

**Additional copies of this document
are available from:
Pearson Inc.
1950 Bulverde Rd.
San Antonio, TX 78259
1-800-211-8378
1-877-576-1816 (fax)
<http://www.HEMWEB.com>**