

The Use of Separate Answer Sheets

As early as 1932, separate answer sheets for standardized tests were hailed as a financial boon for educational testing. Not only are the production costs of a single answer sheet substantially lower than those associated with producing an entire test booklet, but increased accuracy and speed in scoring provide additional financial gains (Petersen & Peterson, 1932). While financial accountability is one important aspect of large scale testing, a vital psychometric issue is raised: how do separate answer sheets affect the reliability and validity of student scores?

The answer to this question remains unclear for a variety of reasons. Very few studies have been conducted to explore this issue and the results of the few studies that exist are mixed. The existing studies employ small sample sizes and few follow-up studies have been conducted to explore findings more thoroughly. Other factors, like training, and its closely related counterparts of experience and student maturity, also may make it difficult to distinguish the impact of using separate answer sheets in standardized testing from other potential explanations of observed differences in student performance.

“How do separate answer sheets affect the reliability and validity of student scores?”

There seems to be consensus among test experts that third grade is the point at which students can answer test questions competently on an answer sheet. Cashen and Ramseyer (1969) demonstrated that 3rd

grade students could accurately respond on separate answer sheets, while 1st and 2nd graders could not. Other researchers extended these findings by showing that the reliability of scores is not adversely affected by the use of separate answer sheets for elementary students in 3rd and 4th grade. Specifically, Beck (1974) compared the performance of 3rd and 4th graders on math and English tests while observing no differences in average test scores. Similarly, Solomon (1971) studied 4th graders and found no differences between scores from tests that utilized separate answer sheets and test booklets on a reading test, even when controlling for gender.

Contrary to these findings, however, some researchers have reported that young examinees in grades two through six obtain higher scores when they provide answers to test questions directly in the test booklet (Muller, Calhoun, & Orling, 1967; Wise, Duncan & Plake, 1984). Muller et al. (1967) pointed out that the additional construct irrelevant variance that is introduced by the use of separate answer sheets may be attributed to a greater demand being placed on students in the areas of encoding, short-term-memory, and motor coordination skills. They found that the average number of errors for students taking an easy test was higher for those who recorded answers on the separate answer sheet than for those who recorded answers in a test booklet for 3rd, 4th, and 6th graders, but that the number of errors decreased as grade level increased.

Several of these researchers offer potential explanations for the observed differences in the results between studies. Specifically, factors that were hypothesized to impact differences in scores for different answer recording modes include sample size, student ability level, and experience using a separate answer sheet. For some studies,

sample sizes were small, ranging from 53 students in one study to 117 in another. As a result, the statistical power to detect real differences in student performance, should they exist, may be relatively weak.

Other researchers hypothesized that only low ability students might be affected by answer recording medium. The Wise et al. (1985) study found no differences attributable to answer recording medium when controlling for prior ability based on student performance on the previous year's test. In a similar fashion, Muller et al. (1967), in an attempt to control for student ability, removed students who scored less than 80% correct on the studied test on a previous test attempt from their sample because they suspected those students to be of low ability.

Finally, some researchers have suggested that experience may account for differences in student performance attributable to answer recording medium and that training students to use separate answer sheets might mitigate any observed differences. Wise et al. (1987), in a follow-up study, provided students with three 45-minute training sessions concerning the use of answer sheets on the three days prior to testing. They found that students who received training and answered with answer sheets performed comparably to students who did not receive training and recorded answers in test booklets for reading and math tests completed by 3rd, 4th, and 5th grade students, supporting the notion that explicit training on answer sheet use removes any potential answering medium differences. In addition to training, experience with using answer sheets increases naturally as students progress through the grades. Students also mature during these years and new cognitive and neurological processes emerge, so it is difficult to disentangle these effects when considering the impact

of separate answer sheets on student scores.

"Experience with using answer sheets increases naturally as students progress through the grades."

While an answer to the question "how do separate answer sheets affect the reliability and validity of student scores?" remains elusive, technology is beginning to move testing onto new platforms. Current research on computer-based testing is now focusing on the issue of how much test mode affects the reliability and validity of student scores. With computer interfaces that allow the examinee the opportunity to directly make responses onto a touch-sensitive screen, the availability of mixed media and varied response formats to make testing more compatible with educational goals, and the ability to score and report test outcomes immediately, answering on a separate answer sheet may become a more costly alternative that decreases accuracy and speed in scoring.

-- Shelley Ragland

REFERENCES

- Beck, M. (1974). Achievement test reliability as a function of pupil-response procedures. *Journal of Educational Measurement*, 11(2), 109-114.
- Cashen, V., & Raymseyer, G. (1969). The use of separate answer sheets by primary age children. *Journal of Educational Measurement*, 6(3), 155-158.
- Muller, D. Calhoun, E., Orling R. (1972). Test Reliability as a Function of Answer Sheet Mode.

- Journal of Educational Measurement, (9)*, 321-324 .
Peterson, J., & Peterson, H. (1932).
Reducing the costs of tests without
impairment of their value.
Transactions of the Kansas
Academy of Science, 35, 132-140.
- Solomon, A. (1971). The effect of answer
sheet format on test performance
by culturally disadvantaged fourth
grade elementary school pupils.
*Journal of Educational
Measurement, 8(4)*, 289-290.
- Wise, S., Duncan, A., & Plake, B. (1985).
The effect of introducing third
graders to the use of separate
answer sheets on the ITBS. *Journal
of Educational Research, 78(5)*,
306-309.
- Wise, S., Plake, B., Eastman, L., & Novak,
C. (1987). Introduction and training
of students to use separate answer
sheets: Effects on standardized test
scores. *Psychology in the Schools,*
24(3), 285-288.