The need established, the next step was to identify a course that could incorporate a project on horticulture internship programs. Seniors in horticulture at Clemson are required to take two semesters of seminar, a course which was limited to researching journal articles in the library and making oral reports in class. Why not allow one of the four required reports to be a problem such as "Strengthening the Horticulture Internship Program"? Several of our seniors, having been through the internship program, were well aware of both the strengths and weaknesses of the program. Two students, one who had been through the program and one who had not, elected to work together as a team on this problem. They contacted horticulturally oriented businessmen over South Carolina and across the nation to determine which businesses would be interested in having a Clemson horticulture major work in conjunction with the internship program. A list of interested employers was compiled, and the name, address, telephone number, job description, salary range, and other pertinent information concerning each job were outlined on a form designed by the students. These forms together with available pictures, pamphlets, and comments from prospective employers, were placed in a notebook in the department head's office for student and faculty reference. This resource notebook provided by the students now acts as an ongoing interface between the university and the horticultural industry.

In addition to providing this resource notebook, the students reworked old administrative forms. The new forms briefly describe the internship program, outline student, faculty and employer responsibilities, and explain the process and requirements for completing the internship program.

Finally, to publicize the internship program, the students gave several departmental presentations, advertised in newsletters, and organized a package for extension leaders to present at horticultural association meetings.

The students who put together this package received credit in a course entitled Senior Seminar. They gained experience working with people, organizing information, writing business letters, and speaking in public. In addition, they made contacts with potential employers and future colleagues.

## Conclusion

One word of caution to those undertaking a project with students at the interface between university and community. Keep "student" as the key word. Every project should be accepted and approached as a learning experience for the student, one that will provide a unique and challenging experience. When attacking an unsolved problem, student curiosity is aroused and with it a drive to satisfy this curiosity which works as a motivating force. When motivated, our undergraduate students have made significant contributions.

The potential exists for creating a strong edge at which universities and communities can meet, communicate, and work toward common goals. Undergraduate students are interested in and capable of acting as a link between the two. With supervision from teaching faculty, projects of various scope and impact can be completed that will benefit the university, the community, and the student.

# Effect of Item Order on Exam Scores

Larry A. Nielsen and
David L. Johnson

## Abstract

*The effect of question order on test scores was examined for a group of 289 students in a natural resources course. Four exams were assembled from the same questions by creating two orders of questions and then creating two page sequences for each order of questions. Comparisons between exams differing only in page order revealed a consistent bias toward lower page scores on the exams with more difficult initial pages. The bias produced differences in mean total scores of 5.9 percentage points for one pair of exams and 3.6 percentage points for the other. This case illustrates the influence of the examination environment on student performance.*

## Introduction

An integral part of education, whether in the classroom, or in continuing education programs (e.g., Denova 1979), is the evaluation of student response in the form of written tests. Test construction and administration, however, is perhaps that portion of the teaching-learning experience treated most casually by college instructors. Frequently, essay exams are created the night before a test and multiple-choice exams are composed hurriedly, without revision, to meet a crowded typing schedule.

Such nonchalance continues despite the knowledge that many factors affect the reliability of individual questions and entire testing situations. The relationship of questions to course objectives (Davis 1968), the subjective or objective question format (Wood 1970), the environment of the room (Green 1975), the order of questions (Nelson 1970), the cognitive levels addressed (Marshall and Hales 1971), the scoring method (Davis 1968), and the examination length (Green 1975) influence the way the test is taken by students. For example, the validity of a multiple-choice test item is influenced by length of the stem,

**Nielsen is a professor with the Department of Fisheries and Wildlife Sciences, VPISU, Blacksburg, VA 24061, while Johnson is with the School of Natural Resources, The Ohio State University, Columbus, OH 43210.**

number of wrong choices, placement of the correct choice, relative length of wrong and correct choices, inclusion of "all of the above" and "none of the above," and other factors (Nelson 1970; Wood 1970; Green 1975).

Without explicit attention to the factors that affect the testing environment, unintentional bias of evaluation may occur. This paper describes a situation in which the desire to make cheating difficult in a crowded room produced biased test scores for a portion of the class. Specifically, the structure of the exams permits analysis of the effect of item order on student performance.

## The Testing Situation

The test was administered to 289 students in an introductory course concerning natural resource ecology and management. Because every seat in the lecture hall was occupied, the orders of test questions were scrambled into four different sequences to make copying difficult (Nelson 1970). First, two random sequences of identical test items were assembled (Green and White Forms). Second, the pages of each test form were arranged randomly in two different sequences (Forms A and B). For this paper, test forms which had the same sequence of questions on each page, but different page sequences, are compared (i.e., Green Forms A and B are compared, and White Forms A and B are compared). Because seating was alphabetical and tests were distributed in a uniform pattern, the student groups taking different test forms were assumed to be equivalent and comparable. All students finished the exam within a 50-minute period.

Test items were objective, including multiple-choice questions, fill-in questions, short mathematical problems, and questions asking for specific short answers. The test contained 25 questions with varying point value; perfect score was 134. Each of two instructors graded approximately half of the test for each student. Although bias due to the order in which exam pages were graded is possible, the objective nature of the questions makes such bias unlikely. Average scores and variances for each page of each test form were computed. Differences between student performance means were tested statistically using one-tailed t-tests with a = 0.05. Because of the manner in which scores were recorded, statistical comparisons of total scores among test forms and standard item analysis for individual questions could not be performed.

## Results

### Green Test

The Green test consisted of six pages. Possible points were similar on five of six pages, ranging from 20 to 27 points, with only seven points being possible on the other page because it contained a large diagram of a food web. Forms A and B of the Green test differed only in page order.

**Table 1. Mean scores, variances, and page position of exam pages for green test Forms A and B. Asterisk indicates significant difference in mean scores (P v 0.05).**

| | | Green Form A (72 students) | | | Green Form B (74 students) | | |
|---|---|---|---|---|---|---|---|
| Page | Points Possible | Mean Score (Variance) | Percent Correct | (Page Position in Test) | Mean Score (Variance) | Percent Correct | (Page Position in Test) |
| 1 | 27 | 23.1* (10.7) | 85.6 | (1) | 21.7* (12.7) | 80.4 | (4) |
| 2 | 26 | 20.3* (17.0) | 78.1 | (2) | 18.5* (21.4) | 71.2 | (3) |
| 3 | 7 | 5.0* (3.5) | 71.4 | (3) | 4.4* (3.2) | 62.8 | (2) |
| 4 | 23 | 15.3* (13.6) | 66.5 | (4) | 14.0* (16.7) | 60.9 | (1) |
| 5 | 21 | 16.1* (0.0) | 76.7 | (5) | 14.7* (13.4) | 70.2 | (6) |
| 6 | 30 | 21.7* (19.3) | 72.3 | (6) | 20.2* (28.6) | 67.3 | (5) |
| Total or Mean | 134 | 101.5 | 75.8 | - | 93.5 | 69.8 | - |

Significant differences in average scores existed between Green Form A and Green Form B for every page on the exam (Table 1). Mean scores for Green Form A were higher for every page than for the same page in Green Form B. This consistent difference in mean page scores produced a 8.0-point (6.0-percentage point) difference in mean total scores for the exams of the two groups. Variances were tested and found to be equal in all comparisons (F-test, P v 0.05).

Performance on the exam was related to the difficulty (based upon average page score) of initial pages on the exam (Table 1). The difficulty of pages within the two Green tests differed greatly. Average scores, in percent correct, of the least and most difficult pages differed by 19.1 percentage points on Green Form A and by 19.5 percentage points on Green Form B. The easiest page of the Green exam (average 83.0% correct) was the first page on Form A and the fourth page on Form B. The most difficult page (average 63.7% correct), however, was the first page on Form B and the fourth page on Form A. Students taking Green

**Table 2. Mean scores, variances, and page positions of exam pages for white test Forms A and B. Asterisk indicates significant difference in mean scores (P v 0.05).**

| | | White Form A (73 students) | | | White Form B (70 students) | | |
|---|---|---|---|---|---|---|---|
| Page | Points Possible | Mean Score (Variance) | Percent Correct | (Page Position in Test) | Mean Score (variance | Percent Correct | (Page Position in Test) |
| 1 | 14 | 11.4* (4.0) | 81.4 | (1) | 10.7* (5.1) | 76.4 | (3) |
| 2 | 32 | 24.0 (11.7) | 75.0 | (2) | 23.1 (10.5) | 72.2 | (4) |
| 3 | 28 | 21.9* (18.6) | 78.2 | (3) | 20.6* (19.9) | 73.6 | (5) |
| 4 | 35 | 26.5* (32.5) | 75.7 | (4) | 24.6* (34.0) | 70.3 | (2) |
| 5 | 25 | 18.3 (14.7) | 73.2 | (5) | 18.2 (17.9) | 72.8 | (1) |
| Total or Mean | 134 | 102.1 | 76.2 | - | 97.2 | 72.5 | - |

Form B obviously were confronted with much more difficult material at the beginning of the exam than students taking Green Form A.

**White Test**

The White test consisted of five pages, four of which contained between 25 and 35 possible points. The other page, again containing the large diagram, contained 14 possible points. Forms A and B of the White test differed only in page order.

Difference in average scores between White Form A and B were generally smaller than between forms on the Green test (Table 2). Total mean scores differed by 4.9 points (3.6 percentage points), about two-thirds of the difference for the Green test. Mean scores for three pages in White Form A were significantly larger than scores for the same page in the White Form B. For these three pages, the differences were less than one point. Variances were tested and found to be equal for all comparisons.

The absence of significant differences between performance of the two groups taking the White test presumably resulted from closer similarity in the question-difficulty sequence for White Forms A and B. Average scores for the least and most difficult pages in White Form A differed by 8.2 percentage points and in White Form B differed by 6.1 percentage points. Despite the similarity among pages on the White forms, mean total score for White Form A, which began with the easiest page (81.1% correct) was 4.9 points higher than for White Form B, which began with two relatively more difficult pages (70.2 and 72.8% correct).

## Discussion

Studies of the effect of question sequence on exam performance have produced contradictory results (Marshall and Hales 1971). Brenner (1964) and Marso (1970) found no differences among random, easy-to-hard, and hard-to-easy sequences for multiple-choice exams, but Towle and Merrill (1975) reported that easy-to-hard sequencing resulted in higher average scores. Towle and Merrill suggested that the inability of some students to finish the exam may have affected their results, because the average number of items completed was lower in the hard-to-easy sequence than in the easy-to-hard. For the test presented here, however, all students finished the exam within the allotted time. The discrepancy of these results intensifies our belief that the testing situation is influenced by many factors, some of which can be anticipated and controlled, and some of which cannot.

It is not the purpose of this paper to provide a system for eliminating possible biases, but to stress the need for teachers to devote additional effort to preparation and administration of testing materials. There are, however, several approaches that can help to reduce possible variation in student performance. In a situation like the one described here, an alternative to using different question orders is to use two different styles of answer sheets (Nelson 1970). For example, answer spaces on one sheet may be arranged vertically while answer spaces on the other are arranged horizontally.

Suggestions for the order of arranging test items vary among test construction texts. The most common rule is to begin the test with a few relatively easy questions to reduce anxiety among students (Stanley 1958; Marso 1970; Nelson 1970; Denova 1979). Different authors suggest either that questions be grouped by subject matter, especially if the results will be discussed in class (Nelson 1970) Denova 1979), or that questions be grouped by format (e.g., all true-false together; Stanley 1958; Green 1975).

The need for special care during test construction appears especially relevant to broad environmental and natural resource courses. Subject matter in such courses includes a mix of objective and subjective facts and concepts that require a variety of question types. Combinations of numerical problems (e.g., population changes), multiple-choice items (e.g., effects of oil spills), and conceptual questions (e.g., morality of endangered species legislation) are likely to appear on most exams. For this reason, we recommend that teachers of renewable natural resources courses review a test construction guide, such as one of the cited references, before preparing each exam and that teaching assistants with testing responsibilities be required to read such a guide as part of their assigned duties.

## References


Brenner, M. H. 1964. Test difficulty, relaibility, and discrimination as functions of item difficulty. **Journal of Applied Psychology,** 48, 98-100.

Davis, R. A. 1968. Planning teacher-made tests. **Peabody Journal of Educational Measurement,** 46, 155-158.

Denova, C.C. 1979. **Test Construction for Training Evaluation.** New York: Van Nostrand Reinhold.

Green, J. A. 1975. **Teacher-made Tests.** Second edition. New York: Harper and Row.

Marshall, J. C., and L. W. Hales, 1971. **Classroom Test Construction.** New York: Addison-Wesley.

Marso, R. N. 1970. Test item arrangement, testing time, and performance. **Journal of Educational Measurement,** 7, 113-118.

Nelson, C. H. 1970. **Measurement and Evaluation in the Classroom.** London: MacMillan.

Stanley, J. C. 1958. ABC's of test construction. **NEA Journal,** 47, 224-226.

Towle, N. J., and R. F. Merrill. 1975. Effects of anxiety type and item-difficulty sequencing on mathematics test performance. **Journal of Educational Measurement,** 12, 241-249.

Wood, D. A. 1961. **Test Construction.** Columbus, Ohio: Charles E. Merrill Books.