SOCIOLOGY IN THE CLASSROOM: SINGLE AND PAIRED TEST PERFORMANCE

Wilbert M Leonard, Illinois State University
James J Berry, Western State College

INTRODUCTION

In the past quarter century there was a plethora of innovations in teaching, including television instruction, programmed instruction, authoritarian and non-authoritarian techniques, and large and small classes, all with varying degrees of success. Here we will highlight 1) the meaning and operationalization of *success* and 2) note that regardless of technique, the outcomes are very similar.

It is an axiom in measurement and evaluation that the first step is to define what is to be measured, and to specify a set of operations which captures this dimension. To measure such concepts as *enlightened citizenry, success, intellectual, emotional and social growth* and the like, they must be precisely defined, and operationalized. It seems that educators have been remiss about the end-products of the learning process, and that they have been haphazard in measuring presumably important skills. Often the consequences of learning and of individual courses have been imprecisely measured merely to satisfy bureaucratic requirements. For example, in the typical statement of university goals to serve the citizens of the state and nation through communication of knowledge, no mention is made of the relation of this goal to application of the grading system: A = superior; B = above average; C = average; D = below average; F = failing (Illinois State U 1974-75). The connection is presumably self-evident!

Performance evaluation is intended as an aid to decision-making for the best interests of the student and society. In testing literature, *aptitude tests* are designed to measure the student's *potential* to profit from certain types of experience and is assessed *before* exposure to instruction, and *achievement tests* are used to evaluate what a student has accomplished *after* the instruction. The typical course examination is an assessment of the knowledge which the student has acquired. Numerous hidden dimensons in this process should be revealed.

The most basic step, of defining objectives and goals is often unstated or not carefully thought out. It is only haphazardly measured. We should make the objectives explicit in terms of processes or activities: What is a student expected to display if the objectives have been achieved? The failure to ennumerate such objectives in terms of student behavior accounts for the inadequacy of evaluation and some criticisms of higher education. Even if goals are stated, other evasive dimensions may exist. Some objectives *cannot* be measured within the time constraints of a single semester or even over a four-year college tenure. Objectives cannot always be related to observable behaviors.

Actual testing format exists in two rubrics: 1) essay, and 2) objective tests. Educational researchers generally agree that neither category is satisfactory as the *sole* type of measure for assessing academic achievement. Each has unique advantages and disadvantages.

For the question, "Are there certain techniques of teaching which have been demonstrated to facilitate accomplishment of educational goals?" Extensive experimentation has shown that no significant differences have been found consistently. No one method appears to be categorically superior to others. Of course success with different techniques may vary for individual teachers, but in general, the teaching/testing/learning gestalt does not vary regarding measures of acquisition of knowledge and academic course skills.

THE PROBLEM

In designing our experiment it was necessary to review the factors involved in learning. Educational psychologists Morgan and King (1971) isolate three configurations: 1) individual factors; 2) learning methods; 3) meaningfulness of material to be learned. Category 2 includes amount of practice, knowledge of results, reading vs. recitation, and whole vs. part learning. Category 3 deals with relations between learning and personal/social relevance of content. We did not control for Categories 2 and 3. Category 1 provided the necessary background information for our design.

Five *individual* factors have been shown to be influential in the learning process:
1) intelligence; 2) chronological age; 3) motivation; 4) previous learning; 5) anxiety.

1) Learning and innate intelligence display a positive mutual association. In general, the higher the I.Q. the more rapidly material is grasped. But learning is not always related to intelligence. Consequently it is inappropriate to define learning ability only in terms of intelligence (Woodrow 1946),

2) After roughly 5 years of age, learning ability increases up to about 17-20 years of age, and then plateaus and drops slightly, up to age 50. After after age 50, the ability to learn new material declines (Morgan, King 1971 118).

3) The desire to learn, or *motivation* is directly correlated with learning ability. Psychologists differentiate incidental from intentional learning, and have discovered through research that very rarely does incidental learning approach the magnitude of intentional learning.

4) Previous learning, termed "learning how to learn" relates to acquisition of new materials. Educational researchers treat this component under the heading of *transfer of learning*. Previous experience can either facilitate or hinder new learning, and is termed either *positive transfer* or *negative transfer*.

5) Anxiety has had differential effects on learning depending on whether the task is simple or complex.

In an experiment with a simple task, of eyelid conditioning, high-anxious subjects were more rapidly conditioned than low-anxious subjects (Spence 1964). In the complex task of preparing for college examinations, anxiety has been shown to be a hindrance. Spielberger (1962) compared two groups of college students, of which one had high anxiety scores and the other had low anxiety scores. He subdivided each into five subgroups based on scholastic aptitude test scores. College grade point averages were subsequently correlated with aptitude test scores. He found a high positive association between grade point averages and aptitude test scores. Among students at the extremes — low aptitude and high aptitude, anxiety level made little difference. Both high and low anxiety students with low aptitude did equally poorly. Both high and low

anxiety students with high aptitude did equally well. But the middle three groupings showed marked differences. Low anxiety students performed significantly better than high anxiety students.

METHODOLOGY

Our research problem was to assess the relation between test-taking format by pairs or singly, measuring achievement by objective course examination scores, and anxiety by the students' self reports. Of the five individual factors known to influence learning we held constant intelligence as measured by grade point average, previous exposure to formal course requirements, and anxiety level. Motivation was not controlled. The presumed causal variable was test-taking format.

To answer the research questions, we used a modification of the *classical randomized experiment* (CRE) for the research design. This procedure assured maximum control over the salient variables. The independent causal variable was *test-taking* and the dependent or effect variables were 1) *test score*, or number of correctly answered objective test items, and 2) *anxiety level*.

The first day of class we handed out a conventional biographical sheet requesting such responses as name, year in school, campus address, grade point average, and general anxiety level surrounding the taking of examinations. A random check of reported grade point averages with university records indicated high validity of the responses. To operationalize anxiety level, a 5-point scale ranging from low to high anxiety and scored 1-5 was used. There was no independent criterion used to check the validity of reported anxiety levels, and we assumed face validity of the scale and responses. These two variables were made operational via self reports of 133 students in two separate classes on deviant behavior taught by the same instructor.

To test the hypotheses and avoid methodological pitfalls it was necessary to use two groups which were as identical as possible for experimental and control groups. It was important to assure similarity of the two groups so that any performance differentials would not be due to intrinsic initial differences between the groups. For this purpose we

used random allocation of students to experimental and control groups with one modification. Instead of dividing the total of 133 students into two nearly equal groups, we randomly assigned two students as an experimental pair for each student assigned to the unpaired or "single" group category. Individuals were randomly assigned to experimental and control groups with intervention to produce a nearly equal number of pairs and individuals in the two groups.

To assess the effect of pairing, it was necessary to establish guidelines for pairing. Since a this was a sophomore level course requiring only one introductory course in sociology as a prerequisite, both groups had about the same formal exposure to sociology. The criterion used for devising the paired group was similarity of grade point average. We wanted to avoid the confounding effects of dissimilarity in scholastic achievement between paired students.

These experimental operations were completed during the first two weeks of the term, without the students' awareness of our research project. After establishing the experimental and control groups, we checked statistically to see if the two groups were similar in grade point average and general anxiety level. The results, shown in Table 1 indicate a t-value of 1.39 which indicates a probability of .93 that the two groups are the same on these measures.

It is salient to repeat that the students' data in both classes were aggregated, so that separate comparisons were not made. Other controls to eschew confounding effects included 1) both classes received the same lectures from the same professor; 2) both used the same textbook; 3) both took identical objective examinations. The controls minimized the effects of extraneous factors. The research design is as shown in Figure 1.

FINDINGS

The statistical data for the results of the first examination, given at midterm, are summarized in Table 1. The experimental group correctly answered 2.5 more questions than the control group, with slightly less dispersion, and a t-value of 2.22, (p = .025) for the difference. It is not likely that sampling variation could account for the difference, and we conclude that

the difference is real. A series of Pearson product-moment coefficients (r) were calculated to contrast the experimental and control group results. The correlation between test scores and anxiety was significant and negative for the experimenal group (r = − .21; n = 86; p = .05), and significant and positive for the control group (r = + .33; n = 47; p = .02). The correlation values are not high, but to a limited degree, this means that for the singles of the control group, the higher the score, the higher the anxiety, and conversely, for the paired experimental group, an inverse relation is revealed. The higher the score, the lower the anxiety level. Exactly why this is the case leads to some speculation. Perhaps if students in the paired situation are ill-prepared, their anxieties may be shared through social contagion so as to compound the mental stress of both partners. Schacter's (1959) research on social facilitation and social inhibition suggests that if one is not prepared to master a task, as in lack of preparation for an examination, then the presence of another person would act detrimentally to performance. This possibility applies to interaction vis-a-vis co-acting effects and the experimental condition certainly is one which would elicit interaction effects.

The relation between anxiety and grade point average for the single and paired groups was − .10 and − .08 respectively, and suggests no measurable association. The temporal ordering of these variables has been explored by educational psychologists. Research has suggested that a reciprocal effect. Initial anxiety can inhibit examination performance, and poor examination performance can in turn increase anxiety level. Our results would tend to support this "vicious circle" contention.

The Pearson r correlations between test score and grade point average were statistically significant at + .37 (p = .001) for the experimental group and + .33 (p = .02) for the control group. Thus it appears that examination scores and grade point averages are positive and significant, as reported by Spielberger (1962).

We now turn to a comparison of the association values for the two aggregates. We ask, "What are the significant differences between the experimental and control gorups in the relations between 1) examination scores and
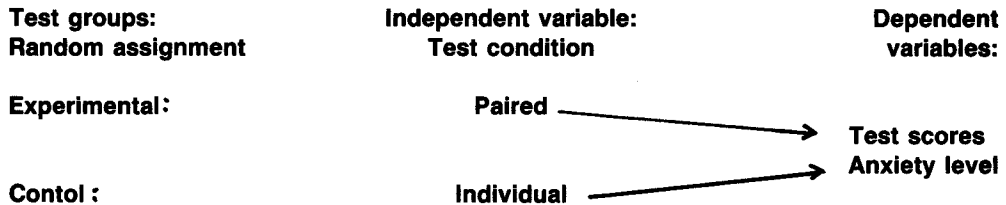
FIGURE 1: EXPERIMENTAL DESIGN

Test groups: **Independent variable:**   **Dependent**
Random assignment   **Test condition**   **variables:**

Experimental: Paired   **Test scores**
                                                                        Anxiety level

Contol :   Individual

TABLE 1: COMPARISON OF SINGLE AND PAIR GROUPS

| | Midterm examination | | | | | Final examination | | | |
	Single n#47	43 pairs n#86	t	p		Single n#47	43 pairs n#86	t	p
	Mean SD	Mean SD				Mean SD	Mean SD		
Test score	35.8 6.17	38.2 5.57	2.22	.03		66.7 10.55	72.8 7.86	2.88	.01
Anxiety	3.4 1.01	3.2 .82	1.39	ns		3.7 .96	3.2 .68	1.97	.05
GPA	2.5 .55	2.6 .71	.73	ns		2.5 .55	2.6 .71	.00	ns

anxiety levels, 2) anxiety levels and GPA's (grade point averages), and 3) examination scores and GPA's?" Does the independent variable of examination format, single versus paired, affect the correlations among the dependent variables of examination score, anxiety, and GPA? To do this statistically we used Fishers z-transformation for the correlation coefficients (Champion 1970). This transformation converts the correlations for

r-value range, − .99:0.00: + .99
to z-score range, − 2.65:0.00: + 2.65
by Algorithm (1).

(1) $z_r = .5 \log \left[(r+1) / (r-1) \right]$

(2) $s_{z1-z2} = \left[(n_1-3)^{-1} + (n_2-3)^{-1} \right]^{.5}$

(3) $z_{z1-z2} = (z_{r1} - z_{r2}) / s_{z1-z2}$

The standard deviation of the mean difference between z_1 and z_2 is given by Algorithm 2, and is a function of the number of cases for each correlation. The z-value for the difference, as determined by Algorighm 3 is evaluated in the cumulative table for the normal curve. If it reaches 1.96, the difference is significant at the .05 level of confidence. This identifies a significant difference between two correlation coefficients, r_1 and r_2. The results or our analysis are shown in Table 2.

The independent variable of test format did *not* produce differential statistically significant outcomes. The magnitudes of correlations coefficient differences for the various comparisons could have been due to sampling variation.

The final examination for the course provided data for replication of the study. The final examination contained twice as many items as the midterm examination. A statistically significant difference was shown between the paired experimental group and singles group, as the experimental group averaged six more correct items than the control group, or a difference significant at the .01 confidence level. The performance of the experimental group was also less variable, as shown by the smaller standard deviation (Table 1).

The self-reported anxiety level was lower for the experimental group at the time of the final examination (p = .05). The paired condition appears to reduce anxiety in conjuction with improved performance on the examination.

The correlation, r of examination score and anxiety was + .33 for the control group and − .21 for the paired experimental group. Both coefficients are statistically significant, and it is therefore noteworthy that they are opposite in sign. For both groups there is a high and significant correlation between examination score and grade point average (GPA).

Let us compare the final examination results with those of the midterm examination.

TABLE 2: CORRELATION DIFFERENCES BETWEEN SINGLE AND PAIRED GROUPS

First examination Variables	Singles Group				Paired Group					
	t	p	r	z_r	t	p	r	z_r	z_{z11z2}	p
Test score: Anxiety	2.36	.025	.33	.34	1.97	.05	−.21	−.21	2.92	.002
Test score: GPA	2.36	.025	.33	.34	3.65	.001	.37	.39	0.27	ns
Anxiety: GPA	0.68	ns	−.10	−.10	0.74	ns	−.08	−.08	0.12	ns
Final examination										
Test score: Anxiety	2.36	.025	.33	.34	2.17	.025	−.23	−.23	3.03	.001
Test score: GPA	5.58	.001	.64	.76	8.07	.001	.66	.79	0.20	ns
Anxiety: GPA	2.84	.005	.39	.41	0.37	ns	−.04	−.04	2.09	.025

The same pattern was discovered with one exception. For the singles group association coefficient between anxiety and grade point average, the correlation was larger, and in the opposite direction. This means that high anxiety and high GPA and low anxiety and low GPA are co-related. Here we must speculate. It may be that those with high GPA's become more tense about maintaining a good record because the final examination has a major effect on the course grade. This grade comes out almost immediately, whereas the midterm examination score is less salient, since its effect on the course grade is more remote.

Regarding the impact of test-taking, the same configurations emerge. The correlations between examination score and anxiety were significantly different for the control and experimental groups as shown in Table 2. The exception was the Pearson r's for anxiety and grade point average. The difference between the correlations for control and experimental groups is significant (Table 2: z = 2.19; p = .025). Substantively, this means that among the single test takers, the higher the anxiety, the higher the GPA, and that for the paired group, the relation is negligible. The practical importance of this is that when students take examinations in pairs, the students with high grade point averages are not hindered by a negative psychological state, whereas it does affect high GPA students taking the examination alone. It appears that a double benefit accrues in the paired experimental condition. The paired group not only performed better, but also experienced less anxiety, while their single counterparts performed more poorly and felt more psychological tension.

CONCLUSIONS

The testing system in the United States usually requires that one individual must work alone, respond to written questions within a fixed period in utter silence. It is assumed that this procudure will result in a reasonably accurate reflection of content mastery or acquired knowledge, and that it permits a valid ordering of individuals in an overall hierarchy of performance. One of the basic claims that sociologists make is that human beings are social, and that they "naturally" operate in interpersonal contexts. Taking this perspective as a starting point, the technique of testing and the validity of methods used in testing become problematic. There are at least three sets of variables that could and probably do intervene between ability and test performance under typical classroom testing conditions: 1) mechanical skills (reading, understanding and applying instructions); 2) psychological states such as anxiety and fear; 3) the perception that the experience is relevant or irrelevant.

The "blackout" noted by some students as they approach an examination is a possible factor in poor test performance even when actual ability among different individuals is held constant. Some students are better able to manage their anxiety than others, and as a partial consequence, leave the examination with higher scores, even though both are equally well prepared. Regarding relevance, one more frequently hears complaints from students that their examination experiences lack meaning, and they they are merely taking part in a ritualized demonstration of their regurgitational ability. If so, we could expect students to be less concerned.

Whereas students working alone are more susceptible to intervening problems, pairs of students could prove less vulnerable. Mechanical errors, anxiety, and relevancy of the testing experience are the kinds of dilemmas that appear to be correctable by cooperation between two problem solvers. The consequences of pairing seem to be more beneficial than harmful. Most students reported a reduction in anxiety and an improvement in learning experience. Both paired student examination situations produced higher grades with some improvement arising in the lower third of the GPA range.

Three students commented as follows:

I thought the pairing was beneficial. Each partner contributed equally. When one didn't know the other one usually did know the answer. Do it again! We need more things like this at this university.

Taking the test in pairs was actually not like taking a test at all ... but it was a learning experience in that we each got a chance to see another person's way of reasoning with practical problems dealing in application of material. Very good experiences; recommend using it again!

It's a form of double learning experience because you can actually retain more because of the discussions.

Our educational system sorts out people in a number of ways. In the end, evaluations depend on ability to execute written examinations. This emphasizes the machine-like qualities of memory while ignoring the social functions of communication, cooperaion and interaction in problem solving. Focusing evlauation on the individual ability to perform alone does not take into account the unnatural problems generated by this very nonsocial endeavor. Perhaps few of us really operate well alone, and further, very little in life is structured to exclude social interaction. If human beings are "social" then our preparations and education should be oriented to developing social skills in problem solving.

REFERENCES

Berry James J, Wilbert M Leonard II 1974 A different approach to classroom teaching. *Today's Education* 63 93

Bloom Benjamin S 1966 Twenty five years of educational research. *Amer Educ Res J* 3 211-221

Champion Dean J 1970 *Basic Statistics for Social Research.* Scranton Penn. Chandler

Horton Paul B, Gerald R Leslie 1974 *The Sociology of Social Problems.* Englewood Cliffs NJ. Prentice Hall

Morgan Clifford T, Richard A King 1971 *Introduction to Psychology.* New York. McGraw Hill

Schacter Stanley 1959 *The Psychology of Affiliation.* Stanford U Press

Sexton Patricia C 1967 ed *Readings on the School in Society.* Englewood Cliffs NJ. Prentice Hall

Spence K 1964 Anxiety (drive) level, performance in eyelid conditioning. *Psych Bltn* 61 129-139

Spielberger C D 1962 Effects of manifest anxiety on the academic achievement of college students. *Mental Hygiene* 46 420-426

Woodrow G 1946 The ability to learn. *Psych Rev* 53 147-158