*Bulletin of the Psychonomic Society*
*1981, Vol. 17(6), 261-263*

# Some differences among students volunteering as research subjects

DAVID O. RICHTER, SANDRA D. WILSON, MICHAEL MILNER, and R. J. SENTER
*University of Cincinnati, Cincinnati, Ohio 45221*

**Two samples of introductory psychology students yielded significantly different performances on serial learning and symbol substitution tasks, in spite of the fact that every effort was made to treat the groups identically. These results bring into question a commonly held assumption that different samples taken from the introductory psychology "population" are statistically equivalent.**

One of the few unqualified assertions experimental psychologists can get away with without serious contestation from their colleagues is that we collect a lot of data from introductory psychology students. By way of quantification of this self-evident truth, Smart (1966) reviewed four volumes of the *Journal of Experimental Psychology* and the *Journal of Abnormal and Social Psychology* and found that 85.7% of the studies reported in the former journal and 73% reported in the latter involved the use of college students, "chiefly male students enrolled in introductory psychology" (p. 115). No matter how we may dignify our sampling of experimental subjects from student "subject pools," samples so drawn are, as pointed out by Jung (1969), clearly not representative of the greater noncollege population, the population of all college students, or even the population of college students at a particular institution. Often, our method of dealing with this chronic bias in much of our research is to acknowledge a restriction of the generality of our findings to the greater population. But, perhaps, a more important sampling question exists. Are different samples taken from our convenient source of student subjects really samples from the same population, whatever the parametric nature of that population might be?

A survey of the existing literature indicated that little concern, in the form of published papers, has been given to the problem of subject sampling from the introductory psychology population. Smart (1966) questioned the wisdom of basing so much of our psychological research on such an atypical sample of the general population. Jung (1969) shared this concern and, in addition, questioned both methodological and ethical aspects of our use, or overuse, of college students in our research.

It would appear that very little in the way of empirical research has been directed toward the evaluation of the suitability of college student subjects for the types of research we commonly carry out. In their 1964 paper, Underwood, Schwenn, and Keppel reported no significant performance differences on a short paired associate learning task among Northwestern University students selected at five different temporal points during each of 2 sequential academic years. These authors concluded that any "selective factors associated with differing performances . . . are quite in balance throughout the quarter" (Underwood et al., 1964, p. 225).

In a study designed to investigate differences among student research participants "volunteering" at the beginning, middle, or end of an academic quarter, Richert and Ward (1976) used both hidden figures ("interesting task") and visual search for specific digits embedded in an array of random digits ("boring task"). The results of this study both confirmed and contradicted those of Underwood et al. (1964). No significant differences were found across the temporally sampled groups for the hidden figures task, but the analysis of the visual search task data showed a significant ($p < .004$) difference apparently associated with a lower performance level exhibited by the subjects who had "volunteered" late in the quarter. Although Richert and Ward offered several quite feasible suggestions to account for their subjects' differential performance on the two tasks, no unequivocal explanation was apparent.

Because of the relative dearth of research designed to investigate the existence of potentially troublesome subject variables, as well as the lack of uniformity of results among the studies that do exist, we have decided to embark upon a series of studies (this report being, hopefully, the first) designed to investigate inherent differences among various samples of student research participants. Inherent in most of the statistical models we use is the presumption that the samples with which we are working are, before the imposition of any treatment variables, representative samples of a common population. If this is the case, then subjects in different samples asked to do the same tasks under the same conditions should produce performances that differ from each other by no more than an amount consistent with sampling error (i.e., not sufficiently different from each other to be statistically significant). The present

Requests for reprints should be sent to R. J. Senter, Department of Psychology, University of Cincinnati, Ohio 45221.



0090-5054/81/060261-03$00.55/0

investigation was designed in an attempt to ascertain whether this implicit statistical assumption would, in fact, be true for a simple two-group experiment patterned, roughly, after the research of Richert and Ward (1976) and Underwood et al. (1964).

## METHOD

### Subject Selection

Subjects were selected via a method often, but certainly not universally, used at the University of Cincinnati for the recruitment of "indifferent" or "undifferentiated" subject samples. This method consists of the posting of "sign-up sheets" on various bulletin boards near the psychology department offices. These sheets announce the nature, time, and location of ongoing or upcoming experiments and provide spaces in which students in the introductory-level classes may sign up for participation. In the academic quarter (fall 1978-1979) during which the present research was conducted, introductory-level students were required to participate in one experiment, but extra course credit could be obtained through participation in additional research projects. For the present research, a sign-up sheet was posted for a 48-h period during the 2nd week of the 10-week academic quarter ("early group") and another was posted for the same time during the 8th week of the same quarter ("late group"). Since the number, sex, ages, and so on, of the subjects were considered to be dependent measures in the present study, these data will be presented in the Results section.

### Procedure

Subjects in both the early and late groups were subjected to the experimental tasks during 1-h periods held 2 days after their sign-up sheets were first posted. The room in which and the time of day at which the subjects were tested were the same for both groups. The experimenter was the same for both sessions, and the instructions were identically presented to both groups.

The subjects were required to perform two experimental tasks. The first of these was an 11-item (three-letter nonsense syllables) serial learning task (serial anticipation). The stimuli were projected via a 35-mm Carousel projector for 5 sec each with an interstimulus interval of 5 sec, during which the subjects were to write their anticipations of the next syllable. The dependent measure was the number of syllables correctly anticipated during each of seven trials.

The second task was a symbol substitution task of local origin. At the top of an 8.5 x 11 in. page was listed a set of 10 "pseudo-German verbs" (i.e., seven-letter nonsense words beginning with "er" or "ek" and ending with "en" or "em"). Each of these nonsense words was paired with a randomly selected two-digit number. The remainder of the page was filled with columns of the 10 nonsense words, arranged in haphazard order, flanked by a blank in which the number associated with each word was to be written. The instructions included an admonition to "work as fast and accurately as you can." The dependent measures for the symbol substitution task were the number of substitutions attempted and the number correct.

Subsequent to the completion of the experimental tasks, the subjects were asked to respond to a brief questionnaire that was designed to collect demographic and personal information.

## RESULTS

We regarded the characteristics of the subjects who signed up for each of the experimental sessions as being, for the present study, dependent measures. The number of students who "volunteered" for each session was almost the same: 34 early and 36 late (there was room on the sign-up sheets for 50 signatures). Two subjects were randomly dropped from the late group for the sake of equal Ns.

The mean ages for the two groups were quite similar (early = 19 years, range = 17-23; late = 19 years, range = 18-24).

The most obvious difference between the subjects signing up for the two sessions was that the early group contained a preponderance of females (20 female, 14 male), whereas the solicitation offered late in the quarter attracted mostly males (11 female, 23 male).

Responses on the questionnaire showed that the motivation for signing up was quite different. Thirty of the 34 early volunteers indicated that they had signed up to fulfill their course requirement instead of for extra credit. In the late group, the ratio was almost reversed, with only three students indicating that signing up was motivated by the course requirement. This is not surprising, since by the end of any given quarter most of the available subject population would have completed their required hour and, hence, extra credit would be the sole available motive for most participants.

Figure 1 shows the relative performance of the two groups on the serial learning task. Analysis of variance showed the superiority of the early group to yield a difference significant beyond the .05 level [$F(1,66) = 4.39$]. The Trials by Conditions interaction was reflected in an $F(6,396)$ of 5.23 ($p < .001$). Apparently, the general superiority shown initially by the late group was diminished by the steeper slope of the early group's learning curve.

With the symbol substitution task, there was also a significant difference between the two groups, but with the early group showing superiority on both the number of items attempted [$t(66) = 3.46$, $p < .01$] and

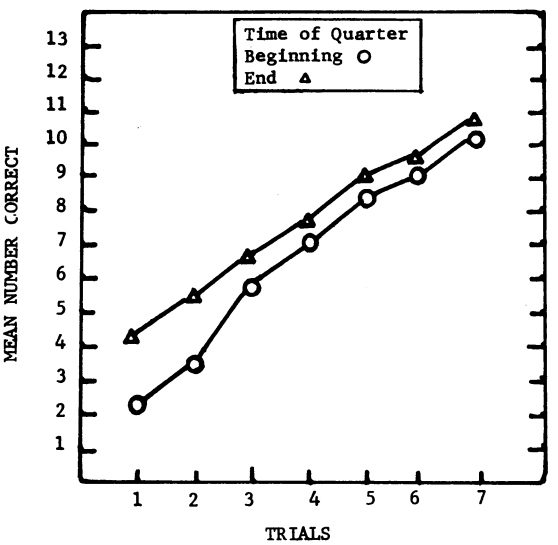


**Figure 1. Mean number of syllables correctly recalled over trials by students participating at the beginning and at the end of the academic period.**

the number of items correctly performed [$t(66) = 3.86$, $p < .01$].

Since there was considerable difference in the sexual composition of the two groups, we investigated the possibility that the group performance difference might be attributed to performance differences between sexes. Comparison of the serial learning performances of male and female subjects in the early group via the t ratio yielded $t(32) = .143$ ($p > .05$), and a like comparison for the late group resulted in $t(32) = .911$ ($p > .05$). Similar "by-sexes" comparisons for the symbol substitution resulted in the following: number of items completed, early $t(32) = .075$ ($p > .05$), late $t(32) = 1.24$ ($p > .05$). There was, then, no significant sex difference for any of our dependent variables for either experimental session.

## DISCUSSION

It would appear that the difference in performance between our early and late volunteering samples is not reasonably attributable to sampling error alone, and, therefore, the groups, as tested, do not appear to have been representative samples of the same population. We cannot, of course, at this time even speculate as to the causes of our nonnull results. There is always the possibility that we have been unlucky enough to have made a series of alpha errors in the same experiment. Although possible, this is probabilistically quite unlikely. It is also possible that we inadvertently administered some "treatment effect" that engendered differences in group performances. Since it was of great importance to the research to keep the treatments of the groups as similar as possible, we believe that every possible effort was made to avoid differential treatment. We can never be sure that we succeeded, but if we did not, whatever differential treatment might have been administered must have been very subtle–too subtle, we think, to account for the rather substantial differences observed.

In our opinion, it would seem most parsimonious to conclude that the samples examined were simply composed of individuals who, with respect to some personal attributes, were not (within the bounds of sampling error) "equivalent." The fact that the observed differences were in opposite directions on the two tasks seems to obviate any simple explanation, such as that the more motivated and ambitious students sign up early, or the students seeking extra credit are more motivated, or the lazy students sign up late, or students who have participated in other experiments do better, for all these conjectures would predict superior performance in a single direction. What we appear to have here is some kind of interaction between type of task performed and the time of sample selection.

It may well be that all we have done is to add more noise to the system. Underwood et al. (1964) found no differences among groups sampled at different times; Richert and Ward (1976) found no difference on one task but a significant difference with another. Now come the present researchers with data yielding significant results in opposite directions for two different tasks. Further research is clearly necessary to test the reliability of our findings and/or to attempt to pinpoint the sources of our observed differences. For the present, though, the possible implications of our findings for a great deal of psychological research conducted over the last 5 or 6 decades is most disquieting. Whatever may be the ultimate resolution of this problem, the immediate caveat is clear: We cannot unquestioningly assume that different subject samples taken at different times during any academic period are statistically "equivalent" in all matters that might affect our dependent measure.

## REFERENCES

JUNG, J. Current practices and problems in the use of college students for psychological research. *Canadian Psychologist,* 1969, **10**, 280-290.

RICHERT, A. J., & WARD, E. F. Experimental performance and self-evaluation of subjects sampled early, middle, and late in an academic term: Sex and task. *Psychological Reports,* 1976, **39**, 135-142.

SMART, R. G. Subject selection bias in psychological research. *Canadian Psychologist,* 1966, **7**, 115-121.

UNDERWOOD, B. J., SCHWENN, E., & KEPPEL, G. Verbal learning as related to point of time in the school term. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3**, 222-225.

(Received for publication May 19, 1981.)