JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 14, NO. 6, PP. 481-498 (1977)

RESEARCH PAPERS

DOES VIDEO-AUTOTUTORIAL INSTRUCTION IMPROVE COLLEGE STUDENT ACHIEVEMENT?

K. M. FISHER,* H. GUENTHER,** and B. MACWHINNEY†

Teaching Resources Center, University of California-Davis, Davis, California 95616

P. SORENSEN and D. STEWART

Stanford Research Institute, Menlo Park, California

A video-autotutorial (video-AT) method of instruction has been developed which utilizes fully illustrated television modules supplemented by a written syllabus to transmit information to students enrolled in an upper-division science course (Fisher, 1974a). How effective is this instructional method?

This paper presents results of some of the efforts to evaluate the method as it is employed in teaching an introductory genetics course having enrollments of 1200-1600 students each year. The criterion of effectiveness considered here is comparative gain in knowledge of the subject as measured through objective tests of achievement preceding and following instruction. Other criteria, such as student attitudes toward the instructional method and its relative cost efficiency, have been reported elsewhere (e.g., Fisher, 1974b).

Achievement of students enrolled in the video-AT course is compared with that of two comparable intact groups of students enrolled in two similar introductory genetics courses taught by the lecture/discussion method. The three courses are offered on three different campuses of the University of California. The lecture/discussion method was selected for comparison because it is the "standard" teaching method for university science courses and because it was the method being employed in the best-matched "sister" courses in the university.

The lecture/discussion method provides a well-established standard for comparison, in that 40 years of research have shown that overall there is no difference in effectiveness between lecture, discussion, lecture/discussion, and/or independent study methods of teaching at the college level as measured by student pretest-posttest achievement gains on written examinations (Dubin & Taveggia, 1969).

While there are no differences *overall,* however, differences were observed in many of the individual two-way comparisons included in the review by Dubin and Taveggia (1969). For this reason, two comparison courses (rather than one) were employed in this study in an effort to increase the likelihood of obtaining a representative sample. The chance that comparison courses would involve exceptionally good or exceptionally poor applications of

*Author to whom correspondence should be addressed.
**Present address: Federal University of Paraiba, Brazil.
†Present address: University of Denver, Denver, Colorado.

Published by John Wiley & Sons, Inc.

the lecture/discussion method, or would be confounded by unusual instructor effects, is lowered by using at least two different comparison courses. The latter problem (unusual instructor effects) is further reduced in this case by the fact that one of the comparison courses was team-taught by several instructors.

Data for the comparisons of genetics achievement test scores both prior to and following instruction are based on paired pretest-posttest scores from a total of 623 students–153 from lecture campus A, 84 from lecture campus B, and 386 from video-AT campus C. The results presented here are unusually strong in suggesting that the nonconventional instructional format may be superior for teaching genetics in the university setting.

The achievement comparisons are based on the assumption that the three classes are equivalent. To some extent they are. All three courses are introductory upper-division genetics courses having similar content and purposes. The achievement comparisons were limited to those portions of the courses that were most alike in content and depth of coverage; the few dissimilar topics contained in each course were not considered. The students are science majors with advanced class standing who have all met University of California entrance requirements and fulfilled similar prerequisites.

In order to evaluate the findings, however, it is important to know what differences, if any, exist between the campus groups. Demographic and attitudinal data were collected from all students who participated in the study. Some of these characteristics, particularly those which might be expected to affect achievement, are summarized for the reader here.

The research literature is rich with studies that have examined socioeconomic and other background factors as predictors of college achievement or academic performance. Findings are not unequivocal, and occasionally contradictory, as a result of varying definitions of achievement, different indices to represent background variables, and a variety of samples. Generally, such variables as family income, father's occupation, type of secondary school attended, and urbanism display low to moderate positive correlations with academic performance; although often statistically significant, the correlations do not tend to account for a high proportion of the variance in the achievement measure. Typically, prior grade point (or, occasionally, class standing) emerges as the best single predictor of achievement when achievement is measured by grade-point average. References that review, summarize, or synthesize such data include Lavin (1965), Duncan et al. (1972), and Trent and Cohen in Travers (1973).

Glass (1976) has very recently cited a "meta-analysis" of the relationship between socioeconomic status (SES) and school achievement in which 636 independent correlations obtained in nearly as many studies were examined. This distribution of correlations revealed an average coefficient of ". . . .25 with a standard deviation of about .20 and positive skew; this means the SES and achievement correlation is below what is generally believed to be the strength of association of the two variables."

Another important factor to consider is community college experience. Students who transfer to the university from a community college usually enter in the junior year and characteristically suffer a pronounced setback in the level of their academic performance for at least one or two quarters (Knoell & Medsker, 1964a, 1964b). If one of the comparison courses under study contained an unusually large proportion of new transfer students, the pretest–posttest gains could be affected significantly.

Finally, regression and commonality analyses are performed to provide bases for estimating how much of the variance in the final test of genetics knowledge could be attributed, uniquely or jointly, to various regressors. With the posttest score as the criterion variable, 47 student background variables and the pretest score were examined for their utility as predictors. The multiple-regression analyses were based on a subset of 555 students.

TABLE I

Number of Students (*n*), Pretest and Posttest Scores (M), and Standard Deviations (SD) on the Three Campuses

Test	Statistic	Campus Lecture A	Lecture B	Video-AT C	Total
	n	153	84	386	623
Pretest*	M	14.9	14.9	12.5	13.4
	SD	3.6	3.6	4.0	4.0
	n	153	84	386	623
Posttest**	M	20.8	21.0	24.5	23.1
	SD	4.1	3.6	3.9	4.3

*Maximum score 25.
**Maximum score 30.

Evaluation Design

The design for evaluation conformed to what Campbell and Stanley (1963) have called the "nonequivalent control group design." Essentially, three treatment groups are compared and there is no "pure" control group.

The student groups consisted of intact classes on three different University of California campuses. The classes were enrolled during a single quarter in introductory genetics courses whose purposes and content were similar, but not strictly identical, to one another. Achievement comparisons were made between portions of each course concerned with classical, molecular, and biochemical genetics; these subject areas accounted for approximately three units of the total introductory units of genetics offered on each campus.

Instructional treatment and campus (including student and faculty) characteristics are confounded in the design. However, all students had met common University of California entrance requirements and were found retrospectively to be quite similar on such dimensions as academic area of major concentration (if not major field *per se*), occupational orientation (e.g., medicine, biological science), and incentives for enrollment in the course. (For most students on all three campuses, the course was a requirement for their major.)

The three regularly scheduled courses were taught by regular ladder faculty in the University of California. Participation in the study by faculty was solicited but voluntary.

Student participation also was solicited and was voluntary for three of the four instruments utilized (pretest, presurvey, postsurvey). The fourth instrument, a posttest, was incorporated into regular examinations and was therefore required for all students taking the course for credit.

Comparative Achievement

Number of students on each campus, group mean scores, and standard deviations of score distributions are shown in Table I for both the pretest and the posttest. Table II

TABLE II

Analysis of Variance on Pretest and Posttest Scores

Source	df	MS	F	p
Pretest				
Between campuses	2	424.9	28.37*	<.001
Within campuses	620	15.0		
Total	622			
Posttest				
Between campuses	2	950.8	61.18*	<.001
Within campuses	620	15.5		
Total	622			

presents the results of analyses of variance for the data summarized in Table I.

The top half of Table I shows that the groups of students at lecture campuses A and B were very similar to one another on the pretest and that both groups were superior to the group on video-AT campus C. The significance of the overall differences is shown in Table II ($F = 28.37$; $p < 0.001$). Pretest raw scores on campus A ranged from 5 to 23 and on campus B from 7 to 23 for the 25-item test. On campus C, the pretest score range was from 1 to 22. Although the mean raw score difference was only about 2.4, this difference represents about 10 percentage points in the 25-item pretest.

As was the case with pretest performance, differences on the final test between campuses A and B were trivial and not statistically significant. However, as shown by the results of the analyses of variance summarized in Table II, the overall differences were significant ($F = 61.18$, $p < 0.001$). The mean raw score difference on the posttest was 3.68 or 12 percentage points. Thus, the lecture campus students earned 59% on the pretest and

TABLE III

Similarities among Campus Groups

Characteristic	Campus A	Campus B	Campus C
Mean grade point average in UC system	3.0	3.1	3.0
% requiring genetics	88.3	97.6	96.4
Number of units that quarter	15.3	14.6	15.1
Average Age	21.8	21.9	21.1
% fathers who are college grads	54.5	53.0	54.8

70% on the posttest, while the video-AT students earned 50% on the pretest and 82% on the posttest.

In evaluating the findings, it is important to examine the extent of comparability among campus groups. Student profiles on the three campuses were compared through a relatively extensive series of demographic and attitudinal items employed in the pilot study (Fisher, 1974b). On the basis of these earlier data, a consolidated list of 47 key items (24 demographic and 23 attitudinal) were selected for use in the present study. The campus groups were similar in many characteristics, some of which are illustrated in Table III.

Table IV shows the most striking differences. The proportion of students whose occupational goals were in the health professions (medical, dental, veterinary) was the highest at lecture campus B. On the average, students at video-AT campus C came from smaller communities and they spent more quarters in junior colleges and correspondingly fewer quarters in the UC system.

The video-AT course is a two-quarter, six-unit, upper-division introduction to genetics that is required for most majors in the natural sciences. The course is characterized by the following features:

1. Fully illustrated instructional modules on video tape, each about 25 minutes in length and devoted to a single topic of genetics.
2. A Learning Center in which viewing can occur throughout the day and evening.
3. A Conference Room adjacent to the viewing areas in which faculty, teaching assistants, and undergraduate tutors are available for individual consultation.
4. A detailed syllabus keyed to the video modules.
5. Frequent quizzes.
6. Lectures for motivation.

The comparison lecture courses were selected on the basis of comparability in content and student composition. The comparative approaches can be characterized as one-quarter, four- and five-unit, upper-division introductions to genetics that are required for most majors in the natural sciences. The lecture courses consist of lectures to large groups with smaller discussion sections led by faculty or teaching assistants, augmented by textbook reading and problem assignments. One of the comparison courses was team-taught and the other was presented by a single lecturer with several teaching assistants.

The three courses are all upper-division introductions to genetics for science majors. Although they differ in total number of units, the variations in content occur primarily in the extent of coverage of quantitative, evolutionary, and population genetics. The numbers

TABLE IV

Differences among Campus Groups

Characteristic	Campus A	Campus B	Campus C
% with health field as goal	42.9	70.2	57.7
% living in large city at age 16	29.4	37.3	15.5
Average quarters in University of California	6.4	6.9	5.0
Average quarters in Comm. College	5.2	4.1	5.8
% of all students who attended community college	45.2	41.2	59.5

of lectures or modules committed to classical, biochemical, and molecular genetics, on which this study is based, are approximately equivalent (about 26 50-minute lectures or 16 75-minute lectures or 24 25-minute modules). Comparison of course outlines and notes reveals extensive overlap in content in these areas, although emphases vary to some extent.

The faculty are similar in that they are ladder faculty in the University, largely tenured and male, under about 45 years of age, and have no apparent distracting or annoying mannerisms. Their approximate equivalence from the student view is apparent in student evaluations of faculty obtained on the postsurvey (Fisher, 1974b).

Identical forms of student background and attitude questionnaires and pretests of genetics knowledge were administered on all three campuses at the outset of the quarter. Likewise, identical forms of postsurveys and posttests were administered on all three campuses at the conclusion of the study. All instruments were pretested in a pilot study conducted the year preceding the experiment described here.

Methods for assuring comparability of administration of the instruments also were devised, tested, and refined through the pilot study. The principal investigator or her primary assistant monitored instrument administration on all campuses to assure that comparability was achieved within controllable limits. Student participation was usually solicited by the reading of a prepared statement and the presentation of an accompanying letter to each student. Consistent guidelines were developed for handling survey returns, latecomers to courses, and so on. The faculty and teaching assistants on each campus worked closely with the authors. These efforts largely preclude the possibility that a trivial difference—for example, excessive cheating on one campus—might account for differences in results.

New instructional methods tend to be accompanied by a Hawthorne effect which could conceivably create a bias in favor of the video-AT campus. At the time of this study, however, the video-AT method was being utilized for the fourth consecutive quarter of genetics instruction on campus C. Further, this instructional method had been preceded by another innovative approach, televised lectures, for eight successive quarters. That the Hawthorne effect had subsided is suggested by the student evaluations of the teaching method. The first quarter in which the video-AT method was utilized, 75% of the students enrolled in the course indicated preference for this method over the conventional lecture approach. In subsequent quarters, the proportion of students favoring the video-AT method subsided and has held steady in the 58–65% range.

Finally, it is necessary to ask whether either of the tests was biased to favor one or another campus. The pretest was constructed by participating faculty to assess the rudimentary genetics knowledge expected upon completion of an introductory biology course, a prerequisite for all of the genetics courses under study. The pretest consisted of 25 multiple-choice questions. It required about 25 minutes on the average and was completed by all students within one class period. Reliability of the pretest was estimated by coefficient alpha as 0.71 for a total of 627 cases on all three campuses. Coefficients ranged from 0.66 to 0.69 for individual campus groups.

On campus C (the video-AT campus), approximately 800 students were enrolled in the course. Student background questionnaires and pretests were administered to a random sample of about half the students; complete data were available for analyses for 386 students at campus C, as shown in Table I. Similarities prior to instruction between students who took the pretest and those who did not cannot be verified directly since comparable test data were not obtained from all. However, the presumptive evidence is strong that both groups—those pretested and those not pretested—were similar, since the two groups did not differ significantly on the final test of genetics knowledge. The mean difference on the final test between the pretested group ($n = 386$) and the group not pretested ($n = 340$) was 0.21 raw score points on a 30-item test. The standard error of this mean difference was 0.29.

Intuitively, one might expect the pretest to be biased in favor of campus C because (1) it was devised largely at campus C and (2) it included a number of items obtained from final examinations given in a biology course at campus C. Efforts to avoid bias included (1) review, with veto power, of each question by faculty on campuses A and B; (2) pretesting of an earlier version of the pretest in other biological science courses on both campuses A and B; and (3) pretesting of the instrument on all three campuses during the pilot study. Empirical evidence that the pretest was not biased is summarized in Table V. The difficulties of pretest items show high correlation among campus groups. This would not be expected if the pretest covered information taught on one campus but not on another.

Construction of the posttest was more cosmopolitan than that of the pretest. Items were obtained from faculty on all three campuses as well as from outside sources, including confidential items from the Professional Examinations Division of The Psychological Corporation and from the Graduate Record Examinations Biology Test of the Educational Testing Service. A rather long list of potential items was generated. Faculty on all three campuses then ranked each item (most applicable, moderately applicable, not at all applicable) with respect to their course contents, and the posttest was compiled in such a way that each campus had equal numbers of "most applicable" and "moderately applicable" questions. All questions that were "not at all applicable" were excluded.

The posttest consisted of 30 multiple-choice questions. These items were included in regularly scheduled examinations which contributed to determination of the students' grades. On each campus, the examinations included additional items which were unique to that campus and which were designed by the responsible faculty. As with the pretest, the posttests were generally comfortably completed in the time allotted. Reliability of the final test was estimated as 0.77 by coefficient alpha for a total of 967 cases on the combined campuses. Coefficients ranged from 0.62 to 0.78 according to campus groups (one group each on campuses A and B and two subgroups on campus C).

The posttest was not as extensively pretested and it contained questions of a more specific nature than the pretest. This may account for the slightly lower intercampus correlations of item difficulties obtained (Table VI). Examination of a scatter diagram reveals that the torque is attributable largely to a few items which may have had greater curricular validity on one or another campus.

TABLE V

Intercamps Correlations between Relative Difficulties of Pretest Items*

	Campus			
Campus	A	B	C_1	C_2
A	-			
B	93	-		
C_1	97	90	-	
C_2	96	91	97	-

*$N = 25$. C_1 and C_2 are subgroups at campus C. Correlations are product-moment. Decimals omitted.

TABLE VI

Intercampus Correlations between Relative Difficulties of Posttest Items*

Campus	Campus: A	B	C_1	C_2
A	-			
B	70	-		
C_1	74	67	-	
C_2	71	67	94	-

*$N = 30$. C_1 and C_2 are subgroups at campus C. Correlations are product-moment. Decimals omitted.

Data for the comparisons of genetics achievement test scores both prior to and following instruction are based on paired pretest-posttest scores from 623 students (153 from lecture campus A, 84 from lecture campus B, and 386 from video-AT campus C). Multiple-regression analyses utilizing variables from both the presurvey and the pretest of genetics knowledge were based on a subset of 555 students (225 from the combination of campuses A and B and 330 from campus C). Altogether, 68 cases used in achievement comparisons were excluded from the regression analysis due to missing data on certain background variables, since incomplete cases were dropped from the analysis in preference to imputing values for missing data. Thus, 5.06% of campus A and B students and 14.51% of campus C students were excluded. The difference in exclusion rates may be attributable to the larger size of the campus C course and to its individualized organization wherein there is necessarily more reliance on individual student initiative in completing surveys.

Effect of Instructional Method

Regression and commonality analyses were performed to provide bases for estimating how much of the predictable variance in the final test of genetics knowledge could be attributed, uniquely or jointly, to various regressors. Three sets of regressors (or predictors, since causation was inferred) were analyzed: (1) student background characteristics derived from the presurvey, (2) the pretest administered prior to instruction, and (3) instructional format or treatment.

The presurveys contained 24 demographic items and 23 attitudinal questions and were administered during the first week of classes. The 47 student background variables derived from the presurveys were examined for their utility as predictors of the final test score. Variables that contributed significantly to the multiple regression coefficient were identified. The stepwise procedure from SPSS was used with an F-to-enter criterion of $p < 0.05$ (Nie et al., 1970). By this method, the set of potential predictors was reduced to seven variables. Brief definitions of these seven variables (plus two additional ones that just failed to satisfy the F-to-enter criterion) appear in Table VII as variables numbered 4–12. Table VII also provides brief definitions of the criterion variable, the instructional format variable, and the pretest variables (variable numbers 1-3).

Table VIII displays the bivariate correlations between all pairs of the seven background variables, the pretest, the instructional format variable, and the final test score as the criterion. The interpretation of negatively signed coefficients in Table VIII will be aided by referring to the variable definitions in Table VII. The overall correlation between the pretest and final is surprisingly low, 0.27. Within treatment groups the pretest-posttest correlations are higher, however, being 0.45 for campuses A and B ($N = 233$) and 0.42 for campus C ($N = 355$).

The seven student background variables identified as the "best" set for predicting final test performance were combined with the pretest and instructional format variables to produce the overall regression solution summarized in Table IX. Table IX shows the changing pattern of partial correlations between predictors and the criterion as successive predictors were entered into the equation. A multiple-regression coefficient of 0.67 was achieved before the inclusion of additional predictors did not contribute significantly to explanation of variance in the criterion. (The coefficient, when corrected for shrinkage, reduced slightly from 0.67 to 0.66.) Beta weights for the predictors shown in Table IX are given in Table X.

TABLE VII

Brief Definitions, Mean Values, and Standard Deviations for Variables Used in the Regression Analysis ($n = 555$)

Variables	Scale Range	Mean	Standard Deviation
1. Posttest	0 - 30	23.1	4.3
2. Instructional format	1 = lecture 2 = video-AT	1.6	0.5
3. Pretest	0 - 25	13.6	4.0
4. Grade point average	0 - 4 (high)	3.1	0.5
5. Qtrs. in univ. system	Actual	4.2	3.8
6. Certainty of occupational preference	1 = low 7 = high	5.1	1.7
7. Perceived necessity of college	1 = low 7 = high	5.8	1.7
8. Expected grade	0 - 4	3.7	0.5
9. Size community in which lived at age 16	1 = small 5 = large	2.8	0.9
10. Qtrs. in community college	Actual	3.0	3.5
11. Educational attainment of father	1 = low 7 = high	4.7	1.7
12. Working part-time	1 = none 5 = much	1.7	1.0

TABLE VIII

Intercorrelations among Predictor and Criterion Variables Used in Multiple-Regression Analysis*

	GPA	Qtrs. UC	Certainty Occup.	College Necessary	Expect Grade	Community @ Age 16	Qtrs. JC	Pretest	Instruct Format	Final
GPA	---	-15	13	08	29	-02	-04	25	06	36
Qtrs. UC	-15	---	-02	01	-10	15	-63	26	-32	05
Certainty occup.	13	-02	---	20	18	00	05	-06	07	-05
College necessity	08	01	20	---	16	12	04	16	-03	13
Expected grade	29	-10	18	16	---	-05	10	14	18	27
Community @ Age 16	-02	15	00	12	-05	---	-13	11	-27	-14
Qtrs. JC	-04	-63	05	04	10	-13	---	-25	19	-14
Pretest	25	26	-06	16	14	11	-25	---	-28	27
Instruct format	06	-32	07	-03	18	-27	19	-28	---	39
Final	36	05	-05	13	27	-14	-14	27	39	---

*$n = 555$. Coefficients rounded and decimals omitted. If coefficient equals 10 or more, $p < 0.01$.

TABLE IX

Multiple and Partial Correlations of Instructional Format, Pretest, and Student Background Variables with Final Test Performance

Predictor Variables	Multiple and Partial Correlations According to Step in which Predictor Variable Entered Regression Equation*											Square of Multiple Correlation
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	
Instructional format	39									**		15
Pretest	43	56										32
Grade point average	37	28	61									37
Qtrs. in univ. system	21	14	21	63								40
Certainty of occupational preference	-09	-07	-12	-14	64							41
Necessity of college	16	10	09	10	13	65						42
Expected grade	22	15	08	10	12	11	66					43
Size community in which lived at age 16	-03	-06	-06	-07	-07	-08	-08	66				43
Qtrs in community college	-24	-17	-18	-06	-05	-06	-07	-08	66			44
Educational attainment of father	11	10	10	08	07	06	06	07	06	66		44
Working part-time	-06	-04	-01	-04	-03	-03	-04	-03	-04	-03	66	44

Note: $n = 555$. Coefficients rounded and decimals omitted.

*Multiple correlation coefficients are underscored and appear on the diagonal.

**Variables entered beyond the ninth step did not contribute significantly to the multiple-regression coefficient ($p > 0.05$ by F test).

TABLE X

Beta Weights for the Predictors Shown in Table IX

Predictor Variable	Regression Coefficient	Ordinal Rank of Variable Potency
Instructional format	.49	1
Grade point average	.26	2
Pretest score	.25	3
Quarters in UC system	.15	4
Certainty of occupational preference	-.14	5
Necessity of college	.11	6
Expected grade in course	.09	7
Quarters in Comm. College	-.08	8
Size community in which lived at age 16	-.07	9

Four of the background predictors were associated positively with performance on the posttest: (1) reported grade-point average, (2) number of prior quarters in the University of California system, (3) perceived necessity of a college education, and (4) the expected grade in the course. Three of the background predictors were associated negatively with the criterion: (1) expressed certainty of occupational preference, (2) size of the community in which the student had lived at age 16, and (3) the number of quarters the student had attended a junior or community college.

The multiple-regression analysis described and summarized above was performed in such a way that estimates could be made of the amount of variance attributable to student background, pretest, and instructional treatment variables. The procedure conformed to the model described by Mayeske et al. (1970). Briefly, the computational steps were as follows:

1. A "best set" of student background variables was defined on the basis of results of a stepwise analysis against the criterion of final test scores (Table X). In succeeding computations involving other predictors as well, this set of seven background variables was used.
2. Three stepwise computations were performed with predictor variables or sets introduced in a prescribed order:
   a. Full solution with predictors entered as follows: (1) instructional format, (2) pretest, (3) student background.
   b. Partial solution: (1) pretest, (2) student background.
   c. Partial solution: (1) student background, (2) instructional format.

The above three stepwise solutions provided the orderings of the three sets of predictors that were necessary to estimate the proportion of total variance in the criterion that was accounted for by different combinations of predictors. The three sets of predictors–student background composed of seven individual variables, the pretest score, and the instructional format variable–together account for 44% of the total variance in the criterion of final test score (i.e., $0.44 = 0.66^2$). By subtraction, one may estimate the following:

1. Student background variables accounted for 12% of variance in the criterion (or 28% of the total explained variance).
2. The pretest of genetics knowledge accounted for 5% of variance in the criterion (or 11% of the total explained variance).
3. The instructional format or treatment accounted for 19% of variance in the criterion (or 43% of the total explained variance).
4. All variables entered into the equation shared some common factors in accounting for 8% of variance in the criterion (or 18% of the total explained variance).

As noted earlier, a conventional correction for shrinkage with nine variables and 555 cases reduced the estimate of the multiple-regression coefficient from 0.67 to 0.66. Table IX and the above comments do not reflect this modest reduction in bias since the main intent of the commonality analysis was to identify the relative potency of the various predictors.

Student Ability versus Instructional Method

It has been suggested that autotutorial instruction particularly favors the slower student (Diederich & Macklin, 1973). To determine whether the video-autotutorial and lecture methods of genetics instruction examined here interacted with student ability, the subjects were divided into three approximately equal strata on the basis of pretest scores, and posttest achievement was examined for each campus group (Table XI).

Within each stratum, differences on the posttest between lecture campuses A and B were trivial and not statistically significant. However, the video-AT campus differed from campuses A and B in all strata. The overall differences are significant (Hi $F = 33.08$, Med $F = 62.5$, Lo $F = 16.48$; in all cases, $p < 0.001$).

The mean posttest scores by pretest strata show no overlap between lecture and video-AT groups. The video-AT group scoring in the lowest third on the pretest performed at least as well on the final as the lecture groups scoring in the highest third on the pretest.

TABLE XI

Posttest Scores Examined by Pretest Strata (Thirds) and Campus Group

Pretest Stratum	Statistic	Campus A	Campus B	Campus C	Total
	n	69	43	89	201
Hi	M	22.9	22.2	26.4	24.3
	SD	3.4	3.7	2.9	3.7
	n	53	27	140	220
Med	M	19.4	20.0	25.0	23.1
	SD	4.1	2.5	3.3	4.4
	n	31	14	157	202
Lo	M	18.7	19.4	22.9	22.0
	SD	3.4	3.9	4.3	4.5

There is no evidence in these data to suggest that autotutorial methods favor the poorer student or that lecture methods favor the better student.

Discussion

The evaluation provides strong evidence for the effectiveness of the video-autotutorial instructional approach in the setting in which it was used in contrast to more conventional approaches to large-group university instruction in a science field. The video-AT group began the instructional quarter with somewhat less knowledge of the subject matter than students in the lecture groups ($p < 0.001$ for the significance of mean differences on the pretest). They completed the instructional quarter with greater knowledge of the subject matter than students in the lecture groups ($p < 0.001$ for the significance of mean differences on the posttest). Furthermore, the regression and commonality analyses suggest that instructional format accounted uniquely for more variance in the final test criterion than did either student background variables or level of prior knowledge of subject matter as measured by the pretest.

Before discussing possible reasons for these apparently strong findings, certain weaknesses in the evaluation design and analysis deserve comment for they may affect interpretation of the strength of the instructional format variable.

An ideal evaluation design would consist of one or more within-campus studies in which both students and faculty are randomly assigned to treatment. A variety of considerations, including student preferences, faculty preferences, artificiality, and problems of policing potential crossing-over between treatments, argued against this approach. As a consequence, student groups on each campus were not as similar to one another in key characteristics as might have been possible under conditions idealized for evaluation.

Of the seven student background variables which were found to be useful predictors of the posttest performance, intercampus mean differences were statistically significant on four: (1) size of community in which the student lived at age 16, (2) number of prior quarters in the University of California system, (3) expected grade in the genetics course, and (4) number of prior quarters in a community or junior college. Analysis of the beta weights associated with these seven variables and the direction and magnitude of the correlation between each variable and the final test criterion suggests that biases due to mismatch between video-AT and lecture campuses were more likely to favor the lecture campuses. Campus C (video-AT) students as a group had fewer prior quarters in the University of California system than their counterparts on the lecture campuses; quarters in the university system carried a positive beta weight and was the second most potent of the seven background predictors. Campus C students also reported more quarters in a community or junior college than students on the control campuses; community college attendance had a negative beta weight. Factors that might have contributed to a bias favoring campus C were (1) expected course grade, which was highest for students on campus C, and (2) size of community in which student lived at age 16, which was smallest for students on campus C but a variable that carried negative beta weight. Reference to the tabulation showing regression coefficients will show that both these variables carried less weight than most other regressors. We therefore conclude that the confounding of instructional approach and student characteristics was more likely to have reduced differences between the video-AT and lecture groups than to have magnified them.

Course content and faculty were also judged, on the basis of both student and faculty evaluations as well as on circumstantial evidence, such as the interchangeability of textbooks, to be very similar.

Another consideration that may influence the findings is the fallibility of the measures. By comparison to carefully standardized and normed achievement tests, the reliabilities of the pretest and posttest were not especially high (i.e., 0.71 for the pretest and 0.77 for the posttest by coefficient alpha, which estimates a lower bound). On the other hand, for faculty-made tests with relatively few items (i.e., 25 in the pretest and 30 in the posttest), the reliability estimates are quite respectable. The reliabilities of both the pretest and the posttest are high enough to protect against serious misinterpretation of the relationship obtained between the pretest and the posttest. Table VIII showed the correlation obtained between the pretest and the posttest to be 0.27 for $n = 555$. If both the pretest and the posttest had been more reliable–for example, 0.90 for both measures–the expected correlation between pretest and posttest would be 0.33 rather than 0.27. A difference of that magnitude would be unlikely to result in markedly changed findings. Furthermore, pretest-posttest correlations within groups are much higher, 0.45 for the lecture group ($N = 233$) and 0.42 for the video-AT group ($N = 355$).

Finally, an unequivocal assertion seems appropriate regarding the curricular validity of the pretest and posttest of genetics knowledge. As noted earlier, participating faculty on all three campuses collaborated in the specification of content coverage and the preparation of the instruments. Test questions covered information in classical, molecular, and biochemical genetics that faculty considered germane to the field and relevant to the instruction. The resulting tests were as valid, against criteria of instructional intent, as the participating faculty were able to produce within the constraints of time and geographical separation. The high intercampus correlations on relative item difficulties shown in Tables V and VI suggest that the pretest and posttest contained relatively little bias.

The fallibility of the self-report information obtained from students in the student background questionnaire also must be acknowledged. Unfortunately, we have no way of estimating the extent to which indices derived from these responses may be biased. The data appeared credible overall, and internal logic checks were applied to exclude cases when data were outside the range (e.g., a reported GPA in excess of 4.0 or terms of attendance beyond plausible bounds). Student expressions of opinion (e.g., certainty of occupational preference) were accepted as reported so long as responses fell within defined scale limits on the instruments. Although we can expect that self-reports of both verifiable fact and opinion will not be error-free, we have no reason to believe that the reports were any more or less biased for students in the lecture groups than in the video-AT group.

The textbook is presumed to be an important component of any teaching method, and the comparative study does not specifically separate out textbook effects from other instructional effects. However, students were asked to rate each course component with respect to its perceived contribution to their knowledge and understanding of genetics. On the video-AT campus, the syllabus rated highest with the TV modules a close second, and the textbook rated substantially lower in third place out of a total of ten components (others were discussion sections, practice problems, faculty contact, TA contact, lectures, quizzes, and student contact). On the lecture campuses, the lectures were rated as being the single most valuable component.

With regard to possible effects of the evaluation design on the findings, we find no compelling reason to believe that the apparent superiority of the experimental instructional approach is an artifact of the design. Despite some nonequivalence among groups, the instructional approach effects, as measured by this particular pretest and posttest with these particular comparison groups, appear to be genuine. The findings add another increment to the accumulating data which suggest that student learning is enhanced by autotutorial teaching (Fisher, 1976; Fisher & MacWhinney, 1976; Kulik, 1975; Kulik et al., 1974).

Autotutorial instruction was first introduced in 1961 by Postlethwait for teaching introductory biology at Purdue (Postlethwait, 1964). The essential features of this approach, called audio-autotutorial instruction, include (1) a general assembly session for motivational lectures, (2) a small assembly session for recitation and quizzing, and (3) an independent study session which utilizes programmed audio tapes to integrate lecture and laboratory material.

In 1968, Keller described a Personalized System of Instruction (PSI), designed to provide reinforcement contingencies or built-in rewards at every step of the educational process (Keller, 1968). The primary features of PSI are (1) self-pacing, (2) mastery grading, (3) lectures for motivation, (4) written instructional materials, (5) tutoring by proctors, and (6) frequent quizzing with immediate feedback.

Recent reviews of the effectiveness of audio-autotutorial and PSI methods in science teaching show that overall, content learning as measured by final examinations nearly always equals, and usually exceeds, performance in lecture sections (Fisher, 1976; Fisher & MacWhinney, 1976; Kulik, 1975; Kulik et al., 1974).

The video-AT method of instruction differs from PSI primarily in its lock-step testing and nonmastery grading, and from the reduced emphasis on tutor-student relationships. It differs from most audio-tutorial methods in its use of the dynamic video medium and in the absence of integrated laboratory materials (except by video demonstrations).

Precisely what it is about AT methods that accounts for their teaching effectiveness remains to be elucidated. The consistency of the findings regarding student achievement with AT (and the consistently high regard students have for these approaches) is particularly interesting when one considers how many variants masquerade under the AT umbrella. Future research efforts might profitably be directed toward analysis of the relative contributions of the various aspects of AT methods. Several attributes seem to be included in nearly all AT approaches: shift of student role from passive recipient to active participant, statement of objectives in behavioral terms, provision of opportunity for students to study at own convenience and (to some degree) pace, and frequent testing and feedback.

The variables other than instructional method which were found to contribute significantly to the multiple-regression analysis in this study hold few surprises. Previous measures of academic success are known to be potent predictors of academic performance (Trent & Cohen in Travers, 1973). Grade-point average and pretest score were the second and third most potent predictors here, exceeded only by instructional method. The temporary decline in academic performance among new transfer students in the University of California has been described previously (Knoell & Medsker, 1964a, 1964b). Two corollary measures, quarters in the UC system (positive beta weight) and quarters in community college (negative beta weight) were the fourth and eighth most potent variables here. The level of motivation among students is known to affect academic achievement (Lavin, 1965). The perceived necessity of college and expected grade in the course, the sixth and seventh most potent variables identified here, are presumably related to motivation. So, perhaps, is the certainty of occupational preference, the fifth most potent variable. What is surprising, however, is that the latter has a negative beta weight. More than half of the students in all three genetics courses identify themselves as pre-medical students. As a group, they are ambitious and highly competitive. The grade in genetics is supposedly one of the important variables considered by medical school admission committees. We would have expected the certainty of occupational preference to have a positive beta weight.

Several researchers have reported trait/treatment interactions, such that autotutorial instruction favors the poorer student and lecture instruction favors the better student (Diederich & Macklin, 1973). There is no evidence for such interaction in the data reported here. When campus groups were divided into thirds on the basis of pretest scores, the video-AT group achieved significantly higher scores than the lecture groups on the posttest in every stratum. The video-AT campus group in the lowest third on the pretest did as well as the lecture campus groups in the highest third on the pretest.

Summary

This study compares student achievement in an upper-division introductory science course taught by the video-autotutorial method with that in two comparable courses taught by the lecture/discussion method. The student groups consisted of intact classes on three different University of California campuses. Three conclusions emerge: (1) the autotutorial instructional method seems helpful to students at all ability levels; at each level, video-AT students outperform those taught conventionally. (2) In determining final performance, instructional method seems to be a more important influence than individual student characteristics. (3) The video-autotutorial method seems to be a particularly effective method of instruction; the performance differential between the video-AT and conventional courses suggest improvements of at least one standard deviation, whereas Kulik and Kulik (1975) report that, on the average, about 2/3 of a standard deviation is observed between individualized and conventional approaches.

What are the implications of these findings for college science teaching? Perhaps the message on the bottom line is that significant improvements in teaching effectiveness can be achieved by meaningful research and development efforts. Much more is needed to ensure that higher education is the quality product we would like it to be.

The video-autotutorial method seems particularly effective in this situation, but it is obviously not appropriate for all situations. Such an approach is probably best suited for introductory classes which emphasize a lot of basic terminology, facts, principles, and relatively simple concepts, and in which the subject matter has intrinsic visual interest. The cost and complexity of producing video modules must be weighed against the potential number of students who would use them.

The success of the video-AT course is presumably not so much a function of its technology, however, as of the *pedagogical strategies* that it employs. The underlying implication is that perhaps college science teachers need to concern themselves with process as well as content. Our data do not allow us to rank the strategies we used in order of potency, but we can identify the principle ways in which the video-AT course differed from the comparison lecture courses, and suggest the theoretical strength of each strategy.

1. *Individualization* allows students to work at times which are optimum for them, determine (within limits) the pace of their learning, vary their pace through the course according to their familiarity with each topic, and adjust their study habits to suit their personal preferences and their own learning styles (see, for example, Cross, 1976).

2. *Frequent testing* permits students to assess their progress often and to adjust their study regimens according to the competency level they desire. It also commits them to study frequently.

3. *Options for improving learning* are an important adjunct to frequent testing. Our course offered (a) unlimited access to faculty and teaching assistants, (b) repetitive viewing of modules, (c) practice problems with answers and often detailed solutions, (d) textbooks, (e) syllabus, and (f) other recommended reading sources.

4. *Careful explication of instructional objectives* provides both teachers and students with a clear idea of the specific teaching/learning goals for each topic and for the course as a whole.

5. *Precision and clarity* of presentation are achieved in instructional materials which are prepared in advance (e.g., textbooks, syllabi, video tapes, audiotape/slides) and which involve peer review and revision *prior* to their use in the classroom.

6. *Analysis and revision* of instructional materials after they are used in the classroom results in an enhancement of quality with time.

7. *Individual feedback and reinforcement* presumably generates high levels of commitment and motivation among the students, so that they may work harder (and perhaps longer) at learning.

8. *Mathemagenic devices* help students learn from prose and perhaps from televised information: (a) presentation of *advance organizers,* (b) *emphasis* on important points (as by underlining), (c) provision of *objectives* specifying what a student is expected to learn from a given unit of material, and (d) periodic presentation of *inserted questions* throughout a prose passage (see, for example, Faw & Waller, 1976).

Creative use of these strategies in college science teaching, combined with careful research and testing, is to be encouraged.

References

Campbell, D. L., & Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally & Co., 1963.

Cross, K. P. Responding to new learning needs. Paper presented at the University of California-California State Colleges and Universities Joint Conference on Entering Freshman, Lake Arrowhead, Calif., December 4, 1976. Available from Center for Research and Development in Higher Education, University of California-Berkeley.

Diederich, M. E., & Macklin, D. B. *A trait-treatment interaction in a college physics course.* Draft version. Ithaca, N.Y.: Cornell University, 1973.

Dubin, R., & Taveggia, T. C. *The teaching-learning paradox.* Eugene, Ore.: The Center for Advanced Study of Education Administration, 1969.

Duncan, O. D., Featherman, D. L., & Duncan, B. *Socioeconomic background and achievement.* New York: Seminar Press, 1972.

Faw, H. W., & Waller, T. G. Mathemagenic behaviours and efficiency in learning from prose materials: Review, critique, and recommendations. *Review of Educational Research,* 1976, **46**, 691-720.

Fisher, K. M. Individualized instruction. *Journal of College Science Teaching,* 1974, **4**, 27-30. (a)

Fisher, K. M. Annual Report to the Alfred P. Sloan Foundation. Davis, Calif.: University of California-Davis, Division of Biological Sciences, 1974. (b)

Fisher, K. M. How effective is AT teaching? *BioScience,* 1976, **26**, 691-697.

Fisher, K. M., & MacWhinney, B. AV autotutorial instruction: A review of evaluative research. *AV Communication Review,* 1976, **24**, 229-261.

Glass, G. V. Primary, secondary, and meta-analysis of research. *Educational Researcher,* November, 1976, **5** (10), 3-8.

Keller, F. Goodbye, teacher.... *Journal of Applied Behavioral Analysis,* 1968, **1**, 79-88.

Knoell, D., & Medsker, L. L. Articulation between 2-year and 4-year colleges. Unpublished manuscript. Berkeley, Calif.: University of California, Berkeley Center for the Study of Higher Education, 1964. (a)

Knoell, D., & Medsker, L. L. Factors affecting performance of transfer students from 2 to 4 year colleges: With implications for coordination and articulation. Unpublished manuscript. Berkeley, Calif.: University of California, Berkeley, Center for the Study of Higher Education, 1964. (b)

Kulik, J. A., & Kulik, C. L. Effectiveness of the Personalized System of Instruction. *Engineering Education,* 1975, **66** (3), 228-231.

Kulik, J. A., Kulik, C. L., & Carmichael, K. The Keller plan in science teaching. *Science*, 1974, **183**, 379-383.

Lavin, D. D. *The prediction of academic performance: A theoretical analysis and review of research.* New York: Russell Sage Foundation, 1965.

Mayeski, G. W., Wisler, C. E., Beaton, A., Jr., Weinfeld, F. D., Cohen, W. M., Okada, T., Proshek, J. M., & Tabler, K. A. *A study of our nation's schools (a working paper).* Washington, D.C.: Department of Health, Education, and Welfare, Office of Education, 1970.

Nie, H. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. *Statistical package for the social sciences.* New York: McGraw-Hill, 1970.

Postlethwait, S. N. Practical audio programmed instruction. *Educational Screen and Audiovisual Guide,* 1974, 82-83.

Trent, J. W., & Cohen, A. M. Research on teaching in higher education. In R. M. W. Travers (Ed.), *Second handbook of research on teaching.* A project of The American Educational Research Association. Chicago: Rand McNally & Co., 1973.

Manuscript accepted June 1, 1977