DOCUMENT ARSUME

ED 031 966 FL 001 049

By-Morrison, H.W.; Adams, E.N.

Pilot Study of a CAI Laboratory in German.

National Federation of Modern Language Teachers Association.
Pub Date May 68
Note-9p.
Journal Cit-The Modern Language Journal; v52 n5 p279-287 May 1968
EDRS Price MF-$0.25 HC-$0.55
Descriptors-Achievement Tests, Aptitude Tests, *Audiolingual Methods, *College Language Programs, *Computer Assisted Instruction, *German, Grades (Scholastic), Language Instruction, *Language Laboratories, Student Opinion, Tables (Data)
Identifiers-State University of New York, Stony Brook

Comparisons are made between two introductory German sections at the State University of New York, Stony Brook, one of which had a conventional language laboratory and the other a computer assisted instruction laboratory. A brief description of the instructional arrangements is followed by descriptions of experimental comparisons, making up the bulk of the article. Sections are compared on language aptitude, overall academic achievement, course grades, achievement in German, and student opinion. For a related document see ED 016 978. (AF)

The Modern Language Journal

Volume LII MAY • 1968 Number 5

Published by

The National Federation of Modern Language Teachers Associations

Pilot Study of a CAI Laboratory in German

H. W. MORRISON* *and* E. N. ADAMS, *IBM Watson Research Center, Yorktown Heights, New York*

INTRODUCTION

THIS paper gives an abridged description of the principal results of an operational experiment in which one section of introductory German students received laboratory practice and remediation through use of computer assisted instruction (CAI), rather than through a conventional language laboratory. This class was taught by Professor F. A. Ruplin at the State University of New York, Stony Brook, during the 1966–1967 academic year, using a preliminary form of a German program prepared by the CAI research group of the International Business Machines Corporation.

The CAI language laboratory program has been described earlier;[1] in this experiment typical lessons contained three or four substitution-transformation modules, one English-German translation module, and either a dictation or aural comprehension module as described there.

Instructional Arrangements

Ruplin and Russell[2] have described the arrangements and rationale for instruction in the CAI section. Students in the CAI section met their instructor for three 50-minute class periods each week where they were taught by

* Now at State University of New York at Stony Brook.

[1] E. N. Adams, H. W. Morrison, and J. M. Reddy, "Conversation With a Computer as a Technique of Language Instruction," *The Modern Language Journal*, Vol. VII, No. 1 (January, 1968), pp. 3–16.

[2] F. A. Ruplin and J. R. Russell, "A Type of Computer Assisted Instruction," *The German Quarterly*, Vol. XVI, No. 1 (January, 1968), pp. 84–88.

the direct method including audio-lingual pattern drills. There was no written homework and practically no class time was spent on writing, translation, spelling, vocabulary, or reading. Recitation which emphasized facility in writing German was scheduled for two 50-minute periods each week at a CAI instructional station (an IBM 1050 with auxiliary tape recorder and slide projector). Students could schedule additional time if terminals were available.

In the CAI laboratory each student proceeded at his own pace, working on one unit of instruction at a time and going on to the next unit only after satisfactory proficiency had been demonstrated. Students were encouraged by their instructor to complete the exercises and to maintain progress if they fell behind, but they were not required to do any CAI work and they were told that performance scores from the CAI exercises would not be used in determining their course grades.

For purposes of comparison, data were collected from a second introductory German section in which the same instructor, Professor Ruplin, taught by the audio-lingual method (ALM) used in all other sections at Stony Brook. This ALM section had three class meetings and two 25-minute conventional language laboratory periods scheduled each week, and used a text[3] designed for ALM instruction. It was intended at the beginning of the year to compare final achievement of students in the two original sections. For this reason students were assigned to both sections through the normal registration procedure, which is effectively random. At mid-year the CAI section was maintained essentially intact through second semester registration; however, there was a substantial turnover of the ALM section as a result of rescheduling for second semester, so that only four of the original ALM students remained in the section at year end.

Experimental Comparisons

The students in the CAI and ALM sections were compared in terms of language aptitude, overall academic achievement, course grades, tests of German achievement at the end of the course, and student opinion of the two kinds of laboratories. In addition, data on CAI utilization and certain performance records were collected and studied in relation to the other variables.

This pilot experiment was conceived as an exploratory comparison rather than as a controlled test of CAI's effectiveness: besides the differences of treatment implied by "CAI laboratory vs. conventional language laboratory" other potentially important differences were: textbooks (hence linguistic content), method of classroom instruction, examinations (different because texts and methods were different), and possible effects of novelty on motivation. In addition, because of the turnover in the ALM section noted above, in effect, pre-testing was done on one ALM section, post-testing on a different one. Thus, an interpretation of year-end test comparisons is dependent on the extent to which the two ALM sections are typical of the Stony Brook population.

Even in these circumstances, comparisons of CAI and ALM sections may at least indicate effects to look for in larger and better controlled experiments. To this end we compared:

1. CAI students with ALM students on the basis of aptitude, grade point average and other characteristics;
2. the two initial sections on the basis of course grades at mid-year;
3. the two final sections on the basis of course grades and standardized test measures at year-end; and
4. patterns of correlation between various final achievement measures.

Taken together our results suggest that the students in the CAI section

1. were comparable to those of the ALM sections in language aptitude and general academic achievement;
2. without using the conventional language laboratory acquired the skills of speaking and listening about as well as ALM students; and
3. without specific classroom instruction in reading and writing acquired these skills as well as or better than ALM students.

[3] G. A. C. Scherer and H. H. Wängler, *Contemporary German*, New York: McGraw-Hill, 1966.

Comparability of Sections

The makeup of the CAI and ALM groups was examined and found to be similar in terms of college class and overall cumulative grade point average (GPA) achieved at the end of the 1966–1967 academic year. (GPA includes grades received in German.) Figure 1 shows the distribution of GPA for each group in the study. To plot the figure the student's GPA is converted into a percentile standing in his *college class* and then plotted above his percentile standing in his *experimental group*. This plot should approximate a diagonal line for a typical group. A statistical analysis shows that there were no significant differences in GPA, which suggests that the three groups, CAI, first semester ALM, and second semester ALM had similar general academic achievement.

The complete Modern Language Aptitude Test (MLAT)[4] was administered to both the CAI and the first ALM section at the beginning of the first semester.[5] Test scores were available for twenty-one CAI students and sixteen ALM students who completed the first semester. The mean and standard deviation of raw scores in the CAI section were 129.9 and 14.2 respectively, in the ALM section 132.7 and 13.3 respectively. These scores were somewhat higher and less variable than for the test standardization group reported by Carroll and Sapon[6] and the experimental groups described by Scherer and Wertheimer.[7] Students for whom MLAT scores were not available received first semester German grades of C, C, D, D in the CAI section, and A, A, C, F, F, F in the ALM section. The distribution of MLAT scores is shown in Figure 2.

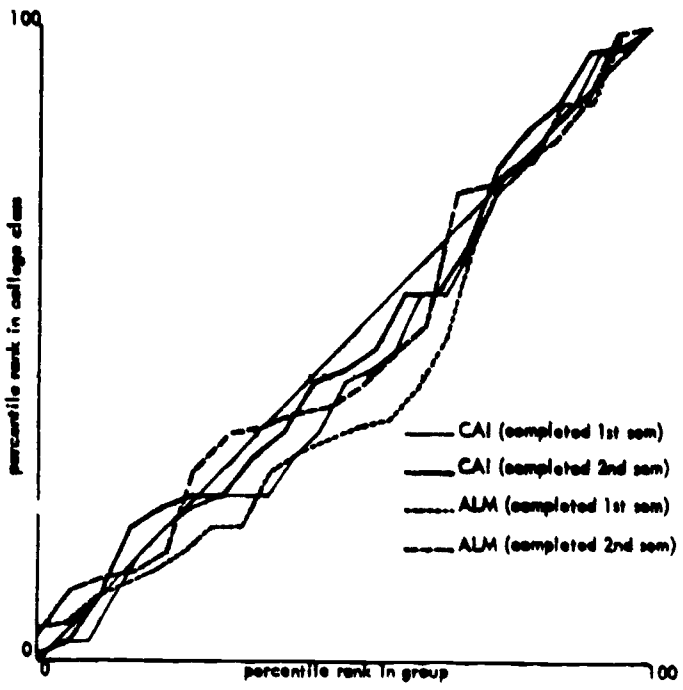

FIGURE 1. Cumulative Grade Point Average at End of 2nd Semester.

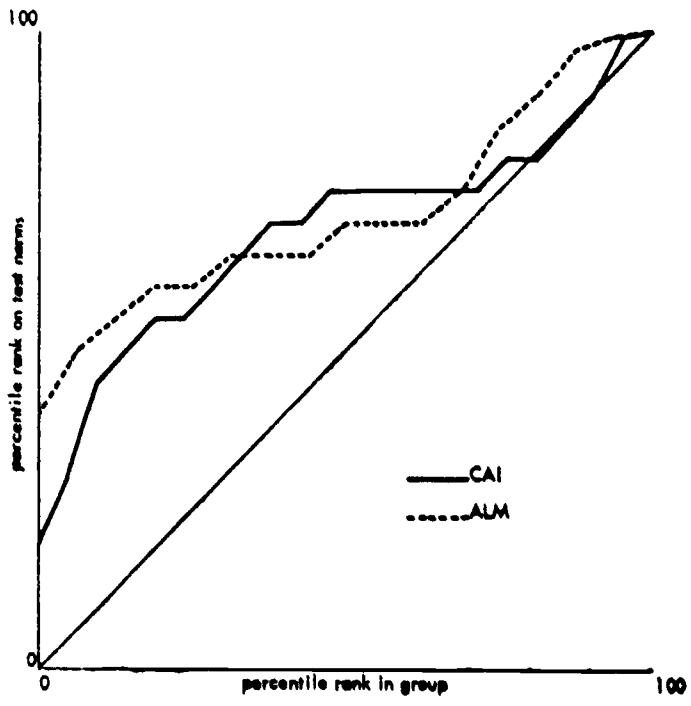

FIGURE 2. Modern Language Aptitude Test Scores.

Both GPA and MLAT are reasonable measures for establishing comparability of the sections. Within each section graded by Professor Ruplin, MLAT scores had product-moment correlations of .37 to .42 with first and second semester final grades; GPA (which is contaminated with the German grades) had correlations between .54 and .74 with first- and second-semester final grades.

In order to further establish the comparability of the first- and second-semester ALM sections and the degree to which each was representative of Stony Brook students in introductory German, a series of comparisons was made of grades received in the German course. The distribution of grades for the vari-

[4] J. B. Carroll and S. M. Sapon, *Modern Language Aptitude Test Manual*, New York: The Psychological Corporation, 1959.

[5] A few of the students were not tested until later in the year; however, instruction apparently does not markedly affect MLAT scores (see *Ibid.*, p. 20).

[6] Carroll and Sapon, *op. cit.*, p. 11.

[7] G. A. C. Scherer and M. Wertheimer, *A Psycholinguistic Experiment in Foreign-Language Teaching*, New York: McGraw-Hill, 1964, p. 43.

TABLE 1A

FIRST SEMESTER GRADES

	W/I	F	D	C	B	A
Stony Brook ALM students outside study	12%	10%	12%	25%	28%	13% (N=226)
First-Semester ALM section	4%	22%	13%	30%	22%	9% (N=23)
Second-Semester ALM section	—	—	6%	29%	35%	29% (N=17)
First-Semester ALM students who registered for second semester	—	—	7%	43%	36%	14% (N=14)
CAI section	4%	0%	8%	38%	35%	15% (N=26)

TABLE 1B

SECOND SEMESTER GRADES

	W/I	F	D	C	B	A
ALM students outside study	14%	3%	8%	31%	26%	18% (N=163)
Second-semester ALM section	6%	0%	24%	35%	24%	12% (N=17)
First-semester ALM students who registered for second semester	14%	7%	14%	36%	21%	7% (N=14)
CAI students who registered in CAI second semester	17%	0%	17%	33%	21%	12% (N=24)

ous groups of interest for comparison are summarized in Tables 1A and 1B.

In summary the above comparisons indicate that within sampling errors in the data:

1. The CAI and first semester ALM students are comparable on both GPA and language aptitude.
2. The second-semester ALM students are comparable to or slightly superior to the CAI students on the basis of GPA, and to first semester ALM students on the basis of German grades.

Achievement Comparisons

Grades: The two pre-tested groups (CAI and first-semester ALM) were compared on the basis of academic achievement at mid-year. Because different examinations were used in the two sections, course grades are not a basis for an objective comparison between the two groups. However, for both sections these grades are a measure of the same instructor's assessment of student achievement. For students who received a first-semester grade, the means based on a 0–4 scale were 2.60 for 25 CAI students and 1.82 for 22 ALM students; this difference indicating better learning by the CAI students was significant at the .05 level.[8] Second-semester grades were not significantly different for the twenty CAI and sixteen ALM students who completed the year course.

Attrition: Another measure of the effectiveness of instruction is the survival rate of students in the course. The data in Table 1 show that the attrition rate over the year was decidedly lower for the CAI section than for the other German students at Stony Brook. Only six of the original twenty-six CAI students were lost during the year including one who registered in an ALM section during the second semester; of the other five one was an early drop, and four of the five worked on fewer than four of the forty lessons in the CAI laboratory; of twenty-six students who began the year twenty finished it and had passing grades, for an overall survival rate of 77%. By contrast in the main body of

[8] All tests of significance in the text of this paper are based on τ or $\tau_{b,c}$; M. G. Kendall, *Rank Correlation Methods*, Second Edition, London: Griffin, 1955. Number of asterisks indicates the level of significance achieved: *designates $p<.05$, and **designates $p.<.01$ all based on 2-tailed tests.

ALM students at Stony Brook 22% were lost in the first semester and a further 17% were lost in the second; of 226 students who began the year only 135 finished it and had passing grades for an overall survival rate of 60%.

Standard Achievement Tests: Shortly before second-semester final examinations students in the three sections (the CAI section and both first and second ALM sections) were each offered $5.00 to attend a special session where the instructor administered Form LA of the MLA Cooperative Foreign Language Test[9] in German (CFLT). These tests were not scored until after final grades were reported. The instructor scored all sub-tests except writing; writing was scored by Professor John Russell, who was not familiar with the students.

Scores on the CFLT were obtained for nineteen of the twenty CAI students and twelve of the 16 ALM students who completed the second semester; in addition scores were obtained for three first-semester ALM students who had registered in other ALM sections and received grades of B, C, C during the second semester. In examining these scores it should be noted that the norms for this test were apparently based on administration about one month earlier.[10] Raw scores were used in comparing the CAI and ALM students, and percentiles based on first-year college norms[11] were used in comparisons with the test standardization group.

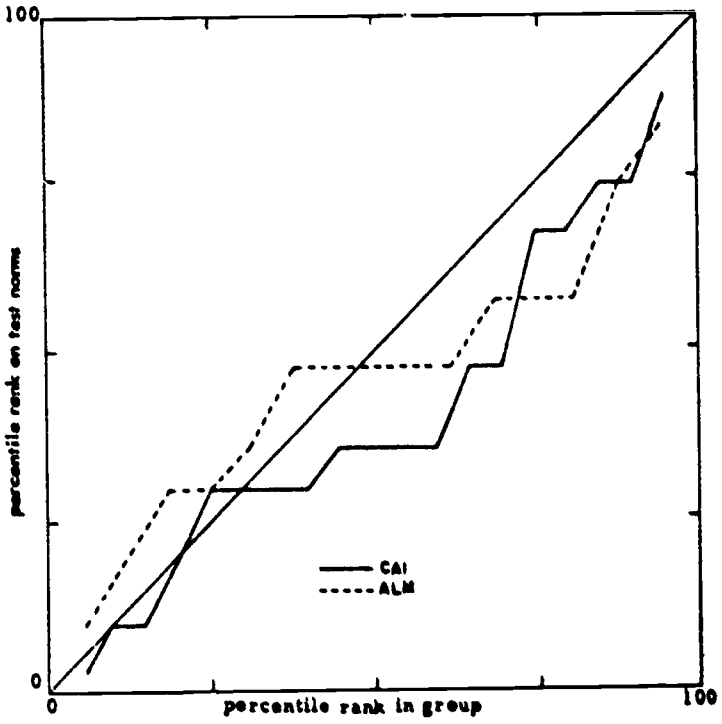

FIGURE 4. Listening Achievement.

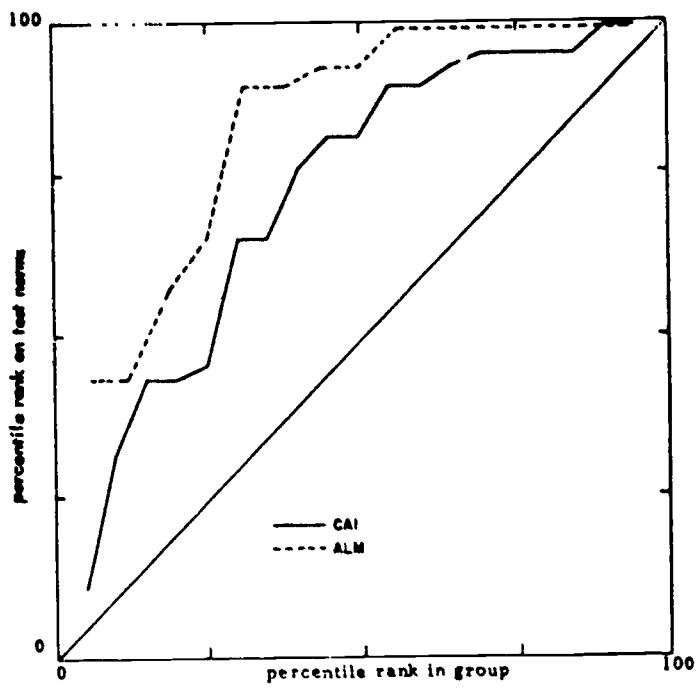

FIGURE 3. Speaking Achievement.

Figures 3–6 display the results of the CAI, ALM, and test standardization groups on the speaking, listening, writing and reading subtests. Several significant differences were observed:

1. On *speaking* both the CAI group** and the ALM group** were higher than the test standardization group.
2. On *writing* the CAI group was higher than both the ALM group** and the standardization group**.
3. On *reading* the ALM group was lower than the standardization group ($p<.06$).

In other comparisons of the two sections there is some indication that the ALM group was better in speaking and listening ($p<.30$) and that the CAI group was better in reading ($p<.13$).

It is doubtful that the observed superiority of CAI students in writing and to a lesser extent in reading is a manifestation of the effect Scherer and Wertheimer,[12] found in a

[9] *MLA-Cooperative Foreign Language Tests Handbook*, Princeton, New Jersey: Educational Testing Service, 1965.

[10] *Ibid.*, p. 9.

[11] *MLA-Cooperative Foreign Language Tests Booklet of Norms*, Princeton, New Jersey: Educational Testing Service, 1965, pp. 17–18.

[12] Scherer and Wertheimer, *op. cit.*, pp. 174–185.

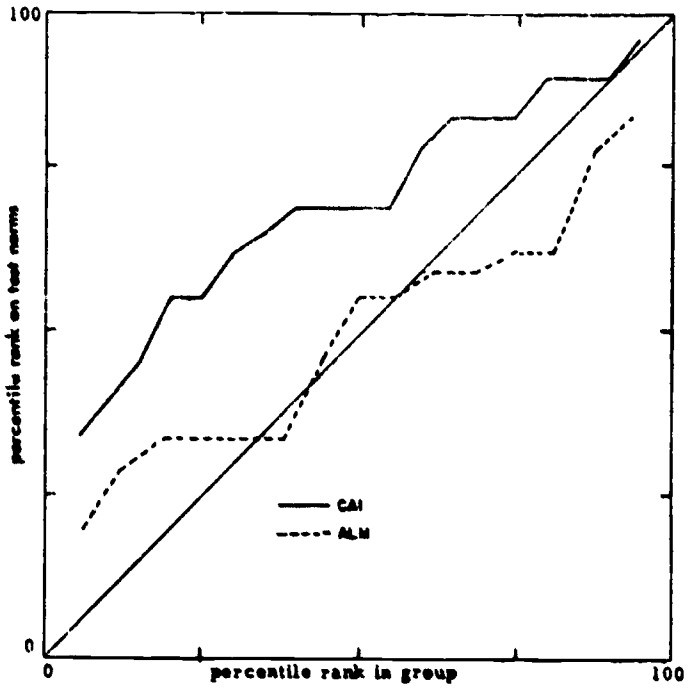

FIGURE 5. Writing Achievement.

comparison of Colorado students taught by a "traditional" method (superior on reading, writing German-English translation, and English-German translation) and "ALM" students (superior on listening and speaking) after two semesters: unlike the "traditional" courses, the CAI course involved no direct instruction in principles of grammar, provided practically no class time on German-to-English translation, spelling, or vocabulary, and did not require written homework. However, the superiority of CAI students on writing is not surprising, since most their laboratory practice involved writing.

Although CAI students spent no time in the ordinary audio-lingual language laboratory, they did not score significantly lower than ALM students in speaking and listening comprehension. If this finding proves valid[13] it represents another real difference from the findings of Scherer and Wertheimer for their "traditional" course. Actually such a result might be expected, since classroom instruction for the CAI section was by the direct method and involved extensive audio-lingual instruction, with its emphasis on pattern mastery, rather than by "traditional" grammar instruction proceeding from rules.

Correlations Between Measures: Using data for the CAI and ALM groups, we examined the correlations of the four language skills at year end with one another and with several predictor and criterion variables. A selection of these data are given in Table 2. Such correlations are a potential source of information about pattern of skills learned in the two groups. We will not discuss most of the data here,[14] since our samples are too small for reliable estimates of the correlations. One finding of interest was that for both CAI and ALM students there was a high correlation between academic grade and performance on the writing subtest of the CFLT. Second, for the ALM students there was a significant correlation between their final achievement on all the various CFLT subtests and their general academic aptitude as evidenced by GPA, whereas for the CAI students there was little correlation of this sort.

Internal Measures of Performance: One of the most important potential benefits of CAI is the effective individualization of remediation. In the most complex case this might include

[13] Because of the small number of students the statistical test is not very powerful. Other reasons for cautious interpretation include the known low reliability of the speaking subtest (*MLA-Cooperative Foreign Language Tests Handbook*, *op. cit.*, p. 22), the possibility that the instructor's scoring was influenced by knowledge of the students and the grades they received, and the possible failure of the test to discriminate effectively at the score level attained by many students.

[14] H. W. Morrison and E. N. Adams, *Pilot Study of a CAI Laboratory in German*, IBM Research Report RC-1974, 1967.

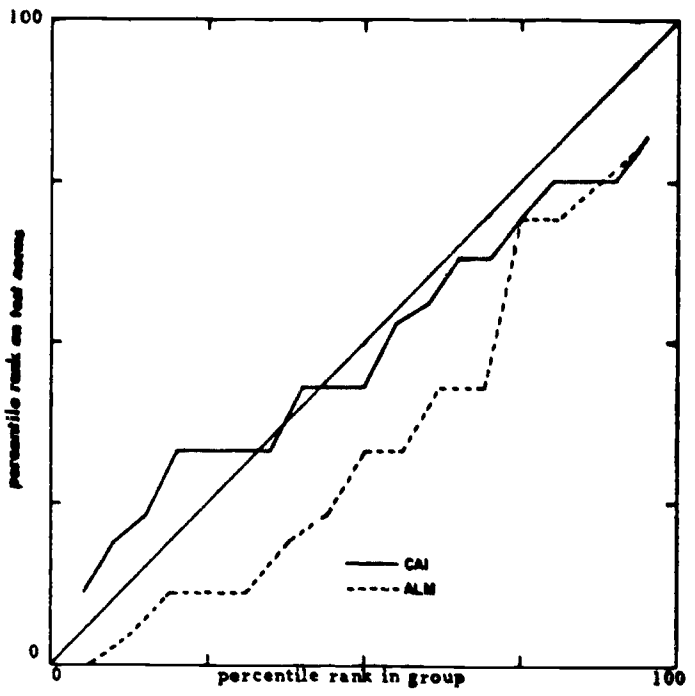

FIGURE 6. Reading Achievement.

TABLE 2

PRODUCT-MOMENT CORRELATIONS WITH ACHIEVEMENT SUBTESTS

	Speaking	Listening	Writing	Reading
A. ALM Section				
Speaking	—	.62*	.60	.76**
Listening	.62*	—	.51	.55
Writing	.60	.51	—	.60
Reading	.76**	.55	.60	—
2nd sem. grade	.55	.47	.73**	.42
GPA	.59*	.70($p < .06$)	.77**	.61*
B. CAI Section				
Speaking	—	.44	.73**	.31
Listening	.44	—	.36	.59*
Writing	.73**	.36	—	.37
Reading	.31	.59*	.37	—
2nd sem. grade	.68**	.20	.68**	.23
MLAT	.25	.13	.66**	.43
GPA	.44*	−.09	.31	−.25

"customized" problems aimed at each student's characteristic weakness, while in the simplest case it might involve no more than regulating the amount of practice. In any case the essential prerequisite for any individualization of instruction is the availability of frequent measures of proficiency by which to direct remedial strategy.

Day-to-day monitoring of proficiency can be accomplished by interspersing many tests throughout the learning program. A better procedure, where feasible, is to monitor progress continuously by means of internal measures of performance made on the learning tasks themselves. For that reason we sought mechanical measures of student activity which could serve as valid internal measures of achievement. A selection of our data are given in Table 3.

The mechanical indices examined were:

1. average entry time per response
2. total time used to enter responses
3. total number of responses
4. total number of items attempted
5. most advanced unit of course reached
6. "degree of match" of first attempt on constructed responses[15]

A rationale for each of the above indices might be found in terms of the importance of speed, accuracy, total amount of practice, total amount of material covered as an indication of progress. However, we did not attempt to develop any particular rationale but only to find indicators that would be pragmatically successful. As it turns out the "most advanced unit reached" and the "degree of match" were indices that correlated most closely with student achievement.

Our analysis was carried out on the performance data for the twenty students who remained at the end of the CAI course and received a final grade. At the end of the second semester these students were located between Lesson 20 and the end of Lesson 40 (mean = 31). In exercises which required entry of an entire German sentence (more than 70% of all modules) students had spent between 10.7 and 45.9 hours actually entering answers, and had made 1020–2862 entries (mean = 1718). Average entry times ranged from 36 to 95 seconds (mean = 56).

Final location in the CAI course correlated .53** with total entry time and .71** with number of attempts. Total entry time correlated .47* with number of attempts and .66** with average entry time. There was a (non-significant) negative relation ($r = -.32$) between total number of attempts and average entry time; this suggests that students who made more attempts were on the average faster. Such a relation might suggest a negative tradeoff between the former variables, as would be in-

[15] See E. N. Adams, H. W. Morrison and J. M. Reddy, *op. cit.*

TABLE 3

CORRELATIONS WITH INTERNAL DEGREES OF MATCH MEASURES

	Lesson 13 translation	Lesson 27 translation	Lessons 14–26 translation	Lessons 14–26 dictation	Lessons 14–26 substitution/ transformation
speaking	.71**	.58	.55**	.44*	.35
listening	.22	−.09	.38	.35	−.25
writing	.82**	.55*	.84*	.76**	.49*
reading	.18	−.10	.32	.30	.09
1st sem. grade	.55*	.44	.60**	.47	.39
2nd sem. grade	.55*	.67*	.69**	.60**	.62**
MLAT	.42	.18	.54*	.50*	.34
GPA	.32	.58	.34	.22	.42
Most advanced unit reached	.35	.15	.23	.10	.22

volved in choosing between "quick and dirty" and "slow but sure" strategies, or it might merely reflect the fact that on attempts subsequent to the first attempt on an item the student has a simpler task, partial copying which can be carried out more quickly. Of these particular internal measures only "final location" had a significant correlation with the achievement measures: with first semester grade (r=.72**), with second semester grade (.54*), with GPA (.61*).

"Degree of match" on constructed answers is a particularly plausible internal measure of current proficiency. We analysed this measure separately for different types of exercises. Lessons 13 and 27 were review or test lessons consisting of thirty translation items, none of which could be skipped. The measure for each of these lessons was the mean score (calculated by the feedback algorithm; see)[16] on the first attempt to answer each item. The same measure was calculated for each translation, dictation, and substitution-transformation module in Lessons 14–26, and averaged over modules for each type of exercise. These measures are highly correlated with one another.

Even though the number of cases is small, all the correlations of "degree of match" with final grade are significant: clearly "degree of match" measures seem to be valid internal indicators of achievement. Not surprisingly, the correlations of "degree of match" were higher for speaking and writing scores than for listening and reading tests. On the other hand, the correlations with two other successful "predictors" of final grades, GPA and final location in the course, were not significant.

The high correlations of the "degree of match" indices with the various achievement measures indicate that they may be valid measures of current proficiency. This is an especially valuable finding, since a principal control strategy in the German laboratory is to regulate the amount of practice on the basis of current achievement as assessed by "degree of match" measures.

Student Opinions

After about eight weeks of instruction during the first semester, students received a questionnaire for recording how much time they would ideally spend in various forms of German instruction: classroom, language laboratory, CAI, instructor conference, outside study along or with other students. Students were told that results would not be given to the instructor nor reported in a manner which made it possible to tell how any individual answered. Approximately half the students in each section responded, and even fewer replied on a second administration three weeks later. A tabulation of results showed CAI students reported more use of their laboratory, compared with ALM students, and CAI students wanted a greater increase in laboratory time (in an ideal course). The latter indirect measure of perceived useful-

[16] *Ibid.*

ness of the laboratory is consistent with the student's opinions expressed in the optional comments.

The few free comments made by CAI students on the laboratory were judged to be positive. ALM student attitudes toward the conventional language laboratory were not so clear cut; they could be interpreted as neutral or negative on balance; thus three ALM students indicated that in an ideal course there would be no conventional laboratory time, but did not make any explicit reference to the laboratory in the form of optional comments. In both sections students were positive toward the instructor, and on balance neutral toward the overall method of instruction or course.