A Forum*

Language Laboratory Research: A Critique

IN TWO carefully conceived and painstakingly executed research projects, the Bureau of Audiovisual Instruction of the New York City Board of Education has succeeded in shedding additional light on what was and what remains a highly sensitive issue, the effectiveness of the language laboratory.¹ This article will deal exclusively with a discussion of the second of the two language laboratory experiments conducted by the Bureau, The Relative Effectiveness of Four Types of Language Laboratory Experiences (referred to in the Bureau article as the Time and Equipment Study), conducted during the academic year, 1962–1963. The final report was submitted to the Division of Research of the New York State Education Department—which partially supported both experiments—in July, 1963.

Brief reviews of this research project have appeared in the professional literature and in a number of foreign language newsletters. This experiment and the one which preceded it were summarized in the October-November 1963 issue of A-V Learning, a Bureau publication. The most comprehensive reviews of the Bureau experiments are probably those prepared by Professor Edward M. Stack in the February, 1964 issue of Communicate, the Electronic Teaching Laboratories newsletter and in the April, 1964 issue of The Modern Language Journal. Professor Stack’s second review is largely a rewritten version of his earlier review. In both instances, Professor Stack, who has been severely critical of recent language laboratory research, describes the Bureau experiments in glowing terms, ignoring some basic and rather apparent weaknesses, particularly in the reporting and in the interpretation of some of the findings.

Very briefly, the Bureau report was an attempt to determine in which type of laboratory installation and in which time pattern the most significant results could be achieved: record-playback daily; audio-active daily; record-playback once weekly; audio-active once weekly; and the control group. The pupils in the control group made no use whatsoever of language laboratory equipment. These five variations will be referred to hereafter as RPD, AAD, RPO, AAO, and control, respectively.

The final report contains a "Discussion of Limitations" which was neither included nor referred to in the rather lengthy and detailed Bureau article. In the RPO classrooms, only ten of the pupil stations were equipped to be used for record-playback activities. Hence, only ten pupils in each of these classes were to be directly involved in the experiment. Since the remaining twenty pupils were not in control groups, they were permitted to listen to the programmed tapes and to participate audio-actively. It was decided to add these pupils to the experiment and to form a fourth experimental condition, audio-active weekly (AAO). This determination was made after the experiment was already in progress.² To decide to include these pupils in the experiment while it was already in progress and then to note as an experimental deficiency the fact that two types of experience, RPO and AAO, had to be conducted in the same classroom jeopardizes the scientific quality of this study. Although we need to know more about the simultaneous exploitation of various types of language laboratory equipment, it is certain that no responsible language educators would recommend one audio-active laboratory period per week.

The findings were divided into three areas: (1) speech skills, (2) listening comprehension skills, and (3) pencil and paper skills. In the area of speech skills, pupils were tested (1) by reading a passage aloud on tape and (2) by recording answers to questions. In the oral reading passage, pupils were tested on five separate subscores: (1) global rating of quality of speech; (2) intonation; (3) pronunciation; (4) fluency; and (5) mimicry.

Overall, or global quality of speech is not defined in the article other than to state that it was rated independently of the other four speech subscores. In the final report, however, it is defined on page 12 as including word grouping, rhythm, and intonation, and on page 15 as including fluency, rhythm, intonation,

* It has often been noted in these pages that our main purpose in editing this journal is to present a forum for exchange of ideas in order to achieve a more precise definition of the knowledge which can properly be classified as foreign language teaching and learning. Previous statements in this journal have elicited these studied replies which in accordance with the above considerations are published here. The professional respect which motivates each author increases our own pleasure in creating the possibility of such an exchange of information.—B.F.R.


² The original title of the research project was Study of Three Types of Language Laboratory Experiences.
and pronunciation. When one considers that the most emphasized gains in achievement were obtained in the ill-defined area of global rating of quality of speech, it is doubly unfortunate that the final report failed to include a section devoted to a definition of the terms used throughout the report. In the question and answer test, pupils were measured in terms of fluency of response, appropriateness of response, and grammatical accuracy of response.

A brief examination of the findings in the area of speech skills is necessary at this point. In the oral reading passage, no statistically significant differences were found in the gains made in the separate measures of pronunciation, intonation, mimicry, and fluency among the experimental groups and between the experimental groups and the control groups. It is not quite clear in the Bureau article or in the final report filed in Albany just how fluency was measured, other than to state that it was measured by word count in a thirty-second period. One must assume that the examiner only recorded the number of words uttered by a pupil within the space of thirty seconds. Despite the lack of statistical significance among the differences in the gains made in the four above-mentioned subtests, in the separate measure of global quality of speech the RPD group achieved a statistically significant gain over the four remaining groups. No explanation is offered for this apparent anomaly. It is difficult, of course, to reject the statistical significance of the findings, but it is likewise difficult to accept their pedagogical significance. There is no question, however, of the trend toward greater, if not significantly greater, gains made by the RPD group in each of the other speech subtests. Still another question remains to be answered concerning the overall quality of speech ratings. In Chart V on page 414 and in Chart V-A on page 415 of The Modern Language Journal article, the RPO condition is listed as having made the third greatest gains in overall rating. In the October-November 1963 issue of A-V Learning, however, the AAO condition is listed as having made the third greatest gains in overall rating.

In the oral answers to oral questions, the daily laboratory classes made significantly greater gains than the weekly laboratory classes. What is more surprising, however, is the fact that on each subtest (1) the control group made significantly greater gains than the weekly laboratory classes, and (2) the control group made greater gains than the daily laboratory groups on each subtest. The control group gains over the daily laboratory groups were not statistically significant, admittedly, but they represent a finding which was virtually ignored in the final report and in the Bureau-prepared summary of the experiment. The November, 1964 article acknowledges the superiority of the gains made by the control group over the two daily laboratory groups. If we are to attach major importance to the nonsignificant trends in the speech subtests, then certainly all nonsignificant trends should be reported and discussed openly. Moreover, it is stated that "in several measures, one or more laboratory groups are superior to the control." It appears, however, that no more than one laboratory group is significantly superior to the control in any single measure.

In the pencil and paper skills, the Cooperative French Achievement Test appropriate to this level of instruction was administered in September, 1962 and in May, 1963. Separate scores in reading comprehension, vocabulary, and grammar were obtained, as well as a composite, or total score. In the reading comprehension findings, the final report states that there were no significant differences in gains among the RPD, AAD, RPO, and control groups. However, in the aforementioned issue of A-V Learning and in Chart V of the November, 1964 article, it is stated that in the area of reading comprehension the RPD and the control groups achieved significantly greater gains than the other three groups.

In the "Implications of the Findings," which is contained in the final report and incorporated into the "Discussion" in the Bureau article, it is stated that the superior results in speaking and in rapid listening comprehension were achieved without detriment to the traditional pencil and paper skills, although 50 percent of the class time was devoted to these laboratory activities. The superior results in speaking are those which were achieved in the area of global quality of speech which, as we have seen, was defined rather nebulously in the final report, and not defined at all in the Bureau article. Moreover, the findings in the other four subtests do not corroborate the findings in the global quality of speech subtest. Therefore, one cannot accept the gains made by the RPD group in global quality of speech as being truly indicative of superior results in speaking over the other groups. It is true, however, that there was no apparent deterioration of traditional pencil and paper skills attributable to the considerable amount of class time devoted to language laboratory activities. This is a valid conclusion and one which is far-reaching in implication. The final report attributes the significantly greater gains made by the RPD over the AAD group in the global quality of speech to the record-playback feature. In view of the fact that no statistical significance was found in the gains made by the RPD group in pronunciation, intonation, mimicry, and fluency, it is rather difficult to ac-

*Lorge, p. 415.
cept the conclusion that it was the record-playback feature which produced the significant difference in the gains made in the global quality of speech sub-test. In the “Summary of Results” contained in the final report, the researcher repeats the statistical significance in the gains made by the RPD groups over the control group and over the other experimental groups. Since many school administrators tend to read only the summaries of research reports, it is unfortunate that this summary is not somewhat more detailed. Certainly a busy school administrator reading this summary is likely to equate “overall quality of speech” with general speaking ability, and would doubtless conclude that the laboratory he is planning to install must contain record-playback facilities in each position if the speaking skill is to be developed in his language pupils. In the “Conclusion” of the Bureau article overall quality of speech is referred to as simply “speech.” The findings of this research report do not conclusively substantiate the view that the record-playback feature is necessary in a secondary school language laboratory. Moreover, let us not lose sight of the fact that the overall rating gains were made on the basis of a sight-reading passage which is, at best, an unsatisfactory single measure of speaking ability. Another researcher could conclude that the control group was superior to the experimental groups in speaking ability on the basis of the gains made in the oral answers to oral questions, a no less acceptable measure of speaking ability than sight-reading.

Special-purpose aid in general and the language laboratory in particular are not without their critics. The most vocal critics are not found among the rank and file of the profession, but there are strong undercurrents of skepticism and incipient disenchantment among language teachers vis-à-vis the language laboratory. It is incumbent upon laboratory researchers, therefore, to interpret their findings conservatively and cautiously and to report all but incontrovertible findings as tentative at best.

Jerald R. Green
New York State
Education Department

Comments on “Language Laboratory Research: A Critique”

The experiments conducted by the Bureau of Audio-Visual Instruction have been the basis of one report and two articles:

1. The official report, submitted to the Research Division of the New York State Education Department Albany, New York, in July, 1963. This is not available for distribution; there are no remaining copies.

2. A four-page article giving in substance the content of the original report. This article appeared as a supplement to the Bureau’s publication A-V Learning in November, 1963. This Bureau article is available on request to the Bureau of Audio-Visual Instruction (BAVI), 131 Livingston Street, Brooklyn 1, N. Y.


The article published in The Modern Language Journal (3) is not intended to be coextensive with either the official report (1) or the Bureau article (2).

Unfortunately the author of the critique has in effect chosen to review all three articles at the same time. It is somewhat confusing to follow his critique, as he uses the term “Bureau article” for both items 2 and 3 above. Occasionally he uses the term “Modern Language Journal article,” and sometimes “A-V Learning.” This may disorient the reader who does not know the Bureau article, and even more the reader who does know it, and who will wonder, for example, why it is said that the Bureau article does not contain the section known as Limitations, when in fact this item does appear in the Bureau article.

The writer of the critique objects to the addition of the twenty audio-active students after the beginning of the experiment. This criticism would have been valid if we had taken on a new group of students. But we found, for administrative reasons, that it was easier to pretest the entire class rather than certain individuals in the class. Therefore, we had the necessary data on the audio-active students who had been practicing in the laboratory together with the record-playback students. It seemed a pity to let the data go to waste. We, therefore, established a fourth group, and no change was made in the program of the audio-actives.

The critique implies that no attention was given to the scores which were highest in the control group. On the contrary, it is clearly stated that in the question-answer test, the control made significantly greater gains than two of the laboratory groups, and greater gains to a non-significant degree than the two other laboratory groups. Quotation from the official report, page 23: “In some aspects the control group made greater gains than the once-a-week laboratory group, i.e., in oral responses to questions.”