J. EDUCATIONAL TECHNOLOGY SYSTEMS, Vol. 15(1), 1986-87

REMOTE SITE INSTRUCTION IN PHYSICS: A TEST OF THE EFFECTIVENESS OF A NEW TEACHING TECHNOLOGY

SIDNEY RUDOLPH
Department of Physics

MICHAEL K. GARDNER
Department of Educational Psychology

University of Utah

ABSTRACT

This study tested the effectiveness of a remote site instructional system utilizing audio-graphic teleconferencing technology. The subject matter taught was high school level physics. The test site was Bountiful High School in Bountiful, Utah. The study consisted of a treatment-control group comparison in which the treatment group (9 males and 2 females) received the remotely delivered instruction and the control group (14 males and 3 females) received traditional high school physics instruction. The dependent variable was performance on physics tests that were administered periodically during the intervention period. A two-step hierarchical multiple regression was used to analyze the data. During the first step, the experimental and control groups were equated for pre-existing differences in student ability. During the second step, the two groups were tested for differences in performance due to the experimental manipulation. No significant differences in performance were found. The results lead us to believe that audio-graphic teleconferencing is an effective medium for delivering remote instruction.

A number of presently intractable problems in public education have stimulated recent interest in alternatives to the traditional "live-teacher-in-the-classroom" technique of delivering instruction [1]. Recent reports, such as *A Nation at Risk* [2] and *High School* [3], have heightened public awareness of problems in public education. This awareness has pressured local school districts, state offices of education, and institutions of higher education to respond in ways that have produced additional problems. For example, increased high school graduation requirements have led to increased enrollments in a number of academic disciplines. The effect is to exacerbate an already severe teacher shortage, especially in disciplines such as mathematics and science [4].

Many state offices of education have raised teacher certification requirements in response to public perceptions that too many unqualified teachers populate public school classrooms. This change means that many teachers who have been teaching numerous subjects will be forced to pursue additional academic work in order to gain additional certifications. The problem is especially acute in rural schools. These schools have small class enrollments, and teachers are expected to teach as many as four or five preparations per day.

Remote populations are additionally disadvantaged in that qualified and certified teachers are reluctant to commit their entire teaching careers to isolated locales. In addition, the physical remoteness of such regions typically implies a much reduced resource base with which to support public education. Despite these problems, most residents of rural communities remain strongly attracted to these settings. A pressing question thus emerges: Can rural settings produce an "outstanding and integrated educational system to bridge the gap between time and distance across vast plains and remote mountainous areas" [5, p. 1]?

One of the possible routes to a resolution of the problems described above is the application of electronic telecommunications technology. Such technology can provide instructional opportunities by qualified, certified personnel where before such opportunities either did not exist or were poorly provided. The technology removes the necessity of having instructional providers live in the region in which they teach. Moreover, this instruction can be provided in a cost effective manner; a single instructor can potentially teach to more than one physical setting simultaneously. These advantages lead us to believe that remote site instruction via electronic telecommunications technology will become an increasingly popular choice among educators—especially rural educators—in the years to come.

WIDEBAND VERSUS NARROWBAND TECHNOLOGIES

The explosion of technology over the past few decades has resulted in a large number of different systems for delivering both audio and video information to learners who are removed from the instructor's location [6-10]. One useful technique for sorting these technologies is to consider the bandwidth they use. Bandwidths can be classified into one of two categories: wideband or narrowband.

Wideband techniques are ones requiring the frequency spectrum associated with television transmission. They allow transmission of the full range of audio and video information to distant learners, and potentially allow the use of two-way television. Two-way television would make it possible for instructors to see and hear students, as well as allowing students to see and hear instructors. Although there is already a rich tradition in the use of television and satellite transmission for providing educational opportunities [7], wideband techniques suffer one significant disadvantage: they require enormous sums of money to work effectively. On the hardware side, each remote location requires a number of monitors (and cameras, if the system is a two-way one), as well as a satellite or microwave reception dish (or special cabling, if the sites are linked directly). At the transmission site, a production studio, cameras, and a transmitter are all needed. This hardware is expensive to purchase and expensive to maintain. Additional expenses, such as satellite or microwave time, can also be quite costly. On the personnel side, a number of highly trained technicians are needed at the studio. Over time these personnel costs can overshadow even the large initial expenditures needed for the purchase of the hardware described above.

Narrowband techniques are considerably less expensive in terms of hardware and personnel than wideband techniques. However, narrowband techniques are presently more limiting in what can be transmitted between the learner and the instructor. Full television cannot be transmitted using narrowband techniques; however, a number of useful things, such as text, graphics, still video pictures, electronic overheads, video facsimiles, and computer data, can be transmitted in this way. Narrowband technology uses only a small fraction of the transmission spectrum utilized by wideband technology—that portion that can be transmitted over conventional telephone lines. Approaches that fall within the narrowband rubric are: 1) audio and audio-graphic teleconferencing, 2) freeze-frame or slow scan video, 3) radio broadcasting, 4) electronic mail, 5) teletext or videotext, and 6) computer conferencing.

Because of the limited budgets of most school districts, and the large cost differential between wideband and narrowband technologies, we believe that narrowband approaches hold the most hope for providing a quality educational experience for learners at remote sites. In this article we test an instructional implementation of one such narrowband approach: audio-graphic teleconferencing. Briefly, the system allowed for two-way audio communication between learners and their (remotely located) instructor, as well as for two-way graphic transmission via digitizing tablets, computer terminals, and monitors. Additionally, the instructor had access to graphic files that he had previously prepared. These files contained sample problems and examples. The particular implementation of this technology was developed by the senior author and is described in detail in Appendix A as well as elsewhere [11-13].

EVALUATING INSTRUCTIONAL EFFECTIVENESS OF REMOTE SITE TECHNOLOGIES

The literature on remote site instruction, sometimes called "distance learning" [7], is replete with descriptions of implementations of technologically-based systems [14-23]. Unfortunately, very few systems provide an evaluation of the effectiveness of the system as a means of delivering instruction. Among the exceptions to this trend are Rushton and Branson [24] and Weingand [25].

Rushton and Branson examined whether an audio-conferencing system in conjunction with prepared materials could provide effective continuing education in complex skills [24]. Subjects in this study were medical laboratory technicians. The null hypothesis was that remotely located students would perform at an equal or higher level than those trained traditionally. Of particular interest was whether satisfactory levels of skill attainment could be achieved in circumstances that limited or even eliminated opportunities for practice. The results of the study demonstrated that the null hypothesis could not be rejected; no significant differences were found between remotely instructed students and those taught in the traditional manner.

Weingand tested the effectiveness of using an audio-teleconferencing system, also in a continuing education context [25]. Students at the University of Wisconsin were offered the option of traditional or mediated instruction in public library administration. The null hypothesis was that there would be no significant difference between students provided traditional instruction and those provided instruction via the technological delivery system. Not only was there no significant difference between the two groups of students in Weingand's study, but the direction of the means favored the audio-teleconferencing group!

Though both Rushton and Branson [24] and Weingand [25] found no differences between audio-conferencing systems and live teacher-in-the-classroom instruction, both these studies dealt with adult learners who would be expected to be highly motivated to learn, regardless of the means of instructional delivery. The present study not only sought to test the effectiveness of a particular audio-graphic teleconferencing system [20], but also to extend the findings of Rushton and Branson [24] and Weingand [25] to the high school age group. Given the large number of students in this age range, it would be comforting to know that this group can also benefit from remote instruction via narrowband systems. Indeed, we earlier argued for the use of electronic telecommunication technology to help solve the pressing problem of providing quality educational experiences to rural students. The ability of high school age students to benefit from such delivery systems is a necessary precondition to our argument.

METHOD

Subjects

Subjects consisted of junior and senior students enrolled in the academic track program at Bountiful High School, Bountiful, Utah. One group of subjects meeting at the first school hour (7:45–8:40 a.m.) represented the experimental (treatment) group and a second group of students meeting at the sixth hour (1:15–2:10 p.m.) constituted the control group. The treatment group consisted of eleven students (9 males and 2 females). The control group consisted of seventeen students (14 males and 3 females). Students deciding to take physics were divided into two groups. Students with superior abilities in science and mathematics were encouraged to enroll in an honors section of physics which met during the third period of the school day. The remainder of the students enrolled in a physics section meeting during the first or sixth period depending upon scheduling necessities. Some "honors-capable" students also enrolled in the first and sixth period classes because scheduling problems made it impossible for them to be in the honors section. However, the distribution of such "honors-capable" students was approximately equal between the two classes. Finally, at the conclusion of the first semester, late in January 1985, a few students in both classes dropped physics, and a small number of students switched sections. It should be noted that these changes occurred prior to the start of the experimental intervention, and represented typical end-of-the-semester behavior for such classes.

Pre-instructional Subject Ability Measures

A number of subject ability measures were collected prior to the instructional intervention. With these measures the experimental and control groups could be equated for pre-existing differences in mental ability through a multiple regression procedure (to be described later). The subject ability measures were of two types: proxy and real [26].

Proxy ability measures – Proxy measures are pre-experimental treatment measures on student ability that exist, but are not implemented or induced by the experimenter. Such data are in archival form and are available in the permanent student records. The proxy ability measures used in the present study were:

1. Previous physics grades: raw scores and letter grades for the first term, second term, and end of the first semester in physics.
2. Previous high school science grades: biology, chemistry, and ninth grade science grades.
3. Previous high school mathematics grades: geometry and all algebra course grades.
4. Stanford Achievement Test (SAT) scores: subscores for mathematics, science, and reading from the SAT in percentile form.
5. American College Testing Program (ACT) scores: subscores for English, social science, natural science, and mathematics from the ACT in percentile form.

Real ability measures – Real measures are measures that did not exist prior to the experimental intervention, but are collected by the experimenter for the purpose of equating the control and treatment groups. Nine factor-referenced cognitive ability tests [27] were administered as real ability measures during the week prior to the initiation of the treatment period. These tests were chosen from six of twenty-three identified cognitive factors [27] with no more than two marker tests used per cognitive factor. Aptitude factors and, therefore, marker tests were chosen so as to tap an important and broad selection of skills necessary for success in a high school physics course. Of the test selection used, reliabilities ranged from approximately 0.70 to approximately 0.90 on those tests reporting reliabilities (see Appendix B for a description of the individual tests and their reliabilities). The reference groups ranged from ninth, eleventh, and twelfth grade high school students, through college students and armed service enlistees.

In order to accommodate the entire sequence of tests within a one week period, the tests were shortened and scored on a straight percentage basis. Since the use of the tests was to provide a base of comparison between treatment and control groups, it was not necessary to adhere to recommended testing practices; all that was required was that the testing procedure be identical for the treatment and control groups.

Design

The experiment involved both the experimental and control groups receiving eleven weeks of instruction in physics. The experimental group received its instruction via the remote audio-graphic teleconferencing system (see Appendix A for a description of the technical system used), while the control group received its instruction via traditional live-teacher-in-the-classroom instruction. Two teachers served as "experimenters" in the project. Each teacher taught three sections of physics to both the treatment and control groups. The same teacher presented the same concepts to both groups during the same time frame (with the exception of the difference in meeting hour for the two groups). This procedure eliminated any possible interaction of teachers with content areas and treatment conditions. For instance, if teacher A excelled at teaching mechanics and teacher B excelled at teaching electricity, having teacher A teach the experimental group mechanics and teacher B teach the experimental group electricity would result in the experimental group having an unfair advantage over a control group which had teacher A teach electricity and teacher B teach mechanics, because these subjects are not these teachers' greatest strengths. The current experimental design controlled for this possibility.

Additionally the two teachers alternated in presenting units of material. If teacher A presented unit I to both groups, then teacher B would present unit II to both groups, and so on. This arrangement minimized any cumulative effects due to one teacher being more facile with one of the two presentation modes (remote teaching with the technical system versus traditional teaching). The actual units of physics material presented, dates presented, and the teacher responsible for the presentation were as follows: 1) Phases of Matter, 2/25/85–3/5/85, Mr. Rudolph; 2) Heat Measurements: Part I, 3/6/85–3/18/85, Mr. Ford; 3) Heat Measurements: Part II, 3/19/85–3/28/85, Mr. Rudolph; 4) Heat Measurements: Part III, 3/29/85–4/15/85, Mr. Ford; 5) Heat Engines, 4/16/85–5/1/85, Mr Rudolph; and 6) Waves, 5/2/85–5/10/85, Mr. Ford.

At the end of each of the six units listed above students took a test to assess the effectiveness of their instruction. Each test involved both multiple choice and computational (problem solving) sections. In all of the tests but one the multiple choice section contained thirty questions. The test that did not have thirty multiple choice questions came from a somewhat shorter section of physics material, and therefore had fewer test questions (10). The computational section of each test was designed to query the mathematical problem solving abilities of the students. These test scores served as the dependent variables in the experiment.

Procedure

The procedure for the experimental group went as follows: hardware was activated five to ten minutes before the start of the class. Lighting was adjusted (e.g., rear window shades had to be drawn to eliminate glare on the TV monitors at the front of the class), programs were loaded, and the sound system was quickly activated and tested. Students were positioned near the front of the class in a traditional seating arrangement with no row of students greater than four deep. After roll was taken by the management person, class began. Typically, less than three minutes was required to effect system set-up.

Lesson material was presented in a variety of ways. The remote teacher developed the subject content live in a lecture-type setting, or downloaded the lesson material from a previously constructed lesson file stored on a five-and-one-quarter-inch floppy disk in the computer. Students responded to the material by watching one of two TV monitors in the front of the classroom that displayed the lesson material as if it were written on a chalkboard in front of the classroom. At the same time, students listened to the remote teacher via an audio link that consisted of a pair of speaker/microphones located in the front of the classroom. Since audio communication was fully bi-directional, students were always able to respond to teacher queries or interrupt and ask questions themselves. At appropriate times students were asked singly or in pairs to go to one of the two electronic digitizing tablets, also in front of the classroom just in front of the television monitors, and write in requested information. This information was seen by everyone in the classroom, as well as by the remote teacher.

Lesson presentation proceeded in a conventional fashion, driven by the requirements of the immediate teaching strategy. For example, when the class was involved in a problem solving session, two pairs of students were often directed to the electronic writing tablets to show their work, while the remaining students worked on the problems at their seats or were directed to display their work to the class at an available "real" blackboard. At different times students were asked to take a short quiz that was displayed on the monitor. In this context grading was handled by a paper exchange (the traditional method in this school for dealing with homework and short quizzes). After the graded papers were returned, the students reported their grades orally to the teacher via the audio link.

The control group received similar lesson material. However, it was presented in the traditional manner with direct lecturing and direct use of the available blackboard space.

RESULTS

The null hypothesis in this study was: the experimental and control groups do not differ with respect to their overall score on physics tests given during the instructional intervention period. To test the question of interest, a hierarchical multiple regression was used. Proxy and real ability measures were entered at the first step to equate the experimental and control groups for non-equivalence due to non-random assignment of individuals [26, 28]. At the second step a binary coding variable representing membership in either the treatment or control group was entered to test if the groups differed as a function of their differing treatments.

Dependent Variables

The results from the six unit-end tests were combined in different ways to answer different questions. There were two teachers (Mr. Rudolph [teacher A] and Mr. Ford [teacher B]) and two types of questions (multiple choice and computational). The crossing of these two factors led to four dependent variables per student: 1) the total number of multiple choice questions answered correctly for teacher A; 2) the total number of multiple choice questions answered correctly for teacher B; 3) the total number of points for computational questions for teacher A; and 4) the total number of points for computational questions for teacher B. The correlations among these four variables are displayed in Table 1. As is apparent from the table, these dependent measures are all strongly related (all p's < 0.001, $N = 28$).

Additionally, the data can be collapsed over teachers yielding an overall multiple choice score and an overall computational score, or collapsed over types of test question yielding a total score for teacher A and a total score for teacher B. These measures are also strongly related. The correlation between teachers was 0.84 ($p < 0.001$, $N = 28$), indicating the two teachers were essentially equally effective in instructing students (as demonstrated by students' performance on

Table 1. Intercorrelations Among Physics Test Scores as a Function of Teacher and Question Type

Physics Test Score	*1*	*2*	*3*	*4*
1. Teacher A,[a] Multiple Choice	–	0.84	0.82	0.72
2. Teacher A, Computational		–	0.78	0.76
3. Teacher B,[b] Multiple Choice			–	0.87
4. Teacher B, Computational				–

Note: Physics test scores are total scores summed over tests within a category. The correlation is over all subjects, experimental and control groups combined, with $N = 28$. All r's significant at the $p < 0.001$ level.

[a] Teacher A was Mr. Rudolph.

[b] Teacher B was Mr. Ford.

physics tests given). The correlation between the two types of test question was 0.90 ($p < 0.001$, $N = 28$), showing that both multiple choice and computational questions were measuring the same thing—physics knowledge.

Finally, data could be collapsed over both teachers and types of test question, yielding a single test score per student. Because of the high intercorrelations among the various dependent measures described above, this was done. This dependent variable—overall physics performance—was used in the multiple regression analysis central to our study.

Proxy and Real Ability Measures

The large number of proxy and real ability measures collected in this study presented a problem: capitalization upon chance. With such a large number of independent measures it was likely that some of the measures would explain variation in the dependent variable merely because of spurious correlations with the dependent variable. To guard against this eventuality, a number of techniques were used to reduce the real and proxy measures to some smaller, more manageable set of measures.

Previous physics grades – To reduce the number of scores per student from six to a smaller, more manageable set, a principal components factor analysis was performed. The six scores that served as input into the principal components analysis were: 1) raw scores for the first term in physics; 2) raw scores for the second term; 3) raw scores for the first semester (terms one and two combined);

4) grade point average for the first term (i.e., A = 4.00; A- = 3.67; B+ = 3.33; etc.); 5) grade point average for the second term; and 6) grade point average for the first semester. These scores were highly intercorrelated (lowest $r = 0.89$, $p < 0.001$, $N = 28$), in part because they were not independent.

The principal components analysis revealed a single principal component that accounted for 96 percent of the total variance in the six physics grade measures described above. All six measures loaded 0.97 or higher on this factor. Only this factor was retained, and a factor score was calculated for each student on it. This factor score served as the independent measure of previous physics performance in our subsequent regression analysis.

Previous high school science grades – First and second semesters of ninth grade science, biology, and chemistry (but not physics) were converted to a numerical grade point scale (i.e., A = 4.00; A- = 3.67; B+ = 3.33; etc.) and then averaged. Since each student differed with respect to which science classes he or she had previously taken, a second measure was also calculated for each student. This measure was the number of science classes a student had taken. This was done to avoid giving undo credit to students who had "padded" their grade point average by taking only a few of the easiest science classes.

In reviewing these two measures, however, a problem was noted. Twelve of the twenty-eight students in the study had not taken any science courses while enrolled in high school. This rendered the average grade point measure useless, due to the large proportion of missing data. We decided to drop the average grade point measure for science from further consideration. Therefore, only the number of science classes taken was used in the subsequent regression analysis.

Previous high school mathematics grades – Previous high school mathematics grades were summarized in a manner analogous to that used for science grades. First and second semester grades for all algebra courses and geometry were converted to grade points and then averaged for each student. Additionally, the number of previous mathematics courses taken by each student was calculated. Missing data were not a problem for the mathematics average grade point measure; all students had taken at least one previous class in mathematics. Both the number of mathematics classes taken and the average grade point in previous math classes were used in the subsequent regression analysis to equate the two groups for pre-existing ability differences.

Stanford Achievement Test (SAT) and American College Testing Program (ACT) scores – These were subscores for reading, science, and mathematics from the SAT and for English, social science, natural science, and mathematics from the ACT. These scores were factor analyzed using a principal axis common factor analysis. Because of the fact that some of the students had taken the SAT tests, but not the ACT tests, while others had taken the ACT tests but not the SAT tests, all individuals who had nonmissing data for each pair of tests were included in the calculation of the intercorrelation matrix used in the factor analysis (i.e., pairwise deletion of missing data was used). Two factors, accounting for 79.3 percent of the common variance in the original test score data, were retained and rotated to a varimax criterion. The first factor appeared to be a general scholastic ability factor. All of the standardized test subscores except SAT science loaded on this factor 0.59 or higher. The second factor appeared to be a science ability factor. The SAT science subscore loaded on this factor 0.98, while the ACT natural science subscore loaded on this factor 0.55. Also, the SAT reading subscore contributed substantially to the science factor, loading on it 0.56.

A factor score for each of these two factors (general scholastic ability and science ability) was calculated for each student, and these two factor scores served as summary measures of the students' performance on the SAT and ACT tests. These factor scores were used in the first step of the hierarchical regression analysis to equate experimental and control groups for pre-existing differences in mental ability.

Factor-referenced cognitive ability tests – The real ability measures used in this study were nine factor-referenced cognitive ability tests [27]. The nine tests (followed by the factor they served as a marker for) were: 1) figural classification (induction), 2) following directions (integrative processes), 3) mathematics aptitude test (general reasoning), 4) necessary arithmetic operations test (general reasoning), 5) diagramming relationships (logical reasoning), 6) inference test (logical reasoning), 7) card rotations test (spatial orientation factor), 8) paper folding test (visualization), and 9) surface development test (visualization). Appendix B presents a more complete description of these tests.

Although these nine tests were found to mark six distinct factors [27], it was possible that within our sample of twenty-eight students the tests clustered in other ways. This might allow us to represent the nine factor-referenced cognitive ability tests by some set of factors smaller than six. To determine if the tests could be represented by some smaller number of factors, the nine tests were subjected to a principal axis common factor analysis. Factors with eigenvalues greater than one were retained and rotated to a varimax criterion. Two factors, accounting for 61.4 percent of the variance in the original set of nine tests, emerged from the factor analysis.

The first factor appeared to be a procedural reasoning factor. The tests loading highly on it were: 1) following directions (loading 0.79), 2) diagramming relationships (0.73), 3) mathematics aptitude test (0.67), 4) necessary arithmetic operations test (0.59), and 5) surface development test (0.52). These tests all involved keeping track of a set of serial processes (e.g., a reasoning procedure).

The second factor was a spatial ability factor. The tests loading highly on it were: 1) paper folding test (loading 0.77), 2) card rotations test (0.68), 3) surface development test (0.59), and 4) figural classification (0.55). Common to these tests is the need to manipulate information spatially or to encode spatial relationships.

Factor scores were calculated for each student for each of these two factors (procedural reasoning and spatial ability). These factor scores served as independent variables in the first step of the hierarchical regression analysis described below.

Hierarchical Multiple Regression

The hierarchical multiple regression was a two-step procedure. At the first step proxy and real ability measures were entered to control for pre-existing differences between the experimental and control groups due to nonrandom assignment of individuals to treatment conditions [26, 28]. There were six proxy ability measures: 1) previous physics grade factor scores, 2) number of previous science courses taken, 3) number of previous mathematics classes taken, 4) average grade in previous mathematics classes, 5) general scholastic ability factor scores (from the SAT and ACT subtests), and 6) science ability factor scores (from the SAT and ACT subtests). In addition, there were two real ability measures: the procedural reasoning factor scores and the spatial ability factor scores from the nine tests taken from the *Kit of factor-referenced cognitive tests* [27].

The science ability factor was determined to be redundant with other proxy and real ability measures (i.e., was a linear combination of them), and was dropped from further consideration. Thus, seven variables were entered at step one of the hierarchical multiple regression to equate the experimental and control group for pre-existing differences in mental ability.

The seven proxy and real ability measures accounted for 87.5 percent of the variance in overall physics scores during the instructional intervention. This multiple R^2 was significant ($F(7,20) = 20.06$, $p < 0.001$), and was further investigated by studying the significance of the individual unstandardized regression coefficients (B weights) for each of the proxy and real ability measures. These weights are presented in Table 2.

The statistical significance of the unstandardized regression coefficients reflects whether or not an individual proxy or real ability measure was predicting the dependent variable (i.e., overall physics scores during the instructional intervention) when the effects of all the other proxy and real ability measures were held constant. Table 2 shows that only procedural reasoning, spatial ability, and previous physics grades made unique contributions to the prediction of current physics performance. This finding should not be taken to mean that the other real and proxy ability measures do not significantly predict current physics performance; rather they may overlap and thus not make any unique contribution to prediction.

The second step of the hierarchical regression involved adding a binary coding variable to the regression equation. This coding variable differentiated the

Table 2. Unstandardized Regression Coefficients for Proxy and Real Ability Measures at Step One of the Hierarchical Multiple Regression

Ability Measure	*B Weight*	*Standard Error of B*	*F Ratio*[a]
Average Math Grade	12.108	14.359	0.711
Number of Math Courses	−2.859	3.441	0.690
Number of Science Courses	1.390	2.906	0.229
Previous Physics Grades	32.536	12.333	6.960*
Procedural Reasoning Factor	23.551	5.611	17.559**
Spatial Ability Factor	11.522	5.344	4.648*
General Scholastic Ability Factor	−0.016	0.013	1.427

[a] Degrees of freedom for all F ratios are (1,20).
* $p < 0.05$
** $p < 0.001$

experimental and the control groups. To discover if remote site instruction via audio-graphic teleconferencing differed from traditional classroom instruction, the significance of the binary coding variable was tested. Because this test occurred after all proxy and real ability measures had been entered into the regression equation, it also controlled for any differences in subject ability between the treatment groups. The unstandardized regression coefficient for the binary coding variable was not significant ($F(1,19) = 1.218$, $p > 0.25$), indicating that we cannot reject the null hypothesis of no difference between the two treatment groups.

DISCUSSION

This study was an attempt to assess the effectiveness of a remote site instructional system. The particular system tested was developed by the first author [11-13] and utilized audio-graphic teleconferencing. This technology is a member of the group known as "narrowband technologies"—methods that use only that portion of the transmission spectrum that can be transmitted over telephone lines. Because of the simpler hardware, maintenance, and manpower costs involved, narrowband technologies tend to be more cost-effective than their wideband (i.e., television) counterparts, and therefore a good choice for widespread use in the public school systems.

Our review of the literature found: 1) few of the remote site instructional systems reported in the literature had been evaluated as to their effectiveness in delivering instruction, and 2) those that had tended to use highly motivated post-secondary school groups. The present study dealt with secondary school students at a public high school who were enrolled in a traditional physics course.

The intervention took place over an eleven week period. During that time students in the experimental group received physics instruction remotely from two instructors via the audio-graphic teleconferencing system. Students in a control group received live-teacher-in-the-classroom physics instruction from the same two instructors during the intervention period. After using real and proxy ability measures to control for pre-existing differences between the experimental and control groups through a hierarchical regression procedure, no significant differences between the two groups could be found. Our interpretation of this finding is that remote instruction via the audio-graphic teleconferencing system is no worse than traditional classroom instruction, at least over the time frame used in the present study.

One might object that the large degree of statistical control used to equate the two groups obscured any relationship that might have existed between the treatment effect (experimental versus control groups) and overall current physics performance. Prevailing sentiment strongly supports the notion that getting physics over computer screens and audio links is not as good as having the teacher live right in front of students in the class. Reasonable as this sounds, our data do not support such contentions. The simple correlation between the binary coding variable (which differentiated the experimental from the control group) and overall physics scores was not significant ($r = 0.13$, $F(1,26) = 0.47$, $p < 0.10$). Thus no matter whether statistical control for non-equivalence of groups is used (as was done in the Results section) or is not used (as in the test of the simple correlation between the binary coding variable and current physics performance), the result is the same. The audio-graphic teleconferencing system is not demonstrably better or worse than traditional classroom teaching as a method of delivering instruction. This conclusion concurs with the findings of others in the continuing education field [24, 25].

We began by noting that technology was one way of dealing with increasing learner demands (e.g., providing a quality education for learners in remote locales) and shrinking educational budgets. The results of this study supported our initial optimism. Despite long held opinions about the value of face-to-face contact, evaluation research is demonstrating that there is also value in other less traditional forms of contact mediated through electronic telecommunications technology.

APPENDIX A
Description of the Technical System

SOUND SYSTEM

Establishing remote telephone contact was handled in a conventional dial-up fashion. At the teaching site (the site where the teacher is located) the sound system consisted of a small Radio Shack audio amplifier connected to a nearby telephone. The telephone signals were fed to the amplifier after telephone contact was established. Once the amplifier was turned on the telephone was hung up. A small lapel microphone and an eight-inch full range eight ohm speaker were connected to the audio amplifier. The lapel microphone allowed the teacher some mobility within the teaching space.

At the remote site (the site where the students received the instruction), the sound system consisted of a model TBA-20A Bogen Paging/Talkback 20 watt amplifier which was connected directly to the telephone wall jack once telephone contact was established. Connected to the amplifier were a pair of combination microphone/speaker setups, each sitting on a stool near the front right and left corners of the classroom. Sound transmission was handled by the equipment in a voice-activated-transmit (VOX) mode, in which students merely spoke out loud to communicate with the remote teacher. Conversation was fully bi-directional limited only by the very minor delay associated with a VOX circuit.

GRAPHICS SYSTEM

A second dial-up telephone line was used to establish graphic communication between participating sites. A single graphics installation consisted of the following components:

1. An IBM PC with the following internal configuration:
   a. Two double sided, double density (DS/DD) disk drives, 256 kilobyte (kB) random access memory (RAM), and keyboard;
   b. Hayes 1200B Smartmodem (internal);
   c. Amdek MIA high-resolution color graphics adapter;
   d. Serial adapter for the digitizing tablet.
2. A television monitor (25-inch CM-2501 Mitsubishi color monitor at the remote site and a 13-inch Amdek 710 color monitor at the teaching site).
3. A digitizing tablet (GTCO DIGI-PAD 5), electronic stylus, and power supply.
4. A power strip.
5. The software system, provided by Wasatch Security Research of Salt Lake City, Utah, that managed the functional components of the graphics system: communications, "blackboard" operation, and authoring.

A pair of such systems were placed side by side at the teaching site, and another pair of systems sitting on 40-inch projection carts were placed at the front corners of the remote instructional site. The digitizing tablets and keyboards were placed on oversized, square, tablet-arm chairs directly in front of the projection carts. Students using the digitizing tablets were able to easily see what was written on the screen. Students were seated in a traditional manner in five rows across the classroom with no row deeper than four students. Moreover, the frontmost seats on the extreme left and right of the classroom were removed to make room for the tablet-arm seats with the digitizing tablets and microcomputer keyboards.

FUNCTIONAL COMPONENTS

Communications. The computer was converted into a terminal emulator which permitted the computer to act as a telephone. Dialing, answering, maintaining contact, and hanging up were handled with approaches that used Hayes-compatible communication protocols.

"Blackboard" operation. Once communications were established between the remote and teaching sites, the digitizing tablet served as an intelligent blackboard. Students at the remote location or the instructor at the teaching location could use the stylus to write on the tablet as one would use chalk to write on a traditional blackboard. At its minimum operational level, the graphics system served as a classroom blackboard linking the students to the remote instructor. The blackboard system had four panels ("pages") of graphics information that could be stored in the available computer buffers. These could be displayed one after another, in the same way that a traditional blackboard with four panels could be displayed by sliding one panel up, and revealing a new underlying panel covered with additional material.

The user of the graphics system had additional facilities, made available through program features on the computer. These facilities were: 1) line, 2) circle, 3) rectangle, 4) area fill, 5) erase (partial as well as a complete screen wipe), 6) file download (for transferring files stored on the disk into the communications network), and 7) an active keyboard for simple word processing.

Authoring capability. Finally, the classroom instructor was able to use the graphics system to pre-design screen pages to fit the needs of an upcoming lesson, and store those pages permanently on a floppy disk. In this way the instructor was able to construct complete libraries of lessons to be used at appropriate times during the class. Lessons were created under a pre-named diskfile in a special LESSONS directory. The instructor could then bring to the screen on an available blank "page" any of the pre-stored lesson screens merely by going to the file downloading facility and using the stylus to select the required file under the name given to it at its creation.

APPENDIX B
Factor-Referenced Tests

This appendix describes the six cognitive factors tested in our study, the particular marker tests used, and the reliability of these marker tests. The tests, and the factor structure delineated by these tests, is the result of research work by Ekstrom, French, Harman, with Dermen [27].

INDUCTION

The induction factor identifies the kinds of reasoning abilities involved in forming and trying out hypotheses that will fit a set of data. The marker test used to measure the induction factor was figural classification.

Figural classification. Each item presents two or three groups containing geometric figures that are alike in accordance with some rule. The second row of each item contains eight test figures. The task is to discover the rules and assign each test figure to one of the groups. Reliability estimates for figural classifications are: 1) 0.77 for approximately 300 suburban eleventh and twelfth grade males; 2) 0.74 for approximately 300 suburban eleventh and twelfth grade females; and 3) 0.88 for 119 ninth grade males.

INTEGRATIVE PROCESSES

The integrative processes factor involves the ability to keep in mind simultaneously or to combine several conditions, premises, or rules in order to produce a correct response. The marker test used to measure the integrative processes factor was following directions.

Following directions. The subject is asked to determine the point in a matrix of letters that would be reached by following a complex set of directions. The reliability estimate for following directions is 0.76 based on a sample of approximately 560 Naval recruits.

GENERAL REASONING

The general reasoning factor measures the ability to select and organize relevant information for the solution of a problem. The marker tests used to assess the general reasoning factor were the mathematics aptitude test and necessary mathematical operations.

Mathematics aptitude test. The subject is asked to solve five problems requiring simple algebraic concepts. The reliability estimate for the mathematics aptitude test is 0.81 based on a sample of eighty-three Army enlistees.

Necessary mathematical operations. The task is to determine which numerical operations are required to solve a problem without actually carrying out the manipulations. The reliability estimates for necessary mathematical operations are: 1) 0.73 for 145 college males, and 2) 0.71 for 119 ninth grade males.

LOGICAL REASONING

The logical reasoning factor involves the ability to reason logically from premise to conclusion, or to evaluate the correctness of a conclusion. The marker tests used to measure the logical reasoning factor were diagramming relationships and the inference test.

Diagramming relationships. The task is to select one of five diagrams which best illustrates the relationships among sets of three objects. The reliability estimate for diagramming relationships is 0.79 based on a sample of approximately 560 Naval recruits.

Inference test. The task is to select the correct conclusion which can be drawn from a given statement. Reliability estimates for the inference test are: 1) 0.76 for approximately 300 suburban eleventh and twelfth grade males; 2) 0.78 for approximately 300 suburban eleventh and twelfth grade females; 3) 0.57 for 145 college males; and 4) 0.76 for 189 high school males.

SPATIAL ORIENTATION

The spatial ability factor involves the ability to perceive spatial patterns or to maintain orientation with respect to objects in space. The marker test for the spatial orientation factor was the card rotations test.

Card rotations test. Items are cards cut into irregular shapes. To each card's right are six other drawings of the same card, sometimes rotated and sometimes turned on the other side. The subject indicates whether or not the card has been turned over. Reliability estimates for the card rotations test are: 1) 0.86 for approximately 300 suburban eleventh and twelfth grade males; 2) 0.89 for approximately 300 suburban eleventh and twelfth grade females; 3) 0.80 for forty-six college students; and 4) 0.83 for ninety-nine college females.

VISUALIZATION

The visualization factor measures the ability to manipulate or transform the image of spatial patterns into other arrangements. The marker tests used to assess visualization were the paper folding test and the surface development test.

Paper folding test. For each item successive drawings illustrate two or three folds made in a square sheet of paper. The final drawing shows where a hole is punched into it. The task is to select one of five drawings to show how the punched sheet would appear when fully reopened. Reliability estimates for the paper folding test are: 1) 0.75 for approximately 300 suburban eleventh and twelfth grade males; 2) 0.77 for approximately 300 suburban eleventh and twelfth grade females; 3) 0.84 for forty-six college students; and 4) 0.84 for eighty-two Army enlistees.

Surface development test. Drawings are presented of solid forms that could be made with poster board or sheet metal. With each drawing there is a diagram showing how a piece of paper might be cut and folded so as to make the solid form. Dotted lines show where the paper was folded. One part of the diagram is marked to correspond to a marked surface in the drawing. The task is to indicate which lettered edges in the drawing correspond to numbered edges or dotted lines in the diagram. Reliability estimates for the surface development test are: 1) 0.90 for forty-six college students, and 2) 0.92 for eighty-six Army enlistees.

REFERENCES

1. W. Kirby, The Need to Consider Alternative Delivery, in *Regional Forum on Distance Learning: A Report* (Contract No. 400-83-0007), B. O'Connor and C. Y. Levinson (eds.), Southwest Educational Development Laboratory, Austin, Texas, (ERIC Document Reproduction Service No. ED 247 910), pp. 4-5, 1984.
2. The National Commission on Excellence in Education, *A Nation At Risk: The Imperative for Educational Reform*, U.S. Government Printing Office, Washington, D.C., 1983.
3. E. L. Boyer, *High School: A Report on Secondary Education in America*, Harper and Row, New York, 1983.
4. L. Darling-Hammond, *Beyond the Commission Reports: The Coming Crisis in Teaching*, (Rand Publication Series Report No. R-3177-RC). The Rand Corporation, Santa Monica, California, July, 1984.

5. New York State Education Department, *Educational Telecommunications in Small Rural Schools: State of New York Pilot Program for Assessment and Implementation of Educational Telecommunications Applications as an Alternative to Strengthen Small Rural Schools*, Author, Albany, New York, 1984.
6. B. Ellis, Educational Teleconferencing, *Educational Media International, 1*, pp. 27-28, 1985.
7. H. E. Hudson and C. H. Boyd, *Distance Learning: A Review for Educators*, (Contract No. 400-83-0007), Southwest Educational Development Laboratory, Austin, Texas, (ERIC Document Reproduction Service No. ED 246 872), 1984.
8. F. C. McCormick and E. R. McCormick, Effective Utilization of Microcomputers in Rural and Small Schools, paper presented at the joint meeting of the Fourth Annual Rural and Small Schools Conference and the Kansas Community Educational Association Conference, Manhattan, Kansas, November, 1982.
9. B. O'Connor and C. Y. Levinson (eds.), *Regional Forum on Distance Education: A Report*, (Contract No. 400-83-0007), Southwest Educational Development Laboratory, Austin, Texas, (ERIC Document Reproduction Service No. ED 247 910), 1984.
10. J. A. Repo, *Computer Conference Information Service: Research Report 191*, (Report No. ISBN-951-38-1799-7), Technical Research Centre of Finland, Espoo, Finland, 1983.
11. S. Rudolph, *Remote Site Instruction Project at Bountiful High School: A Proposal*, Utah State Office of Education, School Community Development Section, Salt Lake City, Utah, 1985-86.
12. ———, *Remote Site Instruction Project at Bountiful High School: An Interim Report*, Utah State Office of Education, School Community Development Section, Salt Lake City, Utah, 1985-86.
13. S. Rudolph and M. K. Gardner, *Remote Site Instruction Project at Bountiful High School: A Final Report*, Utah State Office of Education, School Community Development Section, Salt Lake City, Utah, 1985-86.
14. E. Duchon and R. Fortune, Mathematics and Language Drill by Dial Access, in *Regional Forum on Distance Learning: A Report*, (Contract No. 400-83-0007), B. O'Connor and C. Y. Levinson (eds.), Southwest Educational Development Laboratory, Austin, Texas (ERIC Document Reproduction Service No. ED 247 910), pp. 20-21, 1984.
15. Y. A. Ewell, Multiplying the Math Teacher: An Electronic Aid to Help Meet Shortage, *Educational Technology*, pp. 36-37, February, 1983.
16. J. Gerlovich and R. Unruh, Science Teacher Inservice by Teleconference, in *Regional Forum on Distance Learning: A Report*, (Contract No. 400-83-0007), B. O'Connor and C. Y. Levinson (eds.), Southwest Educational Development Laboratory, Austin, Texas (ERIC Document Reproduction Service No. ED 247 910), pp. 15-17, 1984.
17. R. L. Hartz, Two-way Telecommunications: A Viable Technology for Rural Instruction, paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada, April, 1983.

18. J. Kruh, College Credit for the Gifted by Teleconference, in *Regional Forum on Distance Learning: A Report*, (Contract No. 400-83-0007), B. O'Connor and C. Y. Levinson (eds.), Southwest Educational Development Laboratory, Austin, Texas (ERIC Document Reproduction Service No. ED 247 910), pp. 13-15, 1984.
19. P. A. Monteau, Teleconferencing: What It Is, How It Works, and What You Need, *Instructional Innovator*, pp. 20-21, September 1984.
20. S. Rudolph, A New Technology for Physics Instruction, *Journal of College Science Teaching*, *15*, pp. 123-126, 1985.
21. Teaching by Telephone and Telewriter, *The Physics Teacher*, p. 251, May, 1968.
22. F. Todd, B. S. Scott, and E. Gillette, Elementary Math and Foreign Languages by Gemini Blackboard, in *Regional Forum on Distance Learning: A Report*, (Contract No. 400-83-0007), B. O'Connor and C. Y. Levinson (eds.)., Southwest Educational Development Laboratory, Austin, Texas, (ERIC Document Reproduction Service No. ED 247 910), pp. 18-20, 1984.
23. L. Yetter, How to Use the Gemini Blackboard, in *Regional Forum on Distance Learning: A Report*, (Contract No. 400-83-0007), B. O'Connor and C. Y. Levinson (eds.), Southwest Educational Development Laboratory, Austin, Texas, (ERIC Document Reproduction Service No. ED 247 910), pp. 26-27, 1984.
24. A. F. Rushton, Jr. and R. K. Branson, Remote Delivery of Instruction in Complex Skills, *Journal of Educational Technology Systems*, *11*, pp. 315-323, 1982-83.
25. D. E. Weingand, Telecommunications Delivery of Instruction: A Comparison with the Traditional Classroom, *Journal of Education for Library and Information Science*, *25*, pp. 3-13, 1984.
26. T. J. Cook and D. T. Campbell, *Quasi-experimentation: Design and Analysis Issues for Field Setting*, Rand McNally College Publishing, Chicago, Illinois, 1979.
27. R. B. Ekstrom, J. W. French, H. H. Harman, with D. Dermen, *Kit of Factor-referenced Cognitive Tests, 1976*, Educational Testing Service, Princeton, New Jersey, 1976.
28. J. Cohen and P. Cohen, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, Rand McNally College Publishing, Chicago, Illinois, 1975.

Direct reprint requests to:

Michael K. Gardner
Department of Educational Psychology
The University of Utah
Salt Lake City, UT 84112