UMTA-MA-06-0126-84-4
DOT-TSC-UMTA-84-1

U.S. Department of Transportation
Urban Mass Transportation Administration

The CCIS Experiment: Comparing Transit Information Retrieval Modes at the Southern California Rapid Transit District

Robert O. Phillips

Wilson Hill Associates, Inc.
140 Federal Street
Boston MA 02110

March 1984
Final Report

UMTA Technical Assistance Program

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UMTA-MA-06-0126-84-4		

4. Title and Subtitle	5. Report Date
THE CCIS EXPERIMENT: COMPARING TRANSIT INFORMATION RETRIEVAL MODES AT THE SOUTHERN CALIFORNIA RAPID TRANSIT DISTRICT	March 1984
	6. Performing Organization Code: DTS-64
7. Author(s): Robert O. Phillips	8. Performing Organization Report No.: DOT-TSC-UMTA-84-1
9. Performing Organization Name and Address: Wilson Hill Associates, Inc.* 140 Federal Street Boston, MA 02110	10. Work Unit No. (TRAIS): UM464/R4653
	11. Contract or Grant No.: DTRS-57-81-00054
12. Sponsoring Agency Name and Address: U.S. Department of Transportation Urban Mass Transportation Administration Office of Technical Assistance Washington, DC 20590	13. Type of Report and Period Covered: Final Report November 1980 - December 1981
	14. Sponsoring Agency Code: URT-42

15. Supplementary Notes

*Under contract to: U.S. Department of Transportation
Research and Special Programs Administration
Transportation Systems Center
Cambridge, MA 02142

16. Abstract

This report documents the results of a controlled experiment performed in the Telephone Information Section of the Marketing Department at the Southern California Rapid Transit District (SCRTD) in Los Angeles. The Telephone Information Section is the site of a prototype deployment of automated transit information systems (ATIS) technology known as the Computerized Customer Information System (CCIS). Through CCIS, the SCRTD hopes to: (1) reduce the cost of providing telephone information; (2) increase telephone information productivity and (3) improve the level of service provided by the information agents.

A controlled experiment was designed and implemented to measure differences in productivity and response accuracy between information agents using automated and manual retrieval methods. In the course of the experiment, a predetermined set of 36 itinerary-type transit information questions was asked to each of nine test agents.

Analysis of the results showed that statistically significant differences in performance were discernible among agents of varying skill/experience levels and between retrieval modes. It was concluded that CCIS is indeed a viable alternative to manual information retrieval methods and has potential for further application in transit telephone information operations.

17. Key Words	18. Distribution Statement
Automated Transit Information Systems Analysis of Variance Latin Square Experimental Designs	DOCUMENT IS AVAILABLE TO THE PUBLIC THROUGH THE NATIONAL TECHNICAL INFORMATION SERVICE, SPRINGFIELD, VIRGINIA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	112	

PREFACE

This report has been prepared by Wilson Hill Associates, Inc. for the U.S. Department of Transportation, Transportation Systems Center (TSC) in support of the Urban Mass Transportation Administration's Impact Assessment Program sponsored by the Office of Methods and Support. UMTA is cooperating with and supporting the transit industry's efforts to improve the productivity of telephone information/marketing services to the public. The principal focus of this support is on the use of the computer to improve transit telephone information services, under what UMTA calls its Automated Transit Information Systems (ATIS) program. TSC is currently evaluating ATIS deployments in Los Angeles and Washington, DC. This report is an interim product in this ongoing ATIS impact assessment program. Its intended audience is the transit information community and those interested in the application of computer-aided information retrieval systems.

The experiment described in this report was proposed by and conducted under the management of Robert Furniss, Wilson Hill's project manager, who also prepared the first draft. The experimental design and much of the statistical analysis was provided by Dr. Chester H. McCall, a consultant to Wilson Hill. Comments on the first draft were received from I. Michael Wolfe of TSC, John Durham of UMTA and Dr. McCall. The final version of this report was prepared by Robert Phillips of Wilson Hill.

Much gratitude is expressed to the Southern California Rapid Transit District personnel involved with the information system development, especially Mrs. Paddie Brennen, Supervising Systems Analyst, for their cooperation. Special thanks also go to Dr. Arthur S. Priver, the TSC Technical Monitor, and Mr. I. Michael Wolfe for their encouragement and managerial support.

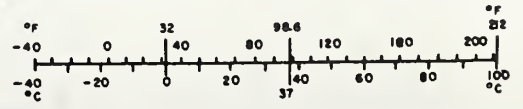
METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



CONTENTS

CONTENTS (CONTINUED)

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (CONTINUED)

EXECUTIVE SUMMARY

This report presents descriptions and analyses of an experiment involving the testing of transit information agents using different information retrieval modes at the Southern California Rapid Transit District (SCRTD). The purpose of this experiment was to provide greater insight into the applicability of automated transit information systems (ATIS) technology currently being implemented in a pilot program at SCRTD.

The ATIS being implemented at SCRTD is known as the Computerized Customer Information System (CCIS). Its applicability is currently limited to those bus routes wholly or partially serving that portion of the SCRTD service area which lies within the San Fernando Valley. A limited number of other frequently-requested landmarks served by these routes, including UCLA, LA International Airport, and "downtown LA", are also included in the system.

The CCIS is used by transit information agents in responding to queries from the public regarding transit service in the San Fernando Valley area. The system is activated by agent input from a computer terminal keyboard located at the agent's work position; system output information is displayed on a cathode ray tube (CRT) located above the agent's keyboard.

For transit information agents the use of ATIS technology represents a radical departure from the standard operating procedure of manually referencing printed material. It requires development on the part of the agent of certain new job skills (e.g., typing). Automation of information agent data retrieval is intended to improve speed and productivity, and also to increase information accuracy, completeness, and consistency.

The experiment documented in this report was designed to measure variations in speed and accuracy between transit information agents using the CCIS and agents using traditional

data retrieval methods. The objective of the experiment was to obtain comparisons (between the various methods) that were not biased by a host of factors present under normal agent working conditions such as call variability, agent experience variability, caller comprehension problems, etc.

In the experiment, nine agents were subjected to a rigorously controlled test. Each agent was queried over the phone with the same set of thirty-six test questions concerning SCRTD transit service provided within the San Fernando Valley. These nine agents represented three different levels of skill and experience (novice, intermediate and advanced) and used three different modes of information retrieval (manual, CCIS, or a mixture of both). The testing was conducted by two members of the contractor evaluation team.

The responses provided by each agent were analyzed in terms of both response speed and quality. In these areas, results were broken down by mode of data retrieval and agent skill/experience level. Where possible, comparisons of individual agent performance were also made. Among the more salient conclusions to be drawn from these analyses are the following:

- Agents using the CCIS mode were fastest overall in providing responses; they were approximately 13% more productive in terms of the number of calls processed per hour. The superior productivity of the CCIS agents as a group was reduced considerably by the performance of the novice CCIS agent, who produced the next to lowest productivity rating of all agents tested. If novices are excluded from the comparison, the superior productivity of CCIS becomes much more pronounced; the experienced CCIS agents processed slightly over 20 calls per hour while the average call processing rates of the experienced manual and mixed mode agents were

approximately 16 calls per hour, a difference of approximately 25%.

- Agents using the mixed mode of data retrieval, in which manual, CCIS, or combinations of the two data retrieval modes, achieved the highest overall ratings in terms of response quality (judged on both accuracy and completeness criteria). Of all the responses provided by this group, 83% were rated as "useful", 12% were rated as "marginal", and only 5% were rated as "unsatisfactory." The manual and CCIS groups were very evenly ranked, with the manual slightly ahead (two percentage points) in all grading categories. Once again, however, the relatively low response quality scores of the novice CCIS agent tend to diminish the overall performance of the CCIS group. If novice data are eliminated, then the CCIS mode moves ahead of the manual mode to second place in the overall qualitative rankings, with scores of 83%, 8% and 8% in the "useful", "marginal", and "unsatisfactory" categories, respectively. These scores are still, however, somewhat lower than those for the mixed mode.
- Average agent performance in terms of both response speed and quality falls below standards (such as 20 calls/hour) established by the SCRTD. Possible explanations of substandard agent performance in these areas are unavoidable experimental effects which hampered agent performance (such as a nervous feeling their job performance was being monitored) as well as the relative difficulty of the test questions as compared with those generally posed by SCRTD customers in normal information center operations. The relative performances of the agents, however, were not invalidated by these conditions.

- As might be expected, the more experienced (intermediate and advanced) agents received higher scores than the novice agents in terms of both response speed and quality in most instances. Differences between intermediate and advanced agents, however, are much less pronounced; in fact, intermediate agents outscored advanced agents in several comparison categories.
- The most important conclusion to be drawn from the experiment is that the CCIS has been shown under rigorously controlled conditions to be indeed a viable alternative to manual information retrieval methods. Under conditions imposed by the experiment, CCIS agents performed quite comparably or better than manual agents in terms of response quality, and were superior in terms of response speed.

Based on these experimental results, it can be stated that the CCIS demonstrates strong potential for assisting information agents in retrieving accurate transit data in a timely manner. The fulfillment of this potential will depend on SCRTD's commitment to provision of necessary system support. This support must include training with respect to any new agent skills required for efficient CCIS operation and any necessary fine tuning of system operating features accomplished through cooperation between the agents and SCRTD management.

1. INTRODUCTION

This report presents descriptions and analyses of an experiment involving the testing of transit telephone information agents at the Southern California Rapid Transit District (SCRTD).

The Urban Mass Transportation Administration (UMTA), in keeping with its commitment to increase public acceptance of transit, has been cooperating with selected properties in the research and development of new tools and methods for improving their public information/marketing services. In this regard, UMTA has funded the implementation and testing of Automated Transit Information Systems (ATIS) technology at SCRTD in Los Angeles and at the Washington, DC Metropolitan Area Transit Authority (WMATA).

ATIS technology is intended to help transit agencies provide accurate, timely and consistent telephone information to callers. With an ATIS, a transit information operator (referred to hereafter as an "agent") receives telephone queries from callers and retrieves the desired information through the use of a remote computer terminal located at his/her work position. The system is activated when query data is entered on the terminal keyboard by the agent. The query is then processed in the computer. This processing involves comparing alternative responses stored in the computer's transit and geographic data bases (containing information on routes, schedules, fares, street addresses, landmarks, etc.) and selection of the most appropriate response. Following processing, which in this particular ATIS deployment is specified to take seven seconds or less, response options are displayed on a cathode ray tube (CRT) located above the agent's keyboard. These responses are then reviewed by the agent and the one judged most appropriate is supplied to the caller.

ATIS operation is quite different from traditional manual methods of data retrieval. Using these traditional methods, an agent typically receives a query from the caller, abbreviates it on a scratch pad, and then consults an indexed set of volumes located at his/her work position. Information contained in these volumes includes printed schedules, headway sheets, and maps of bus routes and landmarks located in the service area. The agent analyzes the relevant information contained in these volumes, selects the appropriate response, and provides it to the caller.

The ATIS being tested by the SCRTD is known as the Computerized Customer Information System (CCIS). The CCIS has been implemented as a pilot project and is currently limited to providing information concerning SCRTD bus routes which operate wholly or partially in the San Fernando Valley. The CCIS services eight peripheral UTS-400 CRT terminals connected to the SCRTD's main UNIVAC 1106 computer. The general objectives of the ATIS-CCIS deployment are to:

- Improve the quality of information provided to the callers in terms of accuracy, consistency, and completeness.
- Reduce costs of the information department to the transit property by reducing training/proficiency times of new agents, and eliminating or minimizing tasks associated with updating of manual reference materials.
- Increase the quantity of information processed by the telephone information marketing department, in terms of both number of calls answered per hour and total information provided per call.

Other longer term objectives associated with a system-wide ATIS deployment are to:

- Provide benefits to other departments in the transit authority (such as those performing scheduling, route planning, and other similar functions) through common data base utilization.
- Provide spin-off benefits to third parties, if possible (e.g., remote terminals located in employment centers, airports, and other transportation nodes).

The experiment documented in this report represents one element of an overall program to assess the extent to which the CCIS is achieving the above objectives. The overall purpose of the experiment was to provide controlled conditions under which qualitative and quantitative differences in the performance of agents using the manual and CCIS modes of data retrieval could be measured.

In Section 2.0, the retionale underlying the design of this controlled experiment is presented. Section 3.0 contains analyses of call time (agent speed) results. Section 4.0 contains analyses of agent response quality (accuracy and completeness) results. Section 5.0 contains other selected analyses and observations relevant to the experiment. In Section 6.0 major findings are summarized. Finally, Appendices A-F present detailed experimental data and calculations.

2. RATIONALE FOR THE EXPERIMENT

This experiment was undertaken to gain additional insight into the differences between the performances of transit information agents using the CCIS and manual modes of data retrieval. Differences between the two groups were measured in quantitative (agent response speed) and in qualitative (agent response accuracy and completeness) terms. The experiment was designed to perform these measurements under controlled conditions, free from external influences or circumstances that might tend to cloud conclusions regarding agent performance under normal operating conditions. When the task of comparing the two modes of data retrieval was attempted during normal operations, it was apparent that a number of uncontrollable outside variables hindered valid modal comparisons. Among these variables were:

- Differences in agent abilities and experience levels (agent proficiency generally increases with job experience);
- Variable agent usage of CCIS;
- The variety of caller queries in terms of both type and content;
- Differing comprehension levels among the callers; and
- Operational factors such as computer malfunctions, shift changes, and agent break periods.

Controlled testing according to an experimental design was recommended by the contractor to overcome these difficulties and provide accurate measures of agent performance for each mode of data retrieval. In this experimental situation, a number of steps were taken to ensure that the agent's environment did not become so artificial as to not represent normal working conditions. To this end a number of important variables were controlled within careful limits. For example:

- A standard questionnaire was developed and presented to each agent. In order to reduce the number of test agents and tests required to produce valid experimental results, this questionnaire was divided into four sets of nine questions each.
- Agents selected for the experiment had call productivities near the average for their skill/experience levels.
- Comparisons were made between agents of relatively equal experience and skill.
- Agents were given at least one practice query in order to get accustomed to the experimental situation.
- The experiment was conducted during normal working hours.
- Agents were instructed to perform as they would normally, and not to consider the experiment as a test of their abilities.
- The experimental setting was designed to be as similar as possible to the agent's normal work station.
- Other potential bias-introducing variables, such as the order of presentation of test queries, were rigorously controlled.

Detailed descriptions of these aspects of the experiment will be presented in the remaining sections of this chapter. It is important to keep in mind that this experiment was not intended to be the sole evaluation tool with which to judge the effectiveness of CCIS. The number of test agents used was small, and the possibility remains that the agents did not perform "normally" under the experiment. However, the experiment was conducted because merely observing agents at their work stations would not have been particularly productive in terms of examining the potential of the CCIS. Its purpose was to provide a controlled setting in which some useful modal comparisons could be made. This effort would not have been successful without the full support and commitment of the SCRTD.

2.1 EXPERIMENTAL DESIGN

The experiment involved the individual testing of nine SCRTD transit information agents. Each agent was queried over the phone with the same set of 36 test questions concerning transit service provided in the CCIS service area. The nine agents represented three modes of information retrieval and three levels of agent skill and experience. The experiment was conducted by two members of the contractor evaluation team; evaluators alternated in performing the roles of caller and observer. As the caller made test calls from a separate room and recorded agent responses on a log form, the observer, located in the same room as the test agent but out of direct view, logged total call times using a stop watch.

2.1.1 Test Environment

The test environment was an SCRTD agent training room made available for the test. The room was furnished with a standard agent work position equipped with both manual and CCIS information retrieval systems. Test agents used a standard telephone to answer calls instead of the headset used in normal operations. This standard phone was required in order to permit direct dialing to the agent. Use of the regular SCRTD information center telephone system would have placed the test call in a queue with assignment to the first available agent, not necessarily the test agent as desired. Although this represented a departure from the norm, none of the agents encountered any problem with the change in equipment. No personnel were allowed in the testing room other than the test agent and observer.

2.1.2 Modes of Information Retrieval

Three agents each were tested using one of the three following modes of information retrieval:

- Manual (M):
  For this mode, each test agent was permitted to use only the manual (printed) reference material provided by the SCRTD. This material is contained in a set of large binders and covers all information on routes, schedules, and fares for the SCRTD system.
- CCIS (C):
  In this mode, each test agent used only the CCIS for query processing and information retrieval. No manual reference verification was permitted. In essence, the CCIS mode agents were captives of the speed and accuracy of the CCIS. During the hours in which testing took place, the SCRTD attempted to minimize other time-shared loading of the main-frame computer supporting the CCIS; employee time sheet and payroll processing, which slows CCIS response time considerably, were deferred during these times.
- Mixed (X):
  This mode gave each test agent the option of using either the manual mode or the CCIS mode, or both. There was no requirement for minimum usage of either mode. Mixed mode agents could, therefore, verify manually any responses obtained through CCIS transactions or vice versa.

In order to simulate normal call processing procedures as nearly as possible, the following guidelines applied to all three test modes:

- All test agents were allowed the option of answering queries without information referencing, if they so chose. Experienced agents frequently respond to familiar queries "off the top of their heads" in this manner. Assuming the agent's memory is accurate, this is by far the quickest method of

providing a query response, since there is no data retrieval time involved.

- All test agents were permitted to use their street map books (Thomas Guides); agents draw bus routes in these Thomas Guides and refer to them frequently to locate a caller's origin or destination or to provide walking instructions.
- All test agents were permitted to use paper and pencil to record the caller query; agents usually abbreviate query data on scratch pads.
- All test agents were instructed to perform as they would normally at their work stations. They were told that both accuracy and speed of their responses were being measured, with no bias placed on one over the other. They were reassured that test results had nothing whatsoever to do with their job status, and that their test results would be strictly anonymous to the graders. It was stressed that the mode of data retrieval, and not the particular agent, was the real subject of the test.

Thus, the major differences between test agents were those of experience/skill level and those created by the implementation of CCIS, i.e. typing and transaction function knowledge, as well as the ability to scan CCIS response screens to select the "best" alternative. These differences were intended to be used as the explanatory variables underlying differences in agent performance in the experiment.

2.1.3 Test Agents

Since one of the major objectives of the CCIS is to reduce overall training requirements for new agents, the experiment attempted to measure performance on the three modes by relative agent experience/skill level. The contractor team identified three major agent categories for testing in each mode:

- Novice:
  A "novice" agent was defined as a recently trained agent (less than six months experience) having an entry level call count (approximately 10 to 14 calls per hour.)
- Intermediate:
  An intermediate agent was defined as an agent with a minimum of one year's experience having a mid-level call count (approximately 15 to 18 calls per hour).
- Advanced:
  An advanced agent was defined as an agent with over two years experience having a high call count (20 or more calls per hour).

Agent qualification and selection for the experiment was performed by the Senior Supervisor of the SCRTD Telephone Information Section. Agents designated to operate the CCIS (CCIS mode and mixed mode) were also required to have a CCIS call count in the range applicable to their experience/skill category.

With the exception of the CCIS mode novice, all agents selected for the test had completed the standard SCRTD entry level agent training course which involved approximately eight weeks of instruction; this basic course was geared to manual information retrieval methods. The novice CCIS agent had not yet finished this course. He was selected for the experiment by the Senior Supervisor because it was her opinion that he most closely fit the above description in the CCIS mode. It should be noted that all test agents, including the CCIS novice, had completed the eight hour CCIS training course which involved both instruction and practice using the CCIS terminal keyboard.

2.1.4 Test Queries

Four question sets of nine queries each were formulated for the experiment. A list of all questions is contained in Appendix A. Question sets I and II (Questions 1-18) were

formulated by the contractor evaluation team; Sets III and IV (Questions 19-36) were formulated by the SCRTD Planning Department and further developed by the contractor. A total of 36 queries were chosen recognizing the SCRTD productivity standard of 20 calls per hour. Total test time was estimated to be two hours, including a short break between Question Sets I-II and III-IV. Four question sets were used because of the length of time necessary to complete the whole test. With the 9X9 Latin Squares (described below) used for each question set, only 9 test agents, and 9 repetitions of the test, were required, rather than 36 agents and repetitions.

Each of the questions in the test included at least one itinerary-type request ("How do you get from Point A to Point B?"). Such requests are the most common type of call encountered by agents in regular telephone information operations. Many of the itineraries requested were rather complex, involving transfers, etc. These itineraries tested not only the overall abilities of manual mode agents, but also the capabilities of the CCIS software and data base in determining the "best" routing between a given origin and destination, one of the major CCIS attributes being evaluated by the contractor. Besides the itinerary request, an average of just over two additional requests were made in each question. These results were for such information as fares, return schedules or itineraries. Some of the additional requests required a change in the original itinerary requested, i.e., a routing which might take longer but involve fewer transfers or a lower fare, etc.

The issue of requesting walking instructions in the queries was a sensitive one due to the fact that CCIS explicitly lacks this capacity. Many of the more experienced manual agents take pride in providing such personalized information to the caller. The contractor team, recognizing this point,

made the decision to request walking instructions in only a few of the 36 test queries, under the assumption that CCIS agents could use their Thomas Guides to provide this information. In any other case where agents volunteered such information, the caller stated that it was unnecessary, and the elapsed time for this exchange was deducted from total call time as recorded by the observer.

The SCRTD reviewed all test questions before the experiment to ensure that geographic points mentioned were indeed included in the CCIS data base, and that all requests were comparable to those handled in regular operations.

Each test query was typed on a five-by-seven-inch index card and numbered for use by the caller during the test. The 36 query cards were divided into four sets of nine cards, each set a different color to prevent error in call sequencing.

The order in which test questions were presented to the agents was controlled through the use of Latin Squares, a statistical technique for performing an analysis of variance among certain variables (this technique is described more fully in Section 3.1). Four such Latin Squares were used to indicate the query order within each of the four question sets. This procedure varied the order of query presentation for each test agent to minimize the impact of this order on test results.

2.2 PERFORMANCE OF THE EXPERIMENT

Actual agent testing took place during the week of September 22, 1980. Agents were scheduled for testing primarily based upon their availability during a standard work shift. No agents were tested after completing a work shift or during a day off. Agent scheduling for CCIS and mixed modes had to work around several breakdowns of the CCIS terminals and mainframe computer; manual agents were tested during these breakdown periods. Thus, computer breakdowns were not a factor in agent testing.

Most agents finished the test in one sitting, although several required testing on two consecutive days due to work shift time constraints. In these cases, two question sets were completed on each day. One agent, the intermediate manual agent, completed three sets during the first sitting; due to this agent's work absence the next day, his fourth question set was not completed until one week later.

Each agent was briefed prior to the test concerning the purpose of the experiment and the guidelines for his or her specific mode of operation as described above. At least one practice query was allowed each agent to permit familiarization with the experimental work position. Following each test, agents were requested not to discuss the nature of the queries with other agents scheduled for testing.

The combined data collection effort by the caller and the observer was successful and accurate. The caller recorded each agent's response on a separate log which identified the agent, the mode of retrieval and the query number. The observer in the test room measured total call times and wrote down qualitative observations concerning each agent's query processing procedures. The observer was seated in a location to avoid distracting the agent during the test. No communication was allowed between the agent and the observer unless the CCIS was unresponsive within its specified retrieval time (seven seconds). In these cases, the caller query was restated and an adjustment made in total call time. Thus, the computer performed up to specification for all CCIS agent testing.

3. ANALYSIS OF AGENT CALL TIMES

This section presents a quantitative analysis of test agent call times without regard to the accuracy or completeness of agent responses to the test questions. This analysis begins with a description of the statistical design used to compare call times. Next, call time results are analyzed for the entire test group; they are subsequently broken down by experience/skill level, mode of data retrieval, and individual test agent. Finally, a set of conclusions drawn from these analyses is presented.

3.1 STATISTICAL DESIGN

The major reason for conducting a controlled experiment such as the one described in this report is to isolate the effects of a particular set of variables of interest (the independent variables) from other extraneous variables that might tend to bias test results. This approach allows valid conclusions to be drawn concerning the impact of these independent variables upon a measure of performance. As noted in Section 2.0, in this experiment the independent variables of interest were the experience/skill level of each agent and the mode of data retrieval used. The measure of performance of interest in this section is the time from the beginning of the presentation of each query by the caller to the time when the agent completed his/her response to that query.

In the test, there were four sets of nine itinerary-type questions. To ensure that the order in which these questions were asked did not bias test results, i.e., that those questions perceived as "easy" by a group of agents did not all appear on the same question set, the test questions were presented in a different order unique to each agent. This unique sequence of presentation for each agent's questions was determined through a statistical design called a Latin Square.

The Latin Square design derives its name from an ancient Roman puzzle that dealt with the number of different ways a given number, say X, of letters could be arranged in a square matrix containing X^2 elements so that each of the X letters appears only once in each row or column of the matrix. For example, if X = 4, a Latin Square could be constructed as follows with the letters (elements) A, B, C and D:

A B C D
B C D A
C D A B
D A B C

Because each element appears only once in each row or column, the Latin Square is useful as a statistical design in applications where unique ordering of certain variables is considered important. The detailed structures of each of the four Latin Squares used for each of the four nine-element question sets are presented in Appendix B. This Latin Square design allowed an analysis of variance to be performed on agent call times. The analysis of variance is a statistical method of dividing patterns or distributions observed in experimental data into different parts, each part assignable to a known source or cause, called an "independent variable". It allows the evaluator to assess the relative magnitudes of importance to experimental results that may be attributed to the presence of each of these independent variables.

3.2 PRESENTATION AND ANALYSIS OF AGENT CALL TIME RESULTS

A review of agent call times, which are presented in detail in Appendix C, shows a considerable range between the shortest and longest responses. The shortest responses overall were posted on Question Set III (Questions 19-27) by the advanced CCIS agent (C3), who averaged 2 minutes and 16 seconds per call. The longest responses were given on Question Set I (Questions 1-9) by the novice mixed mode agent (X1), who averaged seven minutes and four seconds per call. The average for all test agents on all tests was

slightly over four minutes per call. If these call times are translated to calls per hour, the standard agent productivity measure of the SCRTD, they represent a range of 8.5 to 26.5 calls per hour, with an overall average of about 15 calls per hour. This average is only 75% of the SCRTD standard call rate for all agents of 20 calls per hour; possible reasons for this discrepancy will be discussed below.

Table 3-1 presents average agent call times for each test broken down by agent skill/experience level. As might be expected if one assumes job productivity increases with experience, the novice agents took much longer on average to complete their calls than did the more experienced agents. Call times for the novice agents as a group averaged over five minutes while call times for intermediate and advanced agents averaged approximately 3 1/2 minutes, a statistically significant difference. Call time differences between intermediate and advanced agents, however, were less dramatic, with the intermediate agents actually performing slightly faster (seven seconds on average) than the advanced agents over all of the Question Sets. Analysis of variance calculations shows that this difference in performance is not statistically significant. These results might, however, suggest that average call times decrease with increasing job experience only up to a point, beyond which call times remain relatively stable.

Table 3-2 presents average agent call times broken down by mode of data retrieval. The agents using the CCIS mode of data retrieval had the fastest overall call times, averaging approximately 3 minutes and 51 seconds per response. The manual mode agents ranked second overall with an average call time of 4 minutes and 3 seconds, while the mixed mode agents were third overall with an average call time of 4 minutes and 20 seconds. Statistically significant differences

TABLE 3-1. AVERAGE AGENT CALL TIMES
GROUPED BY SKILL/EXPERIENCE LEVEL
(IN MINUTES AND SECONDS)

AGENT EXPERIENCE/ SKILL LEVEL	QUESTION SET				AVERAGE ALL QUESTIONS
	I	II	III	IV	
NOVICE	6:02	5:50	4:14	4:52	5:15
INTERMEDIATE	3:48	3:32	2:58	3:27	3:27
ADVANCED	4:11	3:56	2:41	3:28	3:34
AVERAGE, ALL AGENTS	4:40	4:26	3:18	3:55	4:05

TABLE 3-2. AVERAGE AGENT CALL TIMES
GROUPED BY MODE OF DATA RETRIEVAL
(IN MINUTES AND SECONDS)

MODE OF DATA RETRIEVAL	QUESTION SET				AVERAGE ALL QUESTIONS
	I	II	III	IV	
MANUAL	5:01	4:08	3:28	3:35	4:03
CCIS	3:58	4:30	3:07	3:50	3:51
MIXED	5:02	4:40	3:18	4:21	4:20
AVERAGE, ALL AGENTS	4:40	4:26	3:18	3:55	4:05

between modes appeared only in Question Set I, where the CCIS agents performed much faster on average than the agents of the other two modes.

Tables 3-3 and 3-4 provide a comparison of individual agent performances. Table 3-3 presents the average productivity for each agent on each test. This productivity is determined by dividing an agent's average call time into one hour, producing the number of calls answered per hour on average by that agent. Table 3-4 presents the percentage obtained when an individual agent's call rate is divided by the average call rate for all agents on each test. If this percentage is over 100 for a given agent, then that agent's call productivity was better than the average for all agents. Conversely, a below-100 percentage indicates a below-average call rate. For example, the novice manual agent's (M1) call rate of 13.1 calls per hour was 85.3% of the average call rate of 15.1 calls per hour.

Those agents posting the fastest overall call times were the intermediate and advanced CCIS agents (C2 and C3), followed by the intermediate mixed and manual mode agents (X2 and M2). The advanced mixed and manual agents (X3 and M3) are next, followed by the novice agents in the manual, CCIS, and mixed modes, respectively. The slowest agents, by far, were the novice CCIS and the novice mixed mode agent, who used CCIS most of the time.

A number of detailed calculations leading to certain statistical tests are required for the analysis of variance. These statistical tests show whether a certain independent variable had a "significant" effect on the variable measured (in this case, agent call times). Independent variables tested for significance here were the order of query presentation, the mode of data retrieval, and the experience/skill level of the agents. Results of detailed analysis of variance,

TABLE 3-3. AVERAGE CALL RATES FOR EACH TEST AGENT
(IN CALLS PER HOUR)

AGENT SKILL/EXPERIENCE LEVEL	MODE OF DATA RETRIEVAL		
	MANUAL	CCIS	MIXED
NOVICE	13.1	11.1	10.4
INTERMEDIATE	16.3	19.1	17.1
ADVANCED	15.3	19.9	15.9

TABLE 3-4. PERCENT OF AVERAGE CALL RATE FOR EACH TEST AGENT

AGENT SKILL/EXPERIENCE LEVEL	MODE OF DATA RETRIEVAL		
	MANUAL	CCIS	MIXED
NOVICE	85.3	72.3	68.0
INTERMEDIATE	106.1	124.4	111.4
ADVANCED	99.6	129.6	103.5

or ANOVA, calculations which were performed on call time results are presented in Appendix D. These calculations show the following:

- Controlling the order of presentation of test questions was successful, as no reliable differences can be detected in average call times as a result of this variable.
- The one significant difference between the three modes of retrieval, as mentioned above, when results for each are averaged, occurred in Question Set I which contained the most difficult queries. Here the CCIS agents performed substantially faster than the manual and mixed mode agents. For the other three Question Sets, no statistically significant difference between modes was found. The markedly slower performance of the novice agents in their respective groups was a strong contributing factor to this overall result. When averaged, these slower performances tend to diminish the faster performances of the more experienced agents. This result was especially evident in the CCIS mode.
- Significant differences between agents of different skill/experience level were discernible in the CCIS and mixed mode agent groups for all four question sets as mentioned above. In these modes, the novice agent was substantially slower than the other two agents.
- The average call times of the manual agents exhibit much more variability than those of the other two modes. As a group, only the manual agents were significantly different from each other; this occurs in Question Set III, primarily due to the low variation in response times for this set.

Another variable of interest in comparing agent performance is the variability in the call times recorded. Table 3-5 presents the standard deviations (in seconds) of the call times for each agent over all 36 responses in the experim The lower the standard deviation shown for an agent in this table, the more consistent was that agent at providing respons at or near his or her average call rate. This table shows that the intermediate and advanced CCIS agents (C2 and C3) were the most consistent in their call times, while the novice manual and mixed mode agents (M1 and X2) were the least consistent.

3.3 FACTORS INFLUENCING AGENT CALL TIMES

A number of explanations might be advanced for the fact that seven of the nine test agents posted average call rates below the SCRTD standard of 20 calls per hour (agents C2 and C3 had rates between 19 and 20 calls). Among these are the following:

- Experimental effects.
- Difficulty of test questions.
- Agent inexperience.

These are discussed below:

Experimental effects are those factors influencing agent performance arising out of the experiment itself. Because of the telephone system configuration at SCRTD, it was impossible to test the agents at their normal work stations Instead, they were isolated in one of the training rooms for the experiment, so calls could be placed to them directly. Because of this isolation, agents participating in the experiment may have felt that for some reason they were being singled out for closer observation. Furthermore, because they were advised that call time and response accuracy were being measured, they may have felt, despite consultant

TABLE 3-5. STANDARD DEVIATIONS OF AGENT CALL TIMES (IN SECONDS)

AGENT SKILL/EXPERIENCE LEVEL	MODE OF DATA RETRIEVAL		
	MANUAL	CCIS	MIXED
NOVICE	110.4	103.1	135.6
INTERMEDIATE	84.5	52.3	70.6
ADVANCED	115.1	58.1	90.2

reassurances to the contrary, that the experiment was designed to give SCRTD management an in-depth view of their job performance. For these reasons, agents may well have performed in both more nervous and more cautious manner in their responses than they would ordinarily have at their own work stations. Both of these factors would mitigate towards longer responses.

Agents and SCRTD management both agreed that the queries selected for use in the experiment were generally more complex than average calls encountered by agents during the normal performance of their jobs. Many of the more experienced agents are able to respond to certain familiar and frequently-recurring queries without reference to any data whatsoever. These "off the top of the head" responses obviously take far less time than those for which either manual or computer data retrieval is required, and tend to increase substantially an agent's average call rate. Information Section performance data shows that such "easy" calls occur about 15% of the time. None of the queries in the experiment, however, were so "easy" that data retrieval was unnecessary. In fact, all of the test queries requested construction of a complete itinerary from point A to point B, and most involved further requests beyond this itinerary. As a result, the difficulty of the questions may well have been a factor in producing longer agent call times.

The novice agents ranked seventh, eighth, and ninth in overall call productivity. Their relatively slow performances tended to reduce the overall call rates for each mode. This was especially true for the CCIS mode, where the average call rate for the novice was a full eight calls per hour slower than that of the more advanced agents. Under these circumstances, perhaps a fairer comparison between modes could be achieved if data for the novice agents were alimi-nated from the experiment. Tables 3-6 and 3-7 illustrate

the results of such a comparison. The call time differences between the modes of retrieval are enhanced considerably with novice data eliminated; CCIS agents out-perform the other two modes by better than 30 seconds per call, or approximately 3 calls per hour. In Table 3-7, it can be seen that with novices eliminated, the CCIS agents are the only ones having higher than average call productivities over the entire experiment.

3.4 CONCLUSIONS ON AGENT CALL TIMES

Based on the results of the ANOVA calculations performed on experimental call time results presented above, the following conclusions may be drawn:

1. Average call times reported for each question set ranged from 2 minutes and 46 seconds per call to 7 minutes and 4 seconds per call, with an overall average for all agents on all tests of 4 minutes per call. Translated into calls per hour, this represents a range from 8.5 calls to 26.5 calls per hour, with an average of almost 15 calls per hour.
2. The overall average of 15 calls per hour is 75% of the SCRTD productivity standard of 20 calls per hour. Experimental effects, difficulty of questions, and novice agent inexperience might all be advanced as explanations for this discrepancy.
3. The novice agents in each group performed considerably slower than the intermediate and advanced agents. Call time differences between intermediate and advanced agents are less dramatic, with the intermediates actually performing slightly faster on average than the advanced agents. These results might suggest that call times decrease with

TABLE 3-6. AVERAGE AGENT CALL TIMES
GROUPED BY MODE OF DATA RETRIEVAL
NOVICE AGENT DATA EXCLUDED
(IN MINUTES AND SECONDS)

AGENT SKILL/EXPERIENCE LEVEL	MODE OF DATA RETRIEVAL		
	MANUAL	CCIS	MIXED
INTERMEDIATE	3:41	3:08	3:31
ADVANCED	3:55	3:01	3:46
AVERAGE, BOTH GROUPS	3:48	3:05	3:39

TABLE 3-7. PERCENT OF AVERAGE CALL RATE
NOVICE AGENT DATA EXCLUDED

AGENT SKILL/EXPERIENCE LEVEL	MODE OF DATA RETRIEVAL		
	MANUAL	CCIS	MIXED
INTERMEDIATE	94.4	110.6	99.0
ADVANCED	88.6	115.3	92.1

experience only up to a point, beyond which call times remain relatively stable. Significant differences on agent call times between various levels of agent skill/experience were found in ANOVA calculations for all four Question Sets, primarily as a result of the slow performances of the novice agents.

4. The CCIS mode of data retrieval was clearly demonstrated to be the quickest of the three, with an average call time approximately ten seconds less than the manual mode and thirty seconds less than the mixed mode. The slowness of the mixed mode is due to agents' cross-checking of CCIS responses against manual references. The fastest individual agents were the intermediate and advanced CCIS agents. The mixed mode intermediate agent, who relied heavily on CCIS, had the next highest productivity rate. If novice agent data are eliminated from comparisons among modes, the superiority of the CCIS over the other two modes is enhanced considerably. Significant differences on agent call times between modes were found in ANOVA calculations only for Question Set I, in which the CCIS agents performed much faster than the other two modes.
5. Examination of standard deviations of call times shows that the intermediate and advanced CCIS agents were the most consistent in their call times, while the novice manual and mixed mode agents exhibited the greatest variability.

4. ANALYSIS OF AGENT RESPONSE QUALITY

In this section, the quality of agent responses to the 36 test queries is presented and analyzed without regard to call times. This analysis begins with a description of the qualitative grading procedure and a summary of grading results. Grading results are analyzed first for the entire test group and then broken down by experience/skill level, by mode of data retrieval, and individual agent. Following these analyses, several conclusions regarding agent response quality are presented.

4.1 QUALITATIVE GRADING PROCEDURE

The experiment produced a total of 324 completed agent response logs, one for each question asked of each agent.* These logs included a section for post-test evaluation of agent response quality. Responses were graded independently by an impartial panel selected by SCRTD management which consisted of two members of the SCRTD Planning Department and the Training Supervisor from the Information Section. In the grading process, the panel used scoring criteria developed by the contractor team. Major emphasis in this process was placed on the accuracy and completeness of each response.

To judge response accuracy and completeness, each agent response was compared with a "best" itinerary or other response as determined by the individual grader. The determination of these "best" responses was made using SCRTD manual reference materials available on the date of the experiment. Since the CCIS data base relies on these same manual materials for updating, it was assumed that the manual and CCIS data bases contained comparable information on the date of the experiment. Once grading was completed, the three graders compared their results and discussed those particular responses

---

* See Appendix E for an example of the agent response log.

for which there were major differences in scores assigned. On all but two or three of these responses (there were 60 in all), the graders were able to reach a compromise grade acceptable to all. To provide the final results shown in Appendix F, results from the three graders are averaged.

The primary criterion used by the graders for judging the accuracy of each response focused on the question:

> "Using the response information provided by the agent, can the caller actually make the trip from Point A to Point B in view of query input?"

The answer to this question provided a simple and decisive standard for determining overall response quality. If the caller could indeed make the trip, the response was rated satisfactory (S). If the caller could not make the trip, the response was rated unsatisfactory (U).

Assessing the completeness of the information provided in each response focused on those itinerary instructions volunteered to the caller and on any omission of potentially useful details by the agent. Items emphasized included the following:

- bus route selection
- schedule information
- transfer information
- fare information
- walking instructions (if requested)

All those responses assigned satisfactory (S) ratings in the accuracy grading were evaluated for completeness using these criteria. They were assigned a numerical rating based on a progressive scale of one (least useful information provided) to ten (all necessary information provided). A response with an "S-10" rating was therefore quite comparable to, if not the same as, the "best" response as determined by the SCRTD graders.

Before the grading process began, agent identities and their modes of information retrieval were removed from response logs so as to eliminate any possible bias on grading results induced by these two factors. While this grading process relied on subjective judgements and consistency of the graders, it nevertheless provided the best available method to distinguish the relative quality of agent responses, since it relied on a consensus of "impartial" grader opinion.

4.2 PRESENTATION AND ANALYSIS OF AGENT RESPONSE GRADES

A review of agent response grades, which are presented in Appendix F, reveals that there was some variability to grades assigned, but that the preponderance of these grades fell within the S-6 to S-10 range. Responses in this range, which accounted for 79% of the total, might be termed "useful" to the caller. Responses assigned grades in the S-1 to S-5 range, which might be characterized as "marginal" to the caller, accounted for an additional 12% of the total. Unsatisfactory grades were assigned in only 28 cases, or 9% of the total.

From the frequencies of unsatisfactory and marginal grades assigned, as shown in Table 4-1, it can be seen that Question Sets I and II (Questions 1-18) accounted for 67% of all marginal responses and 75% of all unsatisfactory responses. Question set III clearly produced the most accurate responses from agents; over 90% of the responses to this set were judged "useful" (S-6 to S-10).

Comparison of agent grades assigned broken down by skill/experience level, shown in Table 4-2, demonstrates important differences between the three groups. As might be expected, the agents of the advanced group scored considerably higher than did the other two groups; 86% of all their responses were in the useful range. The intermediate agents were next best at providing accurate responses, with 79% of their

TABLE 4-1. FREQUENCY OF RESPONSE QUALITY GRADES RECEIVED BY ALL TEST AGENTS

RESPONSE QUALITY GRADE	QUESTION SET				TOTAL	PERCENT OF TOTAL
	I	II	III	IV		
S-10	6	6	18	8	38	12
S-9	12	19	26	25	82	25
S-8	22	8	15	20	65	20
S-7	10	13	11	13	47	15
S-6	9	9	2	4	24	7
S-5	7	7	3	3	20	6
S-4	5	3	2	2	12	4
S-3	2	0	0	3	5	2
S-2	1	0	0	0	1	0
S-1	1	0	0	0	1	0
U	5	16	4	3	28	9

TABLE 4-2. RELATIVE FREQUENCY OF
RESPONSE QUALITY GRADES RECEIVED
GROUPED BY AGENT SKILL/EXPERIENCE LEVEL
(% OF TOTAL FOR EACH GROUP)

RESPONSE QUALITY GRADE ASSIGNED	AGENT EXPERIENCE/SKILL LEVEL		
	NOVICE	INTER-MEDIATE	ADVANCED
USEFUL (S-6 to S-10)	73.1	78.7	86.1
MARGINAL (S-1 to S-5)	13.9	13.9	8.3
UNSATISFACTORY (U)	13.0	7.4	5.6

responses in the useful range. The novice agents, by contrast, had 73% of their responses fall in this range. Comparison of unsatisfactory grades received by each group reveals a reverse ranking; that is, the advanced agents received the fewest U's, and the novices received the most.

Comparison of agent response grades by mode of data retrieval also reveals interesting differences among the three modes. Table 4-3 shows the relative frequencies of useful, marginal and unsatisfactory grades grouped by mode of data retrieval. The agents in the mixed mode group, who were allowed to use either CCIS, manual, or a combination of both, were clearly rated the best overall at providing the most accurate and complete responses. The mixed mode group received the most useful (S-6 to S-10) ratings of the three groups, with 83% of all its responses receiving grades within this category. The mixed mode group also received the lowest overall unsatisfactory rating, with only 4.7% of its total in this category. Further examination of Table 4-3 shows that the manual and CCIS modes are very evenly ranked in terms of response quality, with the manual mode only slightly ahead (two or three percentage points) in all grading categories.

Table 4-4 provides a comparison of individual response quality ratings for each agent participating in the experiment. As can be seen in this table, the novice CCIS agent (C1) clearly received the poorest ratings of all the agents, with only 64% of his responses falling into the useful category. The manual and mixed novice and the manual intermediate (M1, X1 and M2) might be grouped next, with between 69% and 78% useful responses. The advanced manual, intermediate CCIS and mixed and advanced CCIS (M3, C2, X2 and C3) agents rank next, with 83% to 86% of their responses in this category. The advanced mixed mode agent (X3) clearly scored the highest in terms of response quality, with 89% useful responses and only one unsatisfactory response.

TABLE 4-3. RELATIVE FREQUENCY OF RESPONSE QUALITY GRADES RECEIVED GROUPED BY MODE OF DATA RETRIEVAL (% OF TOTAL FOR EACH GROUP)

RESPONSE QUALITY GRADE ASSIGNED	MODE OF DATA RETRIEVAL		
	MANUAL	CCIS	MIXED
USEFUL (S-6 to S-10)	77.8	76.9	83.3
MARGINAL (S-1 to S-5)	13.9	10.2	12.0
UNSATISFACTORY (U)	8.3	12.9	4.7

TABLE 4-4. FREQUENCY OF RESPONSE QUALITY GRADES RECEIVED BY EACH TEST AGENT (PERCENT OF TOTAL FOR EACH AGENT)

RESPONSE QUALITY GRADE ASSIGNED	MANUAL			CCIS			MIXED		
	NOV	INT	ADV	NOV	INT	ADV	NOV	INT	ADV
USEFUL (S-6 to S-10)	78	69	86	64	83	83	78	83	89
MARGINAL (S-1 to S-5)	14	19	8	14	8	8	14	14	8
UNSATIS-FACTORY (U)	8	12	6	22	8	8	8	3	3

It is interesting to note in Table 4-4 that there is no difference in the response quality scores of the intermediate and advanced CCIS agents. For the other two modes, however, there is an appreciable difference, especially in the number of marginal responses, between the intermediate and advanced agents. This result might reflect the fact that the "learning period" in which full proficiency is developed is much shorter for the CCIS than for the other two modes.

Table 4-5 presents the percentage obtained when the total number of useful (S-6 to S-10) responses compiled by each agent is divided by the average number of such responses for all nine agents tested. If this percentage is over 100 for a given agent, then that agent performed better than the average at providing high quality responses. Conversely, a below 100 score indicates a below-average performance. As can be seen in this table, all the novice agents and the intermediate manual agent (M1, C1, X1 and M2) scored below average while all the advanced agents and the intermediate CCIS and mixed mode (M3, C3, X3, C2 and X2) scored above average.

4.3 FACTORS INFLUENCING AGENT RESPONSE QUALITY SCORES

While the SCRTD maintains that its information agents give out erroneous information less than one percent of the time, the nine agents in the experiment received unsatisfactory grades on a total of 9% of all responses. This discrepancy might be explained by a number of alternative hypotheses. Among them are:

- experimental effects
- difficulty of questions asked
- lack of familiarity with CCIS

Each of these factors is discussed below.

TABLE 4-5. PERCENT OF USEFUL (S-6 to S-10) RESPONSES PROVIDED

AGENT EXPERIENCE/ SKILL LEVEL	MODE OF DATA RETRIEVAL		
	MANUAL	CCIS	MIXED
NOVICE	98.1	80.5	98.1
INTERMEDIATE	87.5	105.1	105.1
ADVANCED	108.6	105.1	112.1

Experimental effects are influences upon agent performance induced by the experiment itself. Because they were isolated from the rest of their co-workers and because they knew they were being observed, agents participating in the experiment may have felt that for some reason they were considered exceptional. Further, because they knew that their test performances were being measured, they may have feared (despite reassurances to the contrary by the consultant) that their response grades would be reviewed by SCRTD management and might somehow affect their employment status. As a result of these factors, agents may have been more nervous and prone to mental errors than they would ordinarily be in the normal performance of their job at their own work stations. Unfortunately, it was impossible to duplicate the experiment's controlled conditions at these work stations. Although observer logs show that experimental effects were felt by all the agents tested, to measure the overall impact of these effects on each agent's performance is not possible; some agents may well have felt more "pressure" than others.

The transit routing questions chosen for use in the experiment were generally conceded by agents and graders alike to be more difficult than the average caller requests handled by agents under normal working conditions. None of the questions was so "easy" that it could be answered without consulting some manual reference material or processing some computer transaction. On the job, experienced agents are frequently (about 15% of the time) able to respond to average caller queries "off the top of their heads". Each of the questions in the experiment required the agent to construct at least one complete itinerary, and many required construction of two or more. Some of these itineraries involved further complications such as multiple transfers, layovers, walking instructions, and fare or travel time computations. Because

the level of difficulty of the experimental questions was appreciably greater than usual, it is not unreasonable to expect a decline in response quality.

The novice computer mode agent (C1) received response quality scores appreciably lower than the other agents. It would seem reasonable in his case to hypothesize that he had not yet fully mastered the use of the terminal keyboard, the format screens, or the full range of computer transactions available through CCIS. As stated earlier, this agent had received the standard introduction to CCIS, but had not yet fully completed the eight week standard SCRTD agent training course, and had only just begun practice on "live" calls. His response quality scores are so low that they tend to diminish the overall showing of the CCIS agents as a group, particularly in light of the fact that the novice agents for the other two modes posted fairly respectable response quality scores; each of these agents posted five marginal and three unsatisfactory responses.

Perhaps a better comparison between the three modes could be achieved by eliminating novice agent scores from consideration, and using only intermediate and advanced agent scores. The results of such a comparison are shown in Table 4-6, which presents relative frequencies of grades received grouped by mode with novices excluded.

As can be seen by comparing this table with Table 4-3, elimination of novice agent scores makes no difference to the percentage of useful responses in the manual agent group. It does, however, raise the percentage of both the CCIS and mixed mode groups to a level where there is little difference between them, both appreciably ahead of the manual group. The mixed mode group, however, still ranks highest overall in terms of the fewest unsatisfactory responses. This result is somewhat to be expected in light of the considerably

TABLE 4-6. RELATIVE FREQUENCY OF RESPONSE QUALITY GRADES RECEIVED GROUPED BY DATA RETRIEVAL MODE NOVICE AGENT DATA EXCLUDED (% OF TOTAL FOR EACH GROUP)

RESPONSE QUALITY GRADE ASSIGNED	MODE OF DATA RETRIEVAL		
	MANUAL	CCIS	MIXED
USEFUL (S-6 to S-10)	77.8	83.3	86.1
MARGINAL (S-1 to S-5)	13.9	8.3	11.1
UNSATISFACTORY (U)	8.3	8.3	2.8

longer call times of the mixed group caused by double-checking of CCIS response screens against manual reference data.

Table 4-7 presents a comparison of individual agent performance with novice agents excluded from consideration. Table 4-7 contains information similar to that shown in Table 4-5, except that the average score used in percentage computations is the average for intermediate and advanced agents only. Once again, a percentage of over 100 in this table indicates an above-average performance, while a score below 100 indicates a below-average performance. As can be seen form this table, only the intermediate manual agent scored below average, while the rest of the test agents scored above average.

4.4 CONCLUSIONS ON AGENT RESPONSE QUALITY

Based on the response quality results presented above, the following conclusions may be drawn:

1. The quality of agent responses for the entire experiment exhibited a fair degree of variability, although most of them (79%) fell within the "useful" range (S-6 to S-10). An additional 12% of all responses were rated "marginal" (S-1 to S-5) by the graders, while the remaining 9% were assigned "unsatisfactory" ratings.
2. When response quality grades are grouped by skill/ experience level, results are generally as one would expect. The advanced agents scored higher than the other two groups, with 86% of their responses in the useful range. The intermediates were next in the rankings, with 79% of their responses rated useful. The novice agents, by contrast had 73% in this range. Comparison of unsatisfactory grades assigned to each group reveals a reverse ranking; the novice agents had the highest unsatisfactory percentage, the intermediates next, and the advanced agents had the lowest.

TABLE 4-7. PERCENTAGE OF USEFUL (S-6 to S-10) RESPONSES PROVIDED NOVICE AGENT DATA EXCLUDED

AGENT EXPERIENCE/ SKILL LEVEL	MODE OF DATA RETRIEVAL		
	MANUAL	CCIS	MIXED
INTERMEDIATE	84.3	101.1	101.1
ADVANCED	104.5	101.1	107.9

3. Comparison of response quality grades grouped by data retrieval mode shows that the mixed mode group, which was allowed to use either CCIS, manual, or both data retrieval modes scored the highest of the three groups, with 83% of its responses in the useful range and only 5% of its responses rated unsatisfactory. Behind the mixed mode group, the manual and CCIS groups are very evenly ranked, with the manual only slightly ahead (two percentage points) in all grading categories.
4. Analysis of individual agent performances indicates that manual verification of CCIS response screens by mixed mode agents was an important factor in overall response quality. The advanced mixed mode agent (X3) posted the best overall response quality scores, while the other mixed mode agents did at least as well or better than the other agents in their respective experience/skill level categories.
5. The response quality scores of the novice CCIS agent are so poor that his proficiency with the system is suspect. If he had not yet mastered CCIS, then including his scores in the analysis leads to unfair comparisons between modes. If all novices are eliminated from the experiment, then the CCIS mode moves ahead of the manual mode to second place in the overall qualitative rankings while the mixed mode remains in first place.
6. Even with the novice agent scores eliminated from consideration, the best scores posted by agents still remain several percentage points below SCRTD accuracy standards. Possible explanations for this showing are experimental effects and the difficulty of the test questions.

5. SELECTED ANALYSES AND OBSERVATIONS

This section presents some additional analytical and observational insights into experimental results. An attempt is made to integrate the speed and accuracy performance measures developed in the previous two sections.

5.1 OVERALL AGENT PERFORMANCE

An examination of overall agent rankings in terms of both speed and accuracy of responses reveals that the two rankings are quite different. In fact, it can be shown that the correlation, R^2, between the two rankings is not statistically different from zero. Given this disparity between the two rankings, the question arises as to whether there is any method to derive a ranking based on both speed and accuracy in order to ascertain the relative overall performance of each agent.

Several attempts were made to construct "composite" grading schemes in which scores based on speed and accuracy results were added or multiplied. It soon became obvious, however, that rankings based on such composite scores were artificial and could be changed easily by using different scoring assignment schemes for speed and accuracy results.

After composite grading proved to be of little value at measuring overall performance, a two-dimensional graphing method was devised in which both speed and accuracy results could be plotted in the same figure. While this method does not provide detailed rankings, it does permit some conclusions to be drawn concerning overall agent performance.

The graphing method plots the percent-of-average figures reported above in Tables 3-4 and 4-5, 3-7 and 4-7. The two axes of these graphs represent the average call time and the average useful (S-6 to S-10) response frequency for all agents; the intersection point of these axes thus represents the average of both speed and accuracy for the entire experiment. Plotting individual percent of average speed/accuracy

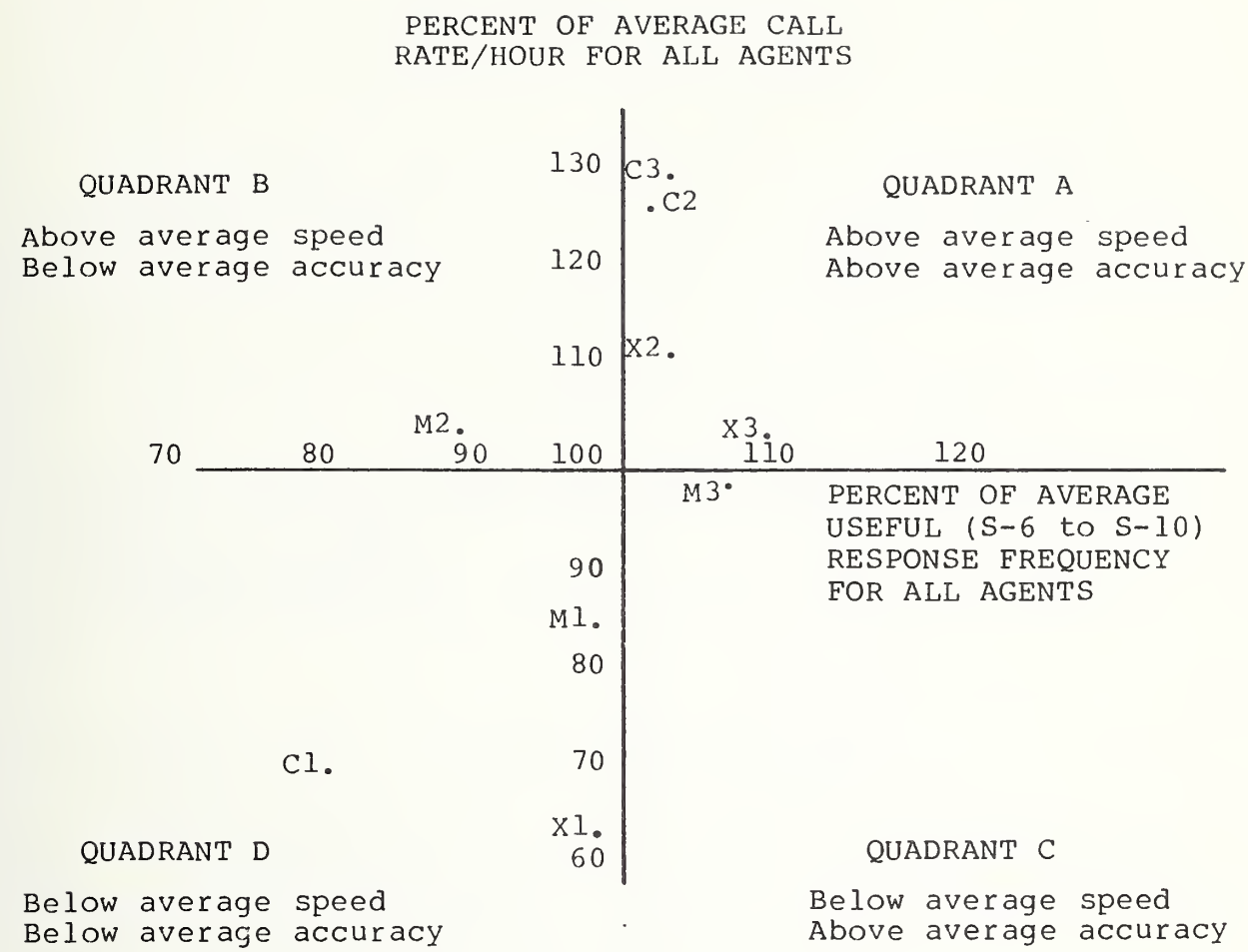
points for each agent thus provides a comparison between individual agent performance and the average performance of all agents.

Figures 5-1 and 5-2 illustrate the results obtained when two different sets of percent of average speed/accuracy data are plotted. In Figure 5-1, these data include all nine test agents; in Figure 5-2, novice agent data are excluded.

If the individual percent of average speed/accuracy point plotted for a given agent falls within the quadrant labeled "A" in these figures, then that agent's performance was above average in terms of both speed and accuracy. Conversely, if an agent's point falls within the quadrant labeled "D", it was below average in both respects. If an agent's point falls within those quadrants labeled "B" or "C", then that agent scored above average on speed but below average on accuracy or vice versa. The distance from the origin along each axis to an agent's particular speed-accuracy point provides a relative comparison of how much above or below average that agent performed in terms of both speed and accuracy.

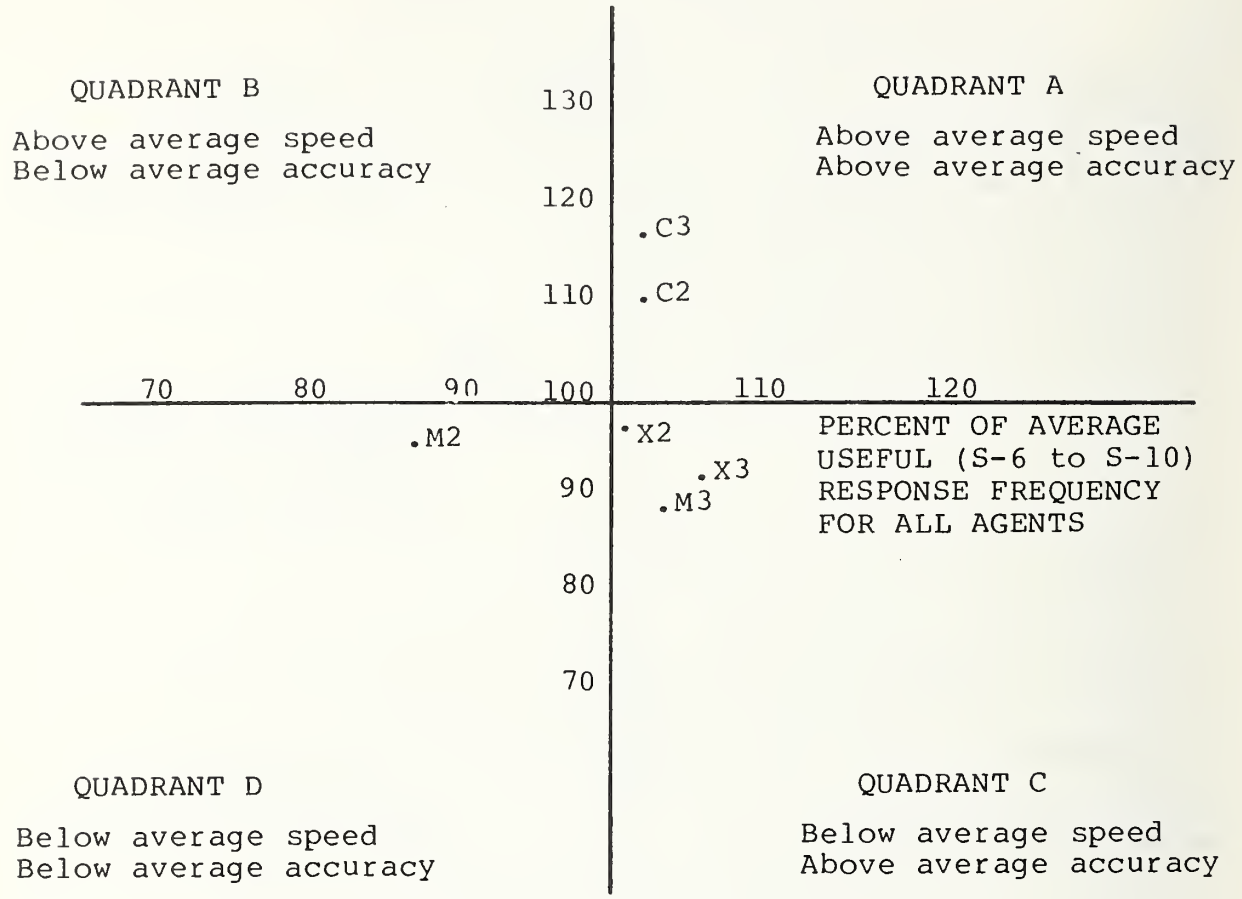
In Figure 5-1, it can be seen that the intermediate and advanced agents for both the CCIS and mixed modes all scored better than the overall average in terms of response speed as well as accuracy. The advanced manual agent (M3) just missed being included with these four agents; although his accuracy ratings were well above average, his call rate was just slightly (four-tenths of one percent) below average. The intermediate manual agent, on the other hand, performed well above average on speed, but his accuracy rating was the second lowest of the nine agents tested. Finally, the novice agents in all three modes scored below average in terms of both speed and accuracy.

With novice data eliminated, modal comparisons become somewhat clearer. In Figure 5-2, it can be seen that the CCIS mode agents (C2 and C3) are the only agents having



NOTATION: X1 = Mixed, Novice;
M2 - Manual, Intermediate;
C3 = CCIS, Advanced; etc.

FIGURE 5-1. PERCENT OF AVERAGE CALL RATE PER HOUR VS PERCENT OF AVERAGE USEFUL (S-6 to S-10) RESPONSE FREQUENCY FOR ALL AGENTS



NOTATION: X1 = Mixed, Novice;
M2 = Manual, Intermediate;
C3 = CCIS, Advanced; etc.

FIGURE 5-2. PERCENT OF AVERAGE CALL RATE PER HOUR VS PERCENT OF AVERAGE USEFUL (S-6 to S-10) RESPONSE FREQUENCY NOVICE AGENT DATA EXCLUDED

above-average call rates; the implication of this result is that their actual call rates were appreciably higher than those of the other agents. The intermediate and advanced CCIS agents, C2 and C3, also scored above average in terms of accuracy. Thus, when novice data are eliminated from consideration, the performances of the CCIS agents are clearly superior to those of the other two modes. The advanced mixed mode and manual agents (X3 and M3) ranked first and second overall in terms of response accuracy, but their call rates are below average for the six agents shown. The intermediate mixed and manual mode agents (X2 and M2) also scored below average in terms of response speed with novices excluded; agent X2 scored slightly above average in terms of accuracy, while agent M2 was below average in both respects.

5.2 PRODUCTIVITY BY MODE FOR USEFUL RESPONSES

The experimental data on speed and accuracy can be used to provide insight into two major issues affecting the transit telephone information industry, namely:

- How much information need be provided to the caller?
- What are reasonable hourly productivity rates for agents, assuming an inverse relationship between the number of calls processed per hour and overall response quality?

The experimental data on call times were reviewed to extract those call times for questions in which useful (S-6 to S-10) answers were provided. For purposes of analysis, the assumption was made that such responses represented the type of information which a caller deserves to receive when making itinerary-type queries as used in the experiment. The times of these useful calls were subsequently examined to highlight differences by mode. The statistical description

of these call times, shown in Table 5-1, provides an indication of how long it takes for agents using each of the three modes to provide a useful level of service to a prospective customer.

A comparison of the three modes indicates that the CCIS agents clearly demonstrated the lowest average call time (3 minutes and 30 seconds) for useful responses. The associated productivity for these responses is 17.1 calls per hour, or 86% of the SCRTD standard of 20. The lower standard deviation for the CCIS mode also reflects a more consistent capability for providing these useful responses in shorter call times. It is clear, however, that the relatively small number of useful responses from the novice CCIS agent contributed to the CCIS mode's overall consistency. Longer but more useful responses from the novice agents in the other two modes tended to reduce their respective group's overall consistency. The corresponding productivity measures for the manual and mixed modes were 15.6 and 14.5 per hour, respectively, or 78 and 73 percent of the SCRTD standard. This latter figure reflects the fact that many of the useful mixed mode responses involved agents first consulting CCIS and then seeking manual verification of the CCIS data, a process which took considerably longer than did retrieval for the other two modes.

The operational productivity implications of this analysis should be assessed in light of the differences between the experimental questions and the actual query mix encountered in daily operations. In all probability, normal queries are less difficult, with fewer complete itinerary requests. Agent productivity is therefore more likely to be higher. Of course, experimental effects influencing agent call times might also explain the substandard productivities for all three modes.

TABLE 5-1. MODAL COMPARISONS OF USEFUL (S-6 to S-10) CALL TIMES (IN MINUTES AND SECONDS)

MODE OF DATA RETRIEVAL	MEAN CALL TIME	STANDARD DEVIATION	PRODUCTIVITY IN CALLS PER HOUR
MANUAL	3:51	1:46	15.6
CCIS	3:30	1:30	17.1
MIXED	4:09	2:10	14.5

5.3 OBSERVATIONS

The preceding analyses do not reveal all the important data elements recorded in the experiment. In addition to the quantitative data recorded, qualitative notes were made by the test room observer to try to recognize the human variables involved in the experiment. A few of the more important observations made during testing are discussed below to provide some additional insight into agent behavior by mode.

The manual agents, in general, tended to offer more detailed information in an attempt to personalize the requested itinerary for the caller. Agent offers to provide walking instructions were quite common. Manual agents also tended to advise the caller on how long a transfer wait might take in light of bus schedules. Helpful details on the itineraries were also frequently provided. Agents identified intersections, landmarks, or bus sign designations (i.e., the destination wording after the bus number). Clearly, the manual agents take pride in providing this comprehensive type of service. Their overall ability highlights the importance of experience in the transit information agent's job.

The performance of the CCIS mode agents was quite varied based on experience level. As noted above, the novice agent had probably not mastered the terminal keyboard, the format screens, or the full range of transactions available through the CCIS. In spite of this lack of intimacy with the system, he did manage to slowly and methodically process most of the queries. The two more experienced agents operated the CCIS much more effectively. It was clear that these agents felt some frustration in not being allowed to perform a manual verification during the system processing hiatus before a CCIS response. On several queries, the advanced agent appeared

suspicious of the accuracy of the response screen. This reaction emphasizes the tendency of more experienced agents to read and evaluate the CCIS response on the CRT before providing the information shown to the caller.

The mixed mode agents using CCIS also spent time reading and evaluating the response screens. Manual reference verification of CCIS responses was quite common for the mixed mode group. This dual referencing gave the impression that the mixed mode agents were working harder to give the same responses during the test than the manual and CCIS agents. In several cases, manual verification led mixed mode agents to process a second CCIS transaction using manual references as input data. As can be seen from their response quality scores, this procedure clearly improved the overall quality of the group's responses, while at the same time slowing their response speeds considerably.

One important qualitative variable across all modes was the manner and sequence of providing the response information to the caller. Manual agents usually initiated the information process earlier in the call but took a longer period for the total response. Agents using CCIS seemed to have a longer initial information retrieval period (data entry) followed by a more concise sequence of bus trip data. Agent responses to the same question involving information on bus stops, bus routes, schedules, and transfers were sequenced in several different ways. The test callers were probably more sensitive to this variation than would the day-to-day caller be as a result of recording response information on call after call.

6. MAJOR FINDINGS AND IMPLICATIONS

This Section presents the major findings of the experiment and discusses some of the implications of the results.

6.1 MAJOR FINDINGS

1. The Experiment. The experiment was carried out successfully and resulted in the evaluation of 36 query responses for each of the 9 test agents. For the entire experiment, average call time was four minutes and five seconds. This call time translates to an average agent call rate of 15.4 calls per hour; this figure is slightly over 75% of the SCRTD productivity standard of 20 calls per hour. In terms of response accuracy, experimental results showed the vast majority (over 90%) of agent responses were judged to be satisfactory in terms of usefulness to the caller. A total of 9% of the test responses, however, were rated unsatisfactory, as compared with an SCRTD quality standard of 1%.

   Several explanations can be advanced for this substandard showing by the test agents in both speed and response quality terms. First, the test queries were more difficult than the average caller query. Second, novice agents were included in the testing. Perhaps these agents were not yet fully experienced in all aspects of data retrieval and/or call management. Third, experimental effects may have caused the test agents to be both more nervous and more cautious and deliberate than they would ordinarily have been at their normal work stations.

2. Test Queries. Based on average agent call times, the queries in Sets I and II were considerably more difficult than those in Sets III and IV. The test was divided into four Question Sets for logistical reasons. In statistical terms, the order of presentation of queries to the agents did not significantly impact call times. Within each set of queries, the average response times were significantly different, as would be expected. Set III had a significantly lower average response time over all agents, and also a significantly smaller amount of variability than the other three sets of queries.

3. Results on Call Times by Test Agent. The lowest average call times in the experiment were achieved by agents C3 (3:01) and C2 (3:08). Third place went to agent X2 (3:31), an agent who relied heavily on CCIS. These average call times translated into hourly productivities of 19.9, 19.2 and 17.0 calls, respectively. Agents X1 and C1, on the other hand, had the two highest average call times for the test and were not significantly different from each other in performance. Each of these two agents substantially increased the average call time and lowered overall productivity for their entire mode. Among both the CCIS and mixed mode agents, the novice agent was significantly slower in processing calls than the intermediate and advanced agents; the intermediate and advanced agents were statistically the same in response times in both of these groups. The overall responsiveness rankings of the nine test agents were generally consistent with their responsiveness ranking within each of the four Question Sets.

4. Results on Call Times by Mode. The CCIS mode had the lowest group average call time per query (3:51). The manual mode (4:03) and the mixed mode (4:21) placed second and third. The only statistically significant difference between the three groups occurred in Question Set I which evidently had more difficult queries. Here the CCIS mode agents performed much faster than the other two modes, who were effectively the same in average response time. When results are averaged over the entire test, however, the manual agents' call times exhibited much less variability than the other two modes; in fact, their call times were significantly different only in Question Set III.

   If novice data are eliminated from experiment results, the superiority of CCIS over the other two modes in terms of speed becomes much more pronounced. The average CCIS call time under these circumstances is 3:05; the mixed and manual modes are considerably slower on average with call times of 3:39 and 3:48, respectively.

5. Results on Response Accuracy by Test Agent and Skill/Experience Level. In the experiment, agent X3 clearly posted the highest overall response quality score with 89% of his responses rated in the "useful" (S-6 to S-10) range and only one unsatisfactory response. Agents M3, C2, C3 and X2 rank next, with 83% to 86% of their responses in this useful range. Agents M1, M2 and X1 had 70% to 78% in this range, while agent C1 scored the lowest of all the agents with only 65% of his

responses in this range. The advanced agents, as might be expected, scored considerably higher than the intermediates and novices; useful (S-6 to S-10) percentages for the three groups were 86%, 79% and 73% of the total, respectively.

6. Results on Response Accuracy by Mode. Comparison of response quality grades grouped by mode of data retrieval shows that the mixed mode group, which was allowed to use either CCIS, manual or both modes of data retrieval, scored the highest of the three groups, with 83% of its responses in the useful range and only 5% of its responses rated unsatisfactory. Behind the mixed mode group, the manual and CCIS groups are very evenly ranked, with the manual only slightly ahead (2 percentage points) in all grading categories. The response quality scores of the novice CCIS agent, however, are such that his proficiency with the system is suspect. If he had not yet mastered CCIS, then including his scores in the analysis leads to unfair comparisons betweens modes. If all novice agent scores are eliminated from the analysis, then the CCIS mode moves ahead of the manual mode to second place in overall qualitative rankings while the mixed mode remains in first place.

7. Overall Agent Performance. The intermediate and advanced agents for both the CCIS and mixed modes (C2 and C3, X2 and X3) all scored better than the overall average in terms of response speed as well as accuracy. The advanced agent (M3) just missed being included with these four agents; although his accuracy ratings were well above average, and his call rate was just slightly (four-tenths of one percent) below the overall average. The inter-

mediate manual agent, on the other hand, performed well above average on speed, but his accuracy rating was the second lowest of the nine agents tested. Finally, the novice agents in all three modes scored below average in terms of both speed and accuracy.

If novice agent data are eliminated, modal comparisons become clearer. The CCIS mode agents (C2 and C3) are the only ones having above-average call rates; the implication of this result is that their actual call rates were appreciably higher than those of the other agents. The intermediate and advanced CCIS agents, C2 and C3, also scored above average in terms of accuracy. Thus, when novice data are eliminated from consideration, the performances of the CCIS agents are clearly superior to those of the other two modes.

8. Level of Service by Mode. The CCIS mode agents clearly demonstrated the lowest average call time for responses that were judged "useful" (in the S-6 to S-10 range). The associated productivity for these responses is 17.1 calls per hour. The corresponding productivity measures for the manual and mixed modes were 15.6 and 14.5 per hour, respectively.

9. Call Management. The nature of an agent's call management style does appear to vary by mode of retrieval. Manual processing tends to lead to an agent response which begins earlier in the call but lasts longer in total time. CCIS processing tends to lead to a more concise response which begins later in the call.

6.2 IMPLICATIONS

The results of this experiment have several important implications for users of ATIS technology and for the transit information business in general. Among these are:

1. ATIS technology appears to have strong potential for assisting transit information agents in providing transit information to the public in an accurate, and timely manner. Experimental results tend to confirm that the objectives of increased call productivity and response accuracy are indeed met.

2. This strong ATIS potential will not be fully realized unless the sponsoring transit agency supports the users of the system through intensive agent training, data base support programs, and a continual fine tuning and enhancement process. Agent feedback should play a critical role in this development process.

3. It appears that the agent experience/skill level plays a major role in determining overall performance in each mode of information retrieval. As a result, novice agents should undergo rather intensive ATIS training to qualify them to operate such a system effectively. This requirement may have a significant impact on the objective of reducing costs through decreased training requirements.

4. Implementation of ATIS systems should clearly take place under mixed mode conditions (i.e., retention of the manual reference for verification purposes).

Whether or not the manual system can eventually be eliminated depends on the continual refinements and enhancements made to ATIS and on agent confidence in system capabilities. This requirement may also have an adverse impact on the objective of reducing costs by eliminating manual update functions.

5. The introduction of ATIS technology changes a transit agent's job in two important ways. First, information retrieval speed, and therefore agent productivity, is no longer a function of an agent's manual dexterity and/or initial choice of appropriate reference materials. Rather, new skills are required, such as typing and memorization of CCIS transaction keys and their intended applications. Second, because the CCIS provides alternative "correct" answers to a specific request, the agent is now faced with the task of interpreting and evaluating this "correct" data provided on the CRT, and selecting the most appropriate response to the caller query. Under the manual system, this evaluation process is generally limited to the agent's initial selection of reference materials, and only one response is searched out unless the caller specifically requests additional information. Thus, automation of data retrieval presumably reduces the need for intensive training to familiarize agents with bus routes and local geographic data so that they make the initial "correct" choice of reference materials.

The typing and function memorization skills required to make the first job change mentioned above are mechanical skills generally achievable through drills and practice on the part of the individual agent. The interpretive and evaluative skills

required for the second job change, however, still presume an intimate, working knowledge of the transit system and of local geography. Selecting the most appropriate response is not an easy task even for experienced agents, as the computer sometimes produces "correct" options that a trained agent would not intuitively select. For example, the computer might "prefer" an itinerary in which a customer walks several blocks to an express bus over an itinerary which involved a local bus and no walking. Agents whose training includes memorization of bus routes and geographic data are clearly better equipped to choose between the "correct" options provided. The implication of this second job change is that agents should receive intensive training with a strong emphasis on decision-making criteria. Learning why the computer selects a particular itinerary over another that might seem more intuitively "correct" is an extremely important part of this decision-making training. For this reason, it does not seem reasonable to expect a substantial reduction in agent training time will result from ATIS implementation.

Perhaps the relatively low call rates of the novice mixed and CCIS agents, particularly the novice CCIS agent, can be attributed to agent inability to choose between computer options, although it must also be remembered that the CCIS novice had not yet completed his initial 8-week standard agent training course. The performances of the intermediate and advanced CCIS agents clearly demonstrated their ability to make choices much more quickly.

6. Call management strategies should be an integral part of the ATIS training process. In order to avoid inconsistency and enhance agent productivity, the transit agency should specify the desired level of information to be provided and the desired sequence for effectively stating it in a given response.

APPENDIX A

TEST QUESTIONS

Appendix A contains a complete list of all 36 routing and schedule questions used in the experiment.

1. I'm staying at the Holiday Inn on Roscoe Boulevard right near the San Diego Freeway. I'd like to visit the Universal Studios for their last tour at 6 p.m. this coming Friday. What's the quickest way for me to return to the Holiday Inn after the tour at 8 p.m.?

2. I expect to be at the Northridge Plaza this Saturday morning. After lunch, say 1:00 p.m., I'd like to return home to 10640 La Tuna Canyon Road (near Wheatland Avenue). What time do I catch my bus? What's it going to cost me?

3. I'd like to get from Van Alden Avenue and Hatteras Street this afternoon to arrive before 5 p.m. at Rayen Street and Burnet Avenue (in Sepulveda). How frequently does the bus run between 3 p.m. and 5 p.m.?

4. I need bus service from Canoga Avenue and Mayall Street (in Chatsworth) to get to a 10 a.m. class at UCLA. How can I do it? My classes are over at 4 p.m. What's the best way to return home? What's the fare each way?

5. I finish my classes on Thursday at 3 p.m. at Cal State Northridge and want to get to the UCLA Medical Center as quickly as possible. When can I catch the first bus?

6. Tomorrow I want to go to the West Hills Hospital (23100 block of Sherman Way) to arrive there shortly before noon. I'm at Genesta Avenue and Chatsworth Street (Granada Hills). How long will it take me to get to the hospital?

7. I'm at Whitsett Avenue and Hamlin Street and want to attend the Grace Community Church on Roscoe Boulevard near Mary Ellen Avenue. I'm interested in attending the 6 p.m. services this coming Sunday evening. Can I make it? What's the fare?

8. What's the best way to get from Kadota Street and Herrick Avenue (in Sylmar) to the Bonaventure Hotel before 9 a.m. weekdays?

9. I'm meeting a friend at the Van Nuys Airport. How do I get there from De Garmo Avenue and Van Nuys Boulevard so I'll arrive about 5 p.m. today? Is there another option for me?

10. I'd like to get to the 9200 block of Hayvenhurst from Saticoy Street and Reseda Boulevard on a weekday afternoon by 4 p.m. How do I do it? What's the latest return routing I can get to return home on a weekday night?

11. How can I get to the Valley Youth Center (Victory and White Oak) from Sherman Oaks, Sunnyslope and Huston Street, this coming saturday morning to arrive by 9 a.m.? Is there an option with fewer transfers?

12. I get out of LA Valley College at 7 p.m. weeknights. Can I catch a bus taking me home to Granada Hills? I live in the 17400 block of Horace Street. (Note: last trip is at 6:51 p.m.)

13. Tomorrow morning, say before 11 a.m., I'd like to leave for Hillrose Street and Sherman Grove Avenue. I'm at Osborne Street and Beachy Avenue (in Arleta). I'd like to make the run with the fewest transfers. What's the one-way fare?

14. I'm on a flight arriving at Burbank Airport at 5 p.m. Friday evening. I'd like to get to Griffith Park and Verdugo Avenue (in Burbank) by the quickest routing. What if I miss the 5:30 bus?

15. I'm going to be at the Northridge Fashion Plaza on Friday and want to go to 650 South Hill Street in Los Angeles to arrive shortly before noon. (At the Plaza I'm near Shirley and Plummer.) What's the fare? How can I return to the Plaza, leaving downtown near 4 p.m.?

16. I would like to go from James Monroe High School to Kennedy High School this Friday to arrive near 4:30 p.m. What's the fare one way?

17. I would like to go from 9100 Reseda Boulevard to Universal and Lancashire Boulevards this Friday arriving around 10 a.m. What's the fare? Can I return at 4 p.m.?

18. I live near Mulholland Drive and Mulholland Highway and want to go to Ventura and Sepulveda Boulevards to meet someone this coming Friday. I'd like to get there about noon.

19. I live near Victory and Balboa Boulevards and want to leave for Kaiser Hospital (Cantara Street and Woodman) this coming Friday morning, say about 11 a.m. What time do I have to catch my bus? How long will the trip take me? I'd like to return home about 2 p.m. from Kaiser. What's my routing?

20. I want to get to LAX (Los Angeles International Airport) in time to meet a 5 p.m. flight this coming Friday evening. A friend can drop me off at the intersection of Hollywood Way and Thornton AVenue in Burbank. Can I make it? What's the travel time? What's the fare one way?

21. Can I catch a bus Thursday afternoon about 4 p.m. going from 10300 Topanga Canyon Boulevard to 18100 Nordhoff Street? What's the one way fare? Is it possible to return back to Topanga Canyon from Nordhoff at approximately 8 p.m.?

22. I want to visit a friend at Sherman Oaks Community Hospital, next Monday morning to arrive before noon. Can I catch a bus near Wilshire and Westwood Boulevards? What's my routing and how long will it take?

23. I'd like to go to the 700 block of S. Hill Street in downtown Los Angeles to arrive about noon Friday. I'm in the 9500 block of Haskell Avenue. When can I catch my bus and what is the routing? What's the one-way fare?

24. I want to meet a friend at the St. Joseph Medical Center (City of Burbank) on Friday afternoon about 3 p.m. I live near the intersection of Kester Avenue and Oxnard Street. What's my routing? Can I return home from the Center, leaving about 6:30 p.m.?

25. Is it possible for me to get from Topanga Canyon and Ventura Boulevards to Vermont Avenue near the Hollywood Freeway this coming Friday morning to arrive about 12 noon? What's an option to that routing? What's the one-way fare on both?

26. I've got to be at 6th and Hill Streets in downtown Los Angeles by 8:15 a.m. on Friday. I live near Van Nuys Boulevard and Sherman Way. What's the quickest routing? What's the one-way fare?

27. I'm meeting a friend at the Van Nuys Airport this coming Friday afternoon. He's got to catch a 6:30 p.m. flight out of Burbank Airport. Could you give me the quickest routing, please. What's the one-way fare?

28. Saturday morning I'd like to get to 4000 Laurel Canyon Boulevard about 11 a.m. I'm close to the intersection of Foothill Boulevard and Sunland Boulevard. What routing can you suggest? I'd like to return to my origin, leaving about 4 p.m. in the afternoon. What's the round trip fare?

29. I'm supposed to be meeting some friends for lunch arriving 12 noon on Friday at 2300 Ventura Boulevard. I'll be leaving from 14700 Brand Boulevard. Is there some bus routing which will help me get there? What if we were to meet on Saturday at the same time instead of Friday?

30. I expect to be downtown in Los Angeles Thursday evening near 4th and Spring Streets. I'd like to catch a bus around 7 p.m. going to San Fernando Road and Roxford Street. Is there an alternate routing?

31. I have a friend who wants to go to the West Valley Adult Occupational Training Center. He lives near 6800 White Oak Avenue. Is there a bus schedule which would permit him to do this after 12 noon on Friday? How frequently do the buses run? Is he any closer, time-wise, to the North Valley Adult Occupational Training Center? (If so, what is the routing?)

32. I'm going to be near Tuxford Street and Telfair Avenue around noon on Friday. I need to get to 7600 Vineland as soon thereafter as possible. Is there a routing which will take me there? If I happen to miss that bus, when does the next one get me there?

33. A friend is dropping me off at the Riverside Hospital next Monday afternoon. I'll need to leave the hospital around 4 p.m. for Woodman and Burbank Boulevards. What's the best routing you can suggest? What's the fare?

34. What's the earliest on Friday morning that I can leave Sherman Way and Van Nuys Boulevard to arrive downtown at 7th and Main Streets by 7 a.m.? How frequently do the buses run during the next two hours?

35. I expect to be in the City of Burbank near Olive Avenue and Golden Mall next Tuesday morning. I'd like to catch a bus so that I can arrive at 16900 San Fernando Mission Boulevard a little before noon. If I miss the first bus, what's the next available bus? We should be finished by 3 p.m. with business and I'd like to return to Olive Avenue and Golden Mall to meet a friend. What's the best routing?

36. Some friends are having a party Saturday evening at 14200 Sylvan Street. I live at 18900 Vanowen Street and would like to arrive by 7 p.m. at the party. Is it possible? Just give one routing. What is the latest bus I can catch on the same route to return later the same evening?

APPENDIX B

LATIN SQUARES USED IN THE EXPERIMENT

As described in Section 3.1, it was necessary to establish a unique ordering for test questions for each test agent in order to perform ANOVA (analysis of variance) calculations on call time results. Results of these calculations are shown in Appendix D.

In order to limit the number of agents tested, the 36 test questions were divided into four sets of 9 questions each. A separate Latin Square was devised for each of these four question sets. These four Latin Squares are shown in Tables B-1, B-2, B-3, and B-4.

In interpreting these four tables, one should follow the test agent designation. For example, on Table B-1, for Question Set I, the Novice CCIS agent, C1, received Test Question #1 first, question #9 second, #8 third, and so on.

TABLE B-1. 9X9 LATIN SQUARE USED TO ESTABLISH ORDER OF TEST QUESTIONS IN QUESTION SET I

ORDER OF TEST QUESTION	TEST QUESTION NUMBER								
	1	2	3	4	5	6	7	8	9
FIRST	C1	C2	C3	M1	M2	M3	X1	X2	X3
SECOND	C2	C3	M1	M2	M3	X1	X2	X3	C1
THIRD	C3	M1	M2	M3	X1	X2	X3	C1	C2
FOURTH	M1	M2	M3	X1	X2	X3	C1	C2	C3
FIFTH	M2	M3	X1	X2	X3	C1	C2	C3	M1
SIXTH	M3	X1	X2	X3	C1	C2	C3	M1	M2
SEVENTH	X1	X2	X3	C1	C2	C3	M1	M2	M3
EIGHTH	X2	X3	C1	C2	C3	M1	M2	M3	X1
NINTH	X3	C1	C2	C3	M1	M2	M3	X1	X2

NOTATION:

C1 = CCIS, Novice;
M2 = Manual, Intermediate;
X3 = Mixed, Advanced; etc.

TABLE B-2. 9X9 LATIN SQUARE USED TO ESTABLISH ORDER OF TEST QUESTIONS IN QUESTION SET II

ORDER OF TEST QUESTION	TEST QUESTION NUMBER								
	10	11	12	13	14	15	16	17	18
FIRST	M3	C2	C3	M2	X2	M1	X1	X3	C1
SECOND	C3	M2	X3	X2	C1	X2	C2	M1	M2
THIRD	C3	X2	X3	C2	M2	C1	M1	M3	X1
FOURTH	X3	M2	M3	X2	C2	X1	C1	C3	M1
FIFTH	X1	C3	M1	M3	X3	M2	X2	C1	C2
SIXTH	C2	X1	X2	C1	M1	X3	C3	M2	M3
SEVENTH	M2	C1	C2	M1	X1	C3	M3	X2	X3
EIGHTH	X2	M1	M2	X1	C1	M3	X3	C2	C3
NINTH	M1	X3	C1	C3	M3	C2	M2	X1	X2

NOTATION:

C1 = CCIS, Novice;
M2 = Manual, Intermediate;
X3 = Mixed, Advanced; etc.

TABLE B-3. 9X9 LATIN SQUARE USED TO ESTABLISH ORDER OF TEST QUESTIONS IN QUESTION SET III

ORDER OF TEST QUESTION	TEST QUESTION NUMBER								
	19	20	21	22	23	24	25	26	27
FIRST	C3	M3	X3	C1	X2	X1	M1	M2	C2
SECOND	X3	C3	M3	X1	M2	M1	C1	C2	X2
THIRD	X2	C2	M2	M3	M1	C3	X3	C1	X1
FOURTH	M3	X3	C3	M1	C2	C1	X1	X2	M2
FIFTH	M1	X1	C1	C2	X3	X2	M2	M3	C3
SIXTH	C1	M1	X1	X2	M3	M2	C2	C3	X3
SEVENTH	M2	X2	C2	C3	C1	X3	M3	X1	M1
EIGHTH	C2	M2	X2	X3	X1	M3	C3	M1	C1
NINTH	X1	C1	M1	M2	C3	C2	X2	X3	M3

NOTATION:

X3 = Mixed, Advanced;
C2 = CCIS, Intermediate;
M1 = Manual, Novice; etc.

TABLE B-4. 9X9 LATIN SQUARE USED TO ESTABLISH ORDER OF TEST QUESTIONS IN QUESTION SET IV

ORDER OF TEST QUESTION	TEST QUESTION NUMBER								
	28	29	30	31	32	33	34	35	36
FIRST	X2	M3	C1	M2	M1	X1	X3	C3	C2
SECOND	M3	M1	X2	C3	C2	M2	X1	C1	X3
THIRD	C3	C1	M2	X3	X2	C2	M1	X1	M3
FOURTH	C2	X3	M1	X2	X1	C1	C3	M3	M2
FIFTH	X3	X1	C2	M3	M2	X2	C1	M1	C3
SIXTH	C1	X2	C3	X1	M3	X3	C2	M2	M1
SEVENTH	X1	M2	X3	M1	C3	M3	X2	C2	C1
EIGHTH	M2	C3	X1	C2	C1	M1	M3	X3	X2
NINTH	M1	C2	M3	C1	X3	C3	M2	X2	X1

NOTATION:

C1 = CCIS, Novice;
X2 = Mixed, Intermediate;
M3 = Manual, Advanced; etc.

APPENDIX B REFERENCES

Federer, Walter T., Experimental Design - Theory and Application (New York: MacMillan Co., 1955).

Kirk, Roger E., Experimental Design: Procedures for the Behavioral Sciences (Belmont, CA: Brooks/Cole Publishing Co., 1966)

Natrella, Mary A., Experimental Statistics Handbook 91 (Washington, DC: National Bureau of Standards, U.S.G.P.O., 1966).

APPENDIX C
AGENT CALL TIME RESULTS

In this Appendix, call time statistics for each agent in each of the four question sets are presented. Means and standard deviations for each agent are shown as calculated over Questions 1-9, 10-18, 19-27, and 28-36 (Question Sets I, II, III and IV).

In interpreting these tables, one should note that each table element represents one agent's results. For example, in Table C-1, the Mixed Mode Novice had a mean call time of 7:04 and a standard deviation of 2:21 for Question Set I.

Results for each test agent were also averaged over the entire experiment (Questions 1-36). These results are shown in Table 3-4 (in which call times in seconds were divided into 3600 to yield agent calls per hour) and Table 3-5.

TABLE C-1. CALL TIME STATISTICS
FOR QUESTION SET I
(IN MINUTES AND SECONDS)

AGENT SKILL/ EXPERIENCE LEVEL	MODE OF DATA RETRIEVAL		
	MANUAL	CCIS	MIXED
NOVICE			
MEAN	5:19	5:43	7:04
STD. DEV.	2:17	1:43	2:21
INTERMEDIATE			
MEAN	4:39	3:13	3:33
STD. DEV.	1:23	0:43	0:19
ADVANCED			
MEAN	5:06	2:57	4:29
STD. DEV.	2:25	0:56	1:26

TABLE C-2. CALL TIME STATISTICS
FOR QUESTION SET II
(IN MINUTES AND SECONDS)

AGENT SKILL/ EXPERIENCE LEVEL	MODE OF DATA RETRIEVAL		
	MANUAL	CCIS	MIXED
NOVICE			
MEAN	5:00	5:53	6:38
STD. DEV.	1:38	2:17	3:03
INTERMEDIATE			
MEAN	3:15	3:59	3:23
STD. DEV.	1:18	0:58	1:10
ADVANCED			
MEAN	4:10	3:38	4:00
STD. DEV.	2:04	1:03	1:31

TABLE C-3. CALL TIME STATISTICS
FOR QUESTION SET III
(IN MINUTES AND SECONDS)

AGENT SKILL/ EXPERIENCE LEVEL	MODE OF DATA RETRIEVAL		
	MANUAL	CCIS	MIXED
NOVICE			
MEAN	3:57	4:41	4:05
STD. DEV.	1:40	1:11	1:51
INTERMEDIATE			
MEAN	3:36	2:24	2:54
STD. DEV.	1:36	0:48	1:09
ADVANCED			
MEAN	2:52	2:16	2:56
STD. DEV.	1:22	0:55	1:16

TABLE C-4. CALL TIME STATISTICS
FOR QUESTION SET IV
(IN MINUTES AND SECONDS)

AGENT SKILL/ EXPERIENCE LEVEL	MODE OF DATA RETRIEVAL		
	MANUAL	CCIS	MIXED
NOVICE			
MEAN	4:00	5:23	5:12
STD. DEV.	1:47	1:42	1:49
INTERMEDIATE			
MEAN	3:14	2:54	4:13
STD. DEV.	1:21	1:01	2:05
ADVANCED			
MEAN	3:32	3:14	3:39
STD. DEV.	1:49	0:59	1:48

APPENDIX D
ANALYSIS OF VARIANCE (ANOVA) RESULTS

The analysis of variance, or ANOVA, is a statistical test used to determine the validity of certain hypotheses concerning the distribution of experimental data. In the case of the CCIS experiment, the data in question were the call times, in seconds, for each response to each test question by each test agent on each Question Set. The hypotheses tested involve the influence of certain variables (the independent variables) on call time results. Controlled testing attempts to remove all extraneous influences which might have a potential effect on experimental results other than these independent variables. In the CCIS experiment, the independent variables were the order in which test questions were asked, differences in test question content, differences between modes of data retrieval, and differences between agents due to varying skill/experience levels.

In ANOVA calculations, a "residual" variable is also postulated to explain distributions observed in experimental data. If this residual is very small, approaching zero, then the independent variables can be said to reliably "explain" variations in experimental data. If the residual is large, then the explanatory power of the postulated variables is diminished. In order to determine the explanatory significance of the variables, a detailed calculation process is followed, culminating in the determination of an F-statistic. The F-distribution is a standard statistical device used in comparing variations between distributions. In this case, the distributions to be compared are the variance of the independent, or "explanatory", variables, and the variance of the residual from their respective average values. Variance is a statistical concept dealing with the dispersion of observed data points.

The first step in calculating the F-statistic is the calculation of differences between individual data points (here, observed call times) and the mean for the data grouping in question (for example, in this experiment, groups of interest included all agents of one particular skill/experience level, all agents using a particular data retrieval mode, etc.). These differences are then squared to eliminate any influence of negative numbers, and summed. These sums of squares are then divided by the "degrees of freedom", which is the number of experimental observations of the variable in question minus one. For example, when considering modes of data retrieval as an independent variable, its number of degrees of freedom is two (or three experimental modes minus one). This division yields the Mean Sum of Squares for the variable in question. A similar calculation is performed for the residual, which for this experiment had 56 degrees of freedom (56 = total observations for each question set minus the degrees of freedom for each of the independent variables minus one). The F-statistic results when the mean sum of squares for the independent variable in question is divided by the mean sum of squares for the residual.

Once the F-statistic has been calculated, it is then compared with values on a standard table of F-statistics to determine the degree to which the residual affects the distribution of data points given the number of degrees of freedom for the independent variable in question and the residual. This comparison can be used to determine reliability of the independent variable as an explanation for variations in observed data (agent call times). The standard statistical measure of reliability is 95% confidence. If a variable exceeds this level, it is by convention considered "statistically significant", or reliable as an "explanatory" variable.

Tables D-1 through D-4 present all agent call times (in seconds) recorded during the course of the experiment

for Question Sets I, II, III and IV. In these tables, the call times are arranged so as to correspond to the Latin Squares shown in Appendix B. The Latin Square arrangement was used so as to eliminate the influence of one of the independent variables, the order in which test questions were asked, on experimental results. These tables show call times (in seconds) posted by each agent on each test question.

Tables D-5 through D-8 present ANOVA calculation results for each of the four Question Sets. They present these results in a standard format used for displaying ANOVA calculations, with independent variables in the left hand column, sums of squares in the next, degrees of freedom in the next, mean sums of squares (MSS) in the next, and F-statistics in the right-hand column. Asterisks next to F-statistics in these tables indicate significance of the variable in question as a reliable explanatory variable; one asterisk indicates 95% confidence of significance, while two asterisks indicate significance at the 99% confidence level.

Table D-9 presents an overall summary of F-statistics, and includes the right-hand columns of Tables D-5 through D-8. The data in this table lead to the following conclusions concerning the independent variables in the CCIS experiment.

- Test Question Order: Controlling the order of presentation of test questions through the use of Latin Squares was successful, as no reliable differences can be detected in call times as a result of this factor.
- Test Question Content: Different test questions, as expected, produced highly significant differences in agent call times.

- Modes of Data Retrieval: The only reliable difference between the three modes of retrieval occurred in Question Set I which contained the most difficult queries. Here the CCIS agents performed substantially faster than the manual and mixed mode agents. For the other three Question Sets, no statistically reliable difference between modes was found. The markedly slower performance of the novice agents as compared to the intermediate and advanced agents in their respective groups was a strong contributing factor to this overall result. When results are averaged by mode, the relatively slower showings of the novices tend to dilute the faster performances of the more experienced agents.
- Agent Skill/Experience Level: Reliable differences between agents of different skill/experience level were discernible in the CCIS and mixed mode results for all four question sets. In these modes, the novice agent was substantially slower than the other two agents. Significant differences between call times for manual agents of different skill/ experience levels were found only in Question Set III. On the other Question Sets, manual agent average call times exhibit a relatively higher variability than do those of the other two modes. As a result, no statistical difference between the call times of the manual agents can be found on these Question Sets.

TABLE D-1. LATIN SQUARE OF AGENT CALL TIMES FOR QUESTION SET I (IN SECONDS)

TEST QUESTION ORDER	TEST QUESTION NUMBER								
	1	2	3	4	5	6	7	8	9
FIRST	C1	C2	C3	M1	M2	M3	X1	X2	X3
	327	238	187	280	165	186	463	237	385
SECOND	C2	C3	M1	M2	M3	X1	X2	X3	C1
	180	184	316	317	122	267	230	240	312
THIRD	C3	M1	M2	M3	X1	X2	X3	C1	C2
	149	282	440	580	179	181	312	299	262
FOURTH	M1	M2	M3	X1	X2	X3	C1	C2	C3
	117	234	425	665	193	209	420	150	195
FIFTH	M2	M3	X1	X2	X3	C1	C2	C3	M1
	152	322	434	204	119	263	216	136	374
SIXTH	M3	X1	X2	X3	C1	C2	C3	M1	M2
	100	571	210	306	205	160	222	367	271
SEVENTH	X1	X2	X3	C1	C2	C3	M1	M2	M3
	469	230	204	439	119	115	650	307	402
EIGHTH	X2	X3	C1	C2	C3	M1	M2	M3	X1
	233	404	555	212	108	248	335	291	444
NINTH	X3	C1	C2	C3	M1	M2	M3	X1	X2
	242	263	197	298	241	291	330	324	201

NOTATION: Agent Designation / Agent Call Time

TABLE D-2. LATIN SQUARE OF AGENT CALL TIMES FOR QUESTION SET II (IN SECONDS)

TEST QUESTION ORDER	TEST QUESTION NUMBER								
	10	11	12	13	14	15	16	17	18
FIRST	M3	C2	C3	M2	X2	M1	X1	X3	C1
	302	233	318	349	213	202	402	174	236
SECOND	C3	M2	X3	X2	C2	X2	C2	M1	M2
	576	349	705	260	219	191	211	247	152
THIRD	C3	X2	X3	C2	M2	C1	M1	M3	X1
	249	298	156	263	213	511	272	98	195
FOURTH	X3	M2	M3	X2	C2	X1	C1	C3	M1
	309	213	480	210	216	325	243	131	180
FIFTH	X1	C3	M1	M3	X3	M2	X2	C1	C2
	274	168	329	320	413	173	112	224	150
SIXTH	C2	X1	X2	C1	M1	X3	C3	M2	M3
	258	424	334	194	300	142	212	197	181
SEVENTH	M2	C1	C2	M1	X1	C3	M3	X2	X3
	266	358	360	437	486	175	100	119	184
EIGHTH	X2	M1	M2	X1	C1	M3	X3	C2	C3
	164	490	133	645	325	123	349	177	173
NINTH	M1	X3	C1	C3	M3	C2	M2	X1	X2
	241	177	512	320	300	280	58	125	189

NOTATION: Agent Designation / Agent Call Time

TABLE D-3. LATIN SQUARE OF AGENT CALL TIMES FOR QUESTION SET III. (IN SECONDS)

TEST QUESTION ORDER	TEST QUESTION NUMBER								
	19	20	21	22	23	24	25	26	27
FIRST	C3	M3	X3	C1	X2	X1	M1	M2	C2
	251	183	190	213	84	464	332	107	89
SECOND	X3	C3	M3	X1	M2	M1	C1	C2	X2
	218	166	312	146	122	360	358	81	97
THIRD	X2	C2	M2	M3	M1	C3	X3	C1	X1
	167	139	264	86	169	185	206	200	112
FOURTH	M3	X3	C3	M1	C2	C1	X1	X2	M2
	206	190	152	143	110	347	265	96	82
FIFTH	M1	X1	C1	C2	X3	X2	M2	M3	C3
	241	370	289	109	109	266	314	119	120
SIXTH	C1	M1	X1	X2	M3	M2	C2	C3	X3
	355	389	313	165	94	331	176	83	116
SEVENTH	M2	X2	C2	C3	C1	X3	M3	X1	M1
	269	281	233	75	225	347	238	161	161
EIGHTH	C2	M2	X2	X3	X1	M3	C3	M1	C1
	166	317	224	111	175	252	113	87	185
NINTH	X1	C1	M1	M2	C3	C2	X2	X3	M3
	198	359	248	136	79	190	187	93	58

NOTATION: Agent Designation / Agent Call Time

TABLE D-4. LATIN SQUARE OF AGENT CALL TIMES FOR QUESTION SET IV. (IN SECONDS)

TEST QUESTION ORDER	TEST QUESTION NUMBER								
	28	29	30	31	32	33	34	35	36
FIRST	X2	M3	C1	M2	M1	X1	X3	C3	C2
	300	397	442	282	128	196	80	213	299
SECOND	M3	M1	X2	C3	C2	M2	X1	C1	X3
	183	288	204	180	88	126	187	426	300
THIRD	C3	C1	M2	X3	X2	C2	M1	X1	M3
	167	312	60	324	89	107	225	416	255
FOURTH	C2	X3	M1	X2	X1	C1	C3	M3	M2
	189	232	90	277	159	194	114	318	271
FIFTH	X3	X1	C2	M3	M2	X2	C1	M1	C3
	330	295	139	322	106	184	285	349	264
SIXTH	C1	X2	C3	X1	M3	X3	C2	M2	M1
	298	213	127	329	122	142	147	271	359
SEVENTH	X1	M2	X3	M1	C3	M3	X2	C2	C1
	438	276	84	400	281	132	171	213	449
EIGHTH	M2	C3	X1	C2	C1	M1	M3	X3	X2
	208	142	474	216	144	180	107	335	561
NINTH	M1	C2	M3	C1	X3	C3	M2	X2	X1
	144	164	70	360	110	254	148	276	313

NOTATION: Agent Designation / Agent Call Time

TABLE D-5. RESULTS OF ANOVA CALCULATIONS FOR QUESTION SET I

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SUM OF SQUARES	F-STATISTIC
TEST QUESTION ORDER	60,979	8	7,622	1.10
TEST QUESTION CONTENT	358,644	8	44,833	6.45**
INDIVIDUAL AGENTS	448,950	8	--	--
MODE OF RETRIEVAL	74,714	2	37,357	5.38**
SKILL/EXPERIENCE OF AGENTS:				
Within Mixed Mode	214,690	2	107,345	15.45**
Within CCIS Mode	150,242	2	75,121	10.81**
Within Manual Mode	9,304	2	4,652	0.67
RESIDUAL	389,224	56	6,950	--
TOTAL	1,257,817	80	--	--

* indicates significance at 95% confidence level.

** indicates significance at 99% confidence level.

TABLE D-6. RESULTS OF ANOVA CALCULATIONS FOR QUESTION SET II

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SUM OF SQUARES	F-STATISTIC
TEST QUESTION ORDER	50,015	8	6,252	0.65
TEST QUESTION CONTENT	347,236	8	43,404	4.53**
INDIVIDUAL AGENTS	351,397	8	--	--
MODE OF RETRIEVAL	14,587	2	7,294	0.76
SKILL/EXPERIENCE OF AGENTS:				
Within Mixed Mode	192,054	2	96,027	10.03**
Within CCIS Mode	95,194	2	47,597	4.97*
Within Manual Mode	49,562	2	24,781	2.59
RESIDUAL	536,158	56	9,574	--
TOTAL	1,284,806	80	--	--

* indicates significance at 95% confidence level.

** indicates significance at 99% confidence level.

TABLE D-7. RESULTS OF ANOVA CALCULATIONS FOR QUESTION SET III

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SUM OF SQUARES	F-STATISTIC
TEST QUESTION ORDER	36,328	8	4,541	2.76*
TEST QUESTION CONTENT	403,111	8	50,389	30.63**
INDIVIDUAL AGENTS	175,335	8	--	--
MODE OF RETRIEVAL	6,067	2	3,034	1.84
SKILL/EXPERIENCE OF AGENTS:				
Within Mixed Mode	29,456	2	14,728	8.95**
Within CCIS Mode	120,209	2	60,104	36.54**
Within Manual Mode	19,603	2	9,802	5.96**
RESIDUAL	92,146	56	1,645	--
TOTAL	706,923	80	--	--

* indicates significance at 95% confidence level.

** indicates significance at 99% confidence level.

TABLE D-8. RESULTS OF ANOVA CALCULATIONS FOR QUESTION SET IV.

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SUM OF SQUARES	F-STATISTIC
TEST QUESTION ORDER	49,262	8	6,158	1.06
TEST QUESTION CONTENT	401,128	8	50,141	8.65**
INDIVIDUAL AGENTS	198,198	8	--	--
MODE OF RETRIEVAL	29,292	2	14,646	2.53
SKILL/EXPERIENCE OF AGENTS:				
Within Mixed Mode	40,129	2	20,064	3.46*
Within CCIS Mode	119,027	2	59,514	10.27**
Within Manual Mode	9,750	2	4,875	0.84
RESIDUAL	324,656	56	5,797	--
TOTAL	973,244	80	--	--

* indicates significance at 95% level of confidence.

** indicates significance at 99% level of confidence.

TABLE D-9. SUMMARY OF ANOVA RESULTS
FOR ALL QUESTION SETS
(F-STATISTICS)

SOURCE OF VARIATION	QUESTION SET I	II	III	IV
TEST QUESTION ORDER	1.10	0.65	2.76	1.06
TEST QUESTION CONTENT	6.45**	4.53**	30.63**	8.65**
INDIVIDUAL AGENTS	--	--	--	--
MODE OF RETRIEVAL	5.38**	0.76	1.84	2.53
SKILL/EXPERIENCE OF AGENTS:				
Within Mixed Mode	15.45**	10.03**	8.95**	3.46**
Within CCIS Mode	10.81**	4.97*	36.54**	10.27**
Within Manual Mode	0.67	2.59	5.96**	0.84

* indicates significance at 95% confidence level.

** indicates significance at 99% confidence level.

APPENDIX E

CONTROLLED EXPERIMENT DATA LOG

Complete before Query

REFERENCE DATA SECTION

* TIA Identification:
  * Mode (circle one) M C X
  * Agent (circle one) 1 2 3
* Query Person (circle one) F M O
* Query Identification:
  * Set (circle one) 1 2 3 4
  * Query # (circle one) 1 2 3 4 5 6 7 8 9

Actual Agent Response

TESTING SECTION

Events	Times
Start	* 00
End query	*
End call	*

Observations

X only: Retrieval (circle one)

M C B

Review of Response

EVALUATION SECTION - POST TEST

* Evaluator name _____

APPENDIX F

CCIS EXPERIMENT - AGENT RESPONSE QUALITY GRADES

QUESTION NUMBER	MIXED			CCIS			MANUAL		
	NOV	INT	ADV	NOV	INT	ADV	NOV	INT	ADV
1	S-9	S-5	S-9	S-3	S-8	S-8	S-8	S-8	S-8
2	S-10	S-4	S-10	S-5	S-4	S-5	S-6	S-8	S-9
3	U	S-8	S-4	S-5	S-6	S-5	S-7	S-9	S-9
4	S-4	S-5	S-8	S-3	S-6	S-8	S-6	S-8	S-4
5	S-7	S-6	S-9	U	S-8	S-6	S-8	S-4	S-7
6	S-6	S-9	S-8	S-9	S-10	S-10	S-6	S-6	S-8
7	S-5	S-8	S-8	U	S-7	S-10	U	S-2	S-8
8	S-9	S-6	S-8	S-7	S-8	S-8	S-8	S-8	S-9
9	S-7	S-9	S-10	S-7	S-9	U	S-7	S-7	S-7
10	S-7	S-5	S-7	U	S-5	S-7	U	S-7	S-7
11	S-7	S-9	S-8	U	S-9	S-9	S-5	S-9	S-9
12	U	S-8	U	U	U	U	S-6	U	S-4
13	S-6	S-4	S-4	U	S-8	S-5	U	U	S-6
14	S-6	S-7	S-10	S-8	S-10	S-9	S-10	S-6	S-7
15	S-8	S-8	S-9	S-6	S-7	S-9	S-9	S-7	S-8
16	S-9	U	S-5	U	S-9	S-9	S-9	U	U
17	S-7	S-9	S-9	S-8	S-9	S-10	S-6	S-5	S-6
18	S-7	S-9	S-10	S-9	S-9	S-10	S-5	S-6	S-7
19	S-7	S-9	S-10	S-9	S-10	S-10	S-7	S-5	S-7
20	S-9	S-9	S-10	S-9	S-7	S-8	S-9	S-5	S-10
21	S-4	S-7	S-8	U	S-7	S-7	S-7	S-9	S-7
22	S-10	S-8	S-7	S-9	U	S-8	S-10	S-9	S-9
23	S-9	S-9	S-10	S-10	S-10	S-10	S-10	S-9	S-10
24	S-9	S-8	S-6	S-8	S-8	S-8	S-8	S-9	S-8
25	S-9	S-9	S-10	S-9	S-10	S-6	S-7	S-9	S-9
26	S-5	S-9	S-8	S-8	S-8	S-8	S-4	S-8	S-9
27	U	S-9	S-10	S-10	S-9	U	S-9	S-10	S-9
28	S-9	S-7	S-10	S-9	S-10	S-10	S-4	S-7	U
29	S-9	S-9	S-9	S-9	S-9	S-9	S-9	S-3	S-8
30	S-7	S-8	S-7	S-8	S-8	S-8	S-6	S-7	S-6

QUESTION NUMBER	MIXED			CCIS			MANUAL		
	NOV	INT	ADV	NOV	INT	ADV	NOV	INT	ADV
31	S-3	S-8	S-8	S-3	S-9	S-9	S-8	S-5	S-5
32	S-10	S-7	S-10	S-9	S-10	S-9	S-10	S-9	S-10
33	S-9	S-7	S-9	S-9	U	S-8	S-7	S-9	S-9
34	S-9	S-9	S-8	S-9	S-9	S-7	S-7	S-8	S-9
35	S-8	S-8	S-8	S-8	S-8	S-9	S-8	S-8	S-8
36	S-7	S-6	S-8	S-6	S-4	S-7	S-5	U	S-7

U.S. Department
of Transportation

arch and
ial Programs
inistration

Square
dge, Massachusetts 02142

Official Business
Penalty for Private Use $300

Postage and Fees Paid
Research and Special
Programs Administration
DOT 513

U.S.MAIL