EVALUATION OF AN AUTOMATED FLIGHT TRAINING SYSTEM

Joseph A. Puig, Naval Training Equipment Center
and
Susan Gill, Naval Air Training Command

INTRODUCTION

Statement of Problem

To determine the effectiveness of an automated, adaptive GCA module, an experimental comparison of training with this system and conventional training was performed in the Advanced Jet Phase at NAS, Chase Field, Beeville, Texas.

Review of Previous Work

The Naval Training Equipment Center has been involved in a continuing project of programmed and adaptive training. Digital computer technology and advances in performance measurement techniques have provided a means for implementing these training concepts.

The advantages of automated adaptive training include standardization of instruction, progress tailored to match the individual's abilities, and objective performance measurement. Additionally, costs could be decreased by reducing the number and required experience level of the instructors.

An exploratory study was conducted in 1971 to demonstrate the feasibility of implementing an automated adaptive training program (Charles and Johnson, 1972). Automated ground controlled approach and emergency procedures tasks were implemented on the NAVTRAEQUIPCEN Training Device Computer (TRADEC) and tested with operational pilots.

The results demonstrated the feasibility of automated training and its acceptance by operational personnel. What remained to be done was an evaluation of the GCA module in an operational flight trainer.

Hypotheses to be Investigated

The hypothesis set forth in this investigation was that training with the automated GCA module would be as effective as conventional training. An alternative hypothesis was that the automated system would be more effective than conventional training. Included in this definition of effectiveness was the number of trials that could be completed within a given time period. This alternative hypothesis seemed warranted as the reset capability of the GCA module should make it possible to run more students and cover more material in a shorter period of time than with conventional methods.

METHOD

This section describes the experimental design, selection of subjects, apparatus, and administration of the experiment.

Experimental Design and Subject Selection

A matched subjects design was used with two groups: an experimental group (automatic mode) and a control group (manual mode). Matching was done on the basis of cumulative flight grades upon completion of Basic Jet Training of students from squadron VT-26 (basic) for assignment to advanced squadrons VT-24 and VT-25. However, since only about 25% of the total student population coming into the program was used, this smaller group was matched again. As an added precaution in preventing a bias from the influence of individual squadron training, the students were assigned to the experimental and control groups from alternate squadrons on consecutive months; e.g., one group received two students from VT-24 one month and two from VT-25 the following month. There were 18 students in each group, for a total of 36 in the program. The basic flight grades averaged 3.066 for the experimental group, and 3.064 for the control group.

System Description

The trainer provides full support to student and instructor by performing all routine functions including syllabus and trainer configuration control automatically until overridden by instructor input or student needs.

The implementation requirements for the GCA module include the following hardware systems:

a. A stored training program

b. Voice generation system

c. Graphics display terminal

d. Printer terminal

e. Interface with the basic trainer.

Integration of the GCA module did not require modification of the design of the existing Device 2F90. The automated training hardware interfaced with the basic 2F90 simulation program as shown in Figure 1.

The Cubic Corp. "DIGITALK" speech generation system assembled a fixed vocabulary into phrases and sentences under computer control. This equipment provided the GCA voice command input requirement. The voice system generated words and phrases and assembled them into the standard GCA phraseology for approach control. This system also provided the voice command input for feedback and diagnostic messages concerning student performance.

A graphics CRT (Tektronix Display Terminal Model 4010) and a line printer (Tally Corp. Model 2000) with keyboard were located near one of the instructor stations of Device 2F90. The results of each approach were output to this display subsystem. The CRT provided a time plot of each approach in the format of a GCA Precision Approach Radar (PAR) display, i.e., a vertical profile of aircraft position with respect to the glide slope, and a horizontal profile of aircraft position with respect to the glide slope, and a horizontal profile of aircraft position with respect to course. The printer provided a hard copy summary of conditions and performance for each approach.

Performance Measures

As the GCA has very definitive performance requirements, measures related to operational performance appear feasible. Two separate scores are used. The first reflects performance during the run (path measures); the second reflects position relative to the runway at the end of the control phase of the PAR (gate measures).

Scoring

The performance measures include about 15 path measures and 5 gate measures. These are combined to provide a single score for the adaptive scheduling plan. Figure 2 illustrates the basic logic developed for final scoring. In effect, a path score is computed for all runs. If successful, a gate score is computed and combined with the path scores for a total run score. If the run terminates in a wave-off or a crash, the path score is adjusted to compensate for the position of run completed.

Training Course and Difficulty Levels

The training course was comprised of seven different types of GCA approaches arranged in order of increasing difficulty, as follows: Precision Approach Radar (PAR), Airport Surveillance Radar (ASR), Compass System Failure (CSF), Minimum Fuel (MF), Partial Panel (PP), Partial Panel and Compass System Failure (PP & CSF), and Partial Panel and Minimum Fuel (PP & MF). There were several difficulty levels within each of the seven approach categories, totalling 65 runs. The 65 difficulty levels were generated by assigning weights to four adaptive variables (wind, aircraft weight, air turbulence, and type of approach) and summing across the conditions involved. The course which was developed is shown in Table 1.

Procedures

An initial study was performed during installation, checkout, and preliminary evaluation of the system. This provided an opportunity to train instructors and check procedures. Procedures were modified as a

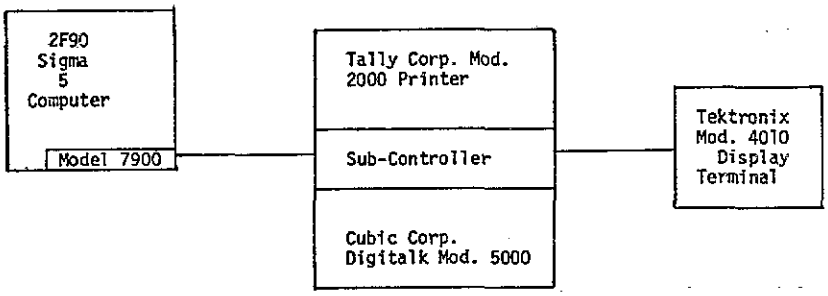

Figure 1. System Interface

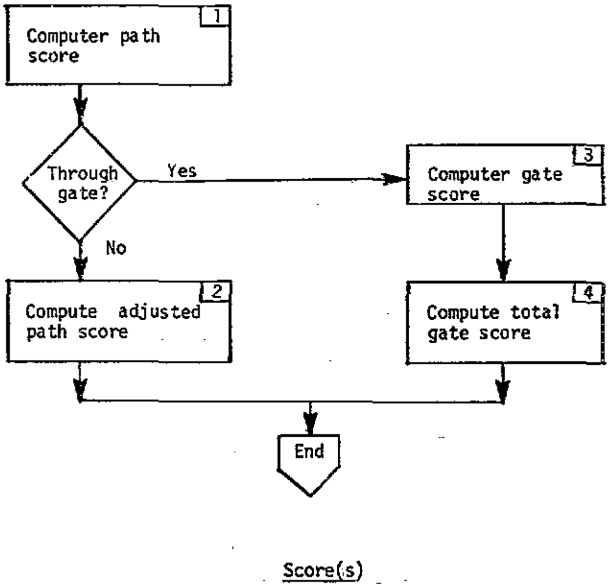

Figure 2. Scoring Logic Flow Chart

result of this initial study and a better grasp of the variability within and between students was obtained. A rough estimate of these variances was helpful in estimating the number of subjects required in the groups being tested. In essence, this study provided information upon which to refine the main experiment in a manner which increased the probability of producing meaningful results with a minimum expenditure of time and money.

The Trainer Experiment

The experimental group was trained using the Automated GCA concept. The control group was trained in the conventional manner (i.e., with instructor and present syllabus as per CNATRA Instruction 1542.20A, 6 July 1973). In each case, training was to proficiency, and the automatic performance measurement capabilities of the GCA module were used in addition to instructor scoring.

Two operation modes are available: a manual mode and an automatic mode. The automatic mode is implemented as the default mode; i.e., the automatic mode will appear as the selected mode when the system is energized. Both modes are similar in that they provide GCA training. However, the automatic mode provides for fully automated training with adaptive syllabus control and simulated GCA controller.

The manual mode is the conventional way of training and requires that the instructor enter the conditions for each GCA approach and act as the GCA controller. In both modes, performance measurement is provided. Table 2 summarizes the two modes.

Since the primary purpose of the GCA training mode was to provide as much training as possible in the final approach phase, students were not required to fly a complete pattern for each approach. Instead, the aircraft was repositioned to base leg after completion of the approach or waveoff for all but minimum fuel approaches. For minimum fuel approaches, the aircraft was positioned to downwind leg after completion of approach or waveoff.

The GCA training mode was normally initiated from an 'in-flight' condition. When airborne and the 'in-flight' condition was attained, the student was 'cleared' (audio instructions) to a specific heading and altitude (2,000 feet, straight ahead). The program detected the aircraft reaching the specified altitude and repositioned the aircraft to one of two geographical locations depending on whether the approach was

Table 1 2F90 AFTS GCA Course

Level	Wind		Weight	Turbulence	GCA Type	Code
	Direction	Speed				
1	130	20	15,000	None	PAR	0000
2	130	5	15,000	None	PAR	1000
3	310	10	15,000	None	PAR	2000
4	130	20	13,000	None	PAR	0100
5	130	20	17,000	None	PAR	0200
6	130	5	13,000	None	PAR	1100
7	130	5	17,000	None	PAR	1200
8	310	10	13,000	None	PAR	2100
9	310	10	17,000	None	PAR	2200
10	130	20	15,000	Light	PAR	0010
11	130	20	15,000	Moderate	PAR	0020
12	130	20	13,000	Light	PAR	0110
13	130	20	13,000	Moderate	PAR	0120
14	130	5	13,000	Light	PAR	1110
15	130	5	13,000	Moderate	PAR	1120
16	310	10	13,000	Light	PAR	2110
17	310	10	13,000	Moderate	PAR	2120
18	310	10	17,000	Light	PAR	2210
19	310	10	17,000	Moderate	PAR	2220
20	130	20	15,000	None	ASR	0001
21	130	5	15,000	None	ASR	1001
22	310	10	13,000	None	ASR	2101
23	130	5	17,000	Light	ASR	1211
24	130	5	17,000	Moderate	ASR	1221
25	130	20	15,000	None	CSF	0002
26	130	5	13,000	Light	CSF	1112
27	310	10	17,000	Moderate	CSF	2222
28	130	20	15,000	None	MF	0003
29	130	5	15,000	None	MF	1003
30	310	10	15,000	None	MF	2003
31	130	20	13,000	None	MF	0103
32	130	20	17,000	None	MF	0203
33	130	20	13,000	Light	MF	0113
34	130	20	13,000	Moderate	MF	0123
35	130	20	17,000	Moderate	MF	0223
36	130	5	17,000	Moderate	MF	1223
37	310	10	17,000	Moderate	MF	2223
38	130	20	15,000	None	PP	0004
39	130	20	13,000	None	PP	0104

Table 1   2F90 AFTS GCA Course (Cont)

Level	Wind		Weight	Turbulence	GCA Type	Code
	Direction	Speed				
40	130	20	17,000	None	PP	0204
41	130	20	13,000	Light	PP	0114
42	130	20	13,000	Moderate	PP	0124
43	130	20	17,000	Moderate	PP	0224
44	130	5	17,000	Moderate	PP	1224
45	310	10	17,000	Moderate	PP	2224
46	130	20	15,000	None	PP&CSF	0005
47	130	5	15,000	None	PP&CSF	1005
48	310	10	15,000	None	PP&CSF	2005
49	130	5	13,000	None	PP&CSF	1105
50	130	5	17,000	None	PP&CSF	1205
51	310	10	17,000	None	PP&CSF	2205
52	130	5	13,000	Light	PP&CSF	1115
53	130	5	13,000	Moderate	PP&CSF	1125
54	130	5	17,000	Moderate	PP&CSF	1225
55	310	10	17,000	Moderate	PP&CSF	2225
56	130	20	15,000	None	PP&MF	0006
57	130	5	15,000	None	PP&MF	1006
58	310	10	15,000	None	PP&MF	2006
59	130	5	13,000	None	PP&MF	1106
60	130	5	17,000	None	PP&MF	1206
61	310	10	17,000	None	PP&MF	2206
62	130	5	13,000	Light	PP&MF	1116
63	130	5	13,000	Moderate	PP&MF	1126
64	130	5	17,000	Moderate	PP&MF	1226
65	310	10	17,000	Moderate	PP&MF	2226

LEGEND: PAR = Precision Approach Radar
ASR = Surveillance approach (Airport Surveillance Radar)
CSF = Compass system failure
MF = Minimum fuel
PP = Partial panel (No AN/AJB-3)

Table 2 Comparison of Automatic and Manual Modes

Function	Automatic Mode	Manual Mode
Student file creation	Manual entry	Manual entry
Student file retrieval	Automatic	Automatic
Level select	Automated-adaptive	Enter each pass
Initialization	Automatic	Automatic on level select
Vector to final	Automatic	Manual
Final controller	Automated	Manual
Performance measurement	Automatic	Automatic
Performance feedback	Automatic	Manual
Hardcopy of performance	Automatic	Automatic
Student file maintenance	Automatic	Automatic

to be normal or a minimum fuel approach. The initialization point for the normal GCA was on base leg, 5 miles from the turn to final and 9 miles from touchdown. Heading and altitude were initialized at 040 degrees and 2000 feet. It was assumed that the aircraft would be in a clean condition, although the aircraft configuration was not critical. The initialization point for the minimum fuel approach was 5 miles from touchdown and 4 miles offset on a heading of 310 degrees and an altitude of 2000 feet.

During both modes, the instructor was not permitted to view the AFTS Instructor's Terminal where the Tektronix Display is located, as this would have provided pictorial information for student feedback and evaluation which would not normally be available to him. At the end of the run, in both manual and automatic modes, the experimenter took a photograph of the display trace (Figure 3) using a Tektronix C-10 oscilloscope camera.

The Transfer Experiment

After completion of GCA training in the Device 2F90, the evaluation of the students who participated in the experiment was continued into the flight training phase in order to assess their performance in the operational situation (i.e., GCA field landing practice). An ATJ Addendum for GCA was filled out by the instructor pilot on a knee pad during flight in the TA-4J aircraft. The information contained in this form was similar to that which was on the ground trainer Tektronix display. This permitted a direct comparison of ground trainer performance with flight performance. In addition, voice recordings of the GCA controllers were recorded on cassettes for analysis and correlation with the ATJ Addendum. Identification of a particular student flight tape recording for transcription to a cassette was made by noting the time of flight and communications channel number. Tapes were held for ten days at the GCA radar shack and identified by date, time, channel, and frequency.

Performance Evaluation

To assess the training value of the GCA Module on Device 2F90, the performance of the students in the experimental and control groups were compared. Performance data on the simulator were extracted from three sources: (1) the computer printout (in both manual and automatic modes); (2) instructor scoring (in both manual and automatic modes); and (3) Tektronix display photographs. These data provided a basis for comparison of automated performance assessment with instructor subjective evaluation on the training device.

The use of instructor ratings as initial guidelines in the development of measures is valuable, and a great deal of insight to the measurement problem can be gained by examination of subjective ratings. However, subjective ratings lack reliable standards and rating-standardization. Another approach

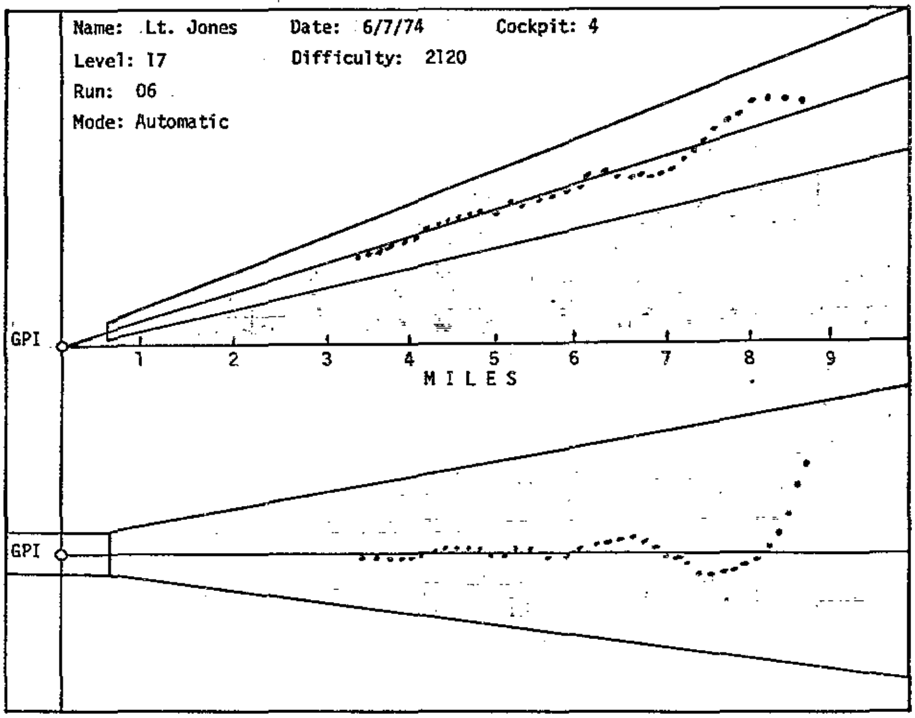

Figure 3. GCA Display Geometry

is to validate measures using within-subject sampling. With this approach, the assumption is that learning will be reflected by student advancement in difficulty level. Individual measures should show performance improvement through their trends from the first practice trial through the last training session for each student. If improvement is evident from the students' performance data, the validity of the measures will be substantiated despite any lack of subjective ratings.

Performance data of actual TA-4J GCA's of students in both the experimental and control groups were used to compare in-flight performance of both groups. In addition, this data provided a comparison of simulator performance with in-flight performance for the two groups.

A summary of the grading scheme is shown in Table 3. Note that the experimental and control groups were treated alike in methods of scoring. Automatic grading and instructor grading were both used to evaluate student performance.

In Figure 4, the relationship of the various data sources used in the study are shown.

Another evaluation device which was used after completion of GCA training was a questionnaire concerning the effectiveness of training with the automated GCA Module. The opinion of instructors and students who were involved in the study was solicited by means of this questionnaire.

Table 3 Grading Scheme

	SIMULATOR		AIRCRAFT	
	Automatic Grading	Instructor Grading	Instructor Grading	Controller Voice Tapes
Exper. Group	X	X	X	X
Control Group	X	X	X	X

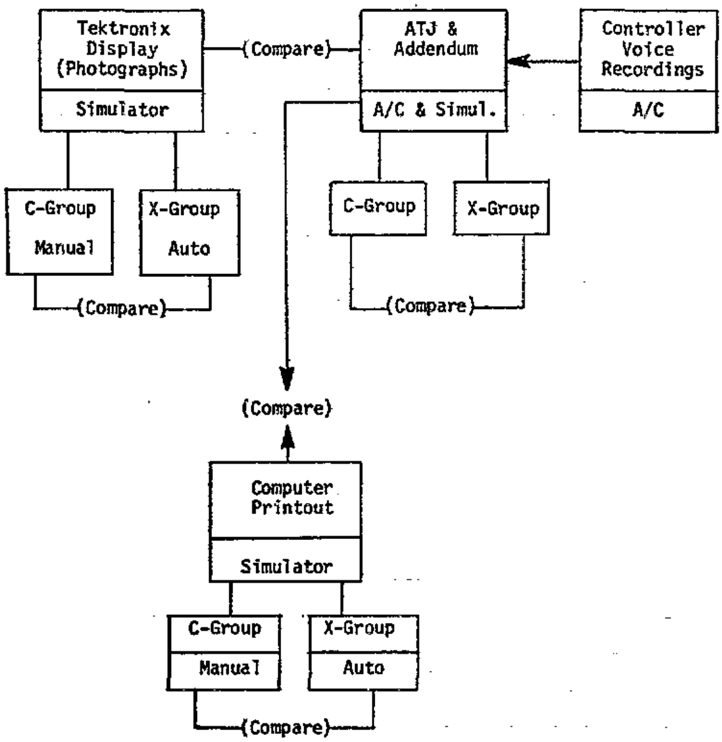

Figure 4. Use of Data Forms

RESULTS

In general, the hypotheses set forth at the beginning of the study were confirmed. Training with the GCA module was as effective as conventional training and, perhaps, more effective when considering the number of additional trials that were completed by the experimental group. However, comparison of performance between the experimental and control groups was not made against the entire GCA course but separately for different types of GCA's. The mean total scores for the experimental group were higher in every type of approach except the minimum fuel (MF) approach. However, no statistically significant differences between groups for ASR, CSF, and PP approaches, were found. For PAR approaches, experimental group scores were significantly higher, whereas, for the MF approaches, the control group scores were significantly higher. These were determined by signed rank tests for paired replicates.

The highest difficulty level reached was 52 in the PP & CSF type of GCA. Time limitations for each student in progressing through the curriculum did not permit more than nine students (five experimental and four control) to reach the PP & CSF level. A comparison of four students from each group did not show any significant differences in scores as determined by a T-test for independent samples; however, the confidence placed on such a small sample is low. See Table 4 for an overall comparison of these factors.

The number of runs completed for each type of approach by the experimental and control groups is shown in Table 5. A greater number of runs (88) was completed by the experimental group using the automated GCA module.

An opinion survey questionnaire was completed by seven students from the experimental group and ten instructors. The answers indicated that the majority of students and instructors appreciated the computer printout as a source of objective scoring data. They also liked the sound quality of

Table 4 Total Score Means for Groups

	PAR	ASR	CSF	MF	PP	PP & CSF
Exper.	178.9	175.7	171.1	95.4	147.2	63.0
Control	147.3	164.3	144.7	171.2	94.8	112.4
Diff.	31.6	11.4	26.4	75.8	52.4	49.4
P =	.01	N.S.	N.S.	.01	N.S.	N.S.

Table 5 Number of Simulator Runs

	PAR	ASR	CSF	MF	PP	PP & CSF	Total
Exper.	270	75	96	120	48	21	630
Control	247	61	44	99	75	16	542
Diff.	23	14	52	21	27	5	88
No. Subjs.	36	36	34	34	26	9	

the voice generation system. The least acceptable item was the course control provided by the computer-generated voice. Between these extremes were glide slope information, general information transmissions, and the value of performance scoring in relation to future student performance. Except for printouts and voice generation sound quality, the instructors seemed to be more generous than students in rating the device. The results may be seen graphically in Figure 5. The scores shown in this graph were adjusted for the unequal numbers of instructors and students.

One item which was not included in the questionnaire concerned the CRT display which provided a graphical representation of the simulated aircraft position at all times during the approach. This display was helpful to the instructor in following the course of a run and also as a debriefing aid at the end of a run; as such, it was very well accepted by the students and instructors.

Instructor grading on the simulator runs generally agreed with the automatic scoring. Automatic scoring, however, was much more critical than instructor grading in the majority of runs. As may be seen by reference to Table 6, most of the instructor grading fell into two categories: "average" (A) and "above average" (AA). There were no "unsatisfactories" (U), and only 2% "below average" (BA) for the control group and 6% "below average" for the experimental group. This resulted in the control students being advanced to the next difficulty level even if they had not satisfactorily completed the previous run. With the experimental group, the computer logic would not permit advancement to the next difficulty level unless a total score of 100 or more had been achieved.

There are fewer number of runs for both the control and experimental groups shown in Table 6 than are shown in Table 5. This is because the instructors did not grade all runs.

Instructor scoring of simulator runs indicates a significantly greater number of below average grades for the experimental group. This is also true of the GCA flight grades shown in Table 7. This trend is not supported by the automatic scoring, except in the minimum fuel approach. A possible explanation for this contradiction may be the unavoidable instructor-student contact in the experiment. A "blind" experiment, in which the instructor did not know whether the student was in the experimental or control group, was not practical in this study. As a result, the instructors might have graded students differentially as a function of their group membership. Despite this possibility, no unsatisfactory grades were received by any student in either group under simulated or actual flight GCA's.

Analysis of actual GCA controller voice recordings did not indicate any differences in performance between the experimental and control groups.

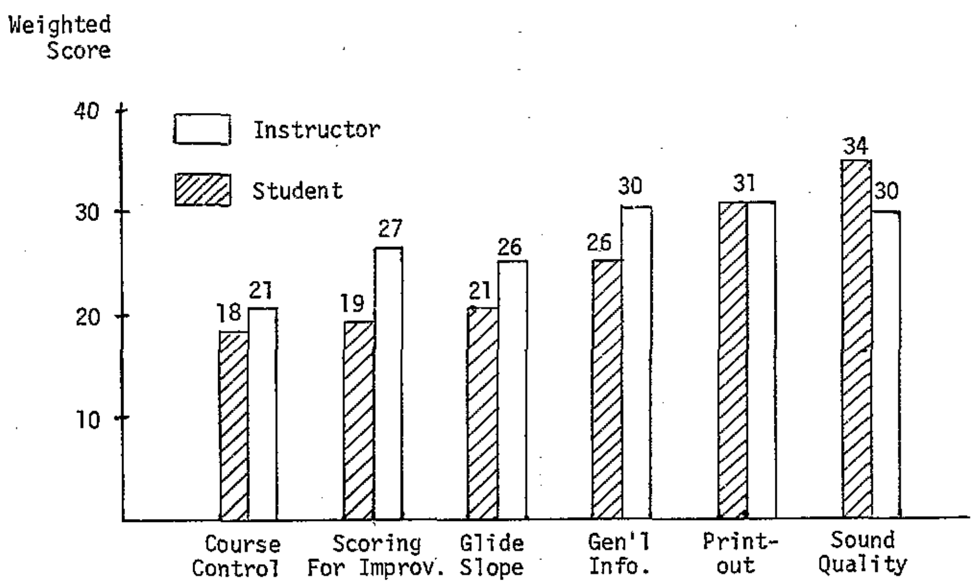

Figure 5. Results of Questionnaire

Table 6 Instructor Grading for Simulator Runs - Frequency Distribution and Percentages of Overall GCA Grades

GCA \ GRADE	CONTROL GROUP (Manual) U	BA N (%)	A N (%)	AA N (%)	TOTAL N	EXPERIMENTAL GROUP (Auto) U	BA N (%)	A N (%)	AA N (%)	TOTAL N
PAR		5 (2)	153 (63)	85 (35)	243		9 (4)	138 (61)	78 (35)	225
ASR			33 (56)	26 (44)	59		2 (3)	49 (73)	16 (24)	67
CSF		1 (2)	25 (61)	15 (37)	41		6 (7)	48 (59)	28 (34)	82
MF		1 (1)	60 (65)	32 (34)	93		8 (8)	82 (79)	14 (13)	104
PP		2 (3)	51 (69)	21 (28)	74		6(15)	24 (62)	9 (23)	39
PP&CSF		1 (8)	9 (69)	3 (23)	13		2(11)	10 (56)	6 (33)	18
TOTAL	0	10 (2)	331 (63)	182 (35)	523	0	33 (6)	351 (66)	151(28)	535

Table 7 Flight Performance Data - Frequency Distribution and Percentages of Overall GCA Flight Grades

GCA \ GRADE	CONTROL GROUP (Manual) U	BA N (%)	A N (%)	AA N (%)	TOTAL N	EXPERIMENTAL GROUP (Auto) U	BA N (%)	A N (%)	AA N (%)	TOTAL N
PAR		1 (2)	38 (65)	19 (33)	58		3 (6)	31 (60)	18 (35)	52
ASR		1 (11)	7 (78)	1 (11)	9		2 (29)	3 (43)	2 (29)	7
CSF			5 (71)	2 (29)	7		3 (33)	2 (22)	4 (44)	9
MF			3 (75)	1 (25)	4		1 (20)	1 (20)	3 (60)	5
PP			3 (50)	3 (50)	6		1 (8)	10 (77)	2 (15)	13
PP&CSF								1(100)		1
TOTAL	0	2 (2)	56 (67)	26 (31)	84	0	10 (12)	48 (55)	29 (33)	87

CONCLUSION

The feasibility of using the GCA module with an operational flight trainer in a training setting has been demonstrated by the current field evaluation. This study has attempted to establish the potential usefulness of an automated adaptive training system to relieve the instructor of minor duties which would interfere with his overall management of the training sessions. In addition to releasing the instructor for more important duties, the performance measuring capability of the system also provided him with a reliable means of objective, quantitative grading, and better aids for performance monitoring and student debriefing.

Some problems were noted in the current software programs. The GCA controller simulation was generally realistic but deficient in some areas which need modification. A great deal of improvement can be made by changing the heading algorithm, providing more glideslope information and rearranging some priorities in the sequence of verbal commands and information messages. Correction of these problems can produce a more effective system than the one used in the present evaluation.

This has been a modest beginning made by automating a small portion of a flight training syllabus. A more ambitious goal is to extend this type of automation to include in-flight maneuvers and air-to-air combat training at Naval flight training facilities. The technology is available under the basic computer program known as the Automated Flight Training System (AFTS). It is a realistic goal, therefore, which will probably be achievable in the very near future.

REFERENCES

Charles, J. P. and Johnson, R.M., Automated Training Evaluation (ATE). Technical Report NAVTRADEVCEN 70-C-0132-1, January 1972.

CNATRA INSTRUCTION 1542.20A, TA-4J Advanced Jet Flight Syllabus, 6 July 1973.

Logicon, Inc., Report: Automated Adaptive Flight Training System (AFTS) for TA-4J Operational Flight Trainer Device 2F90, Volume I, System Operation, Naval Training Equipment Center Contract N61339-73-C-0054, April 1974.

Puig, J. A., Johnson, R. M., and Charles, J. P., Evaluation of an Automated GCA Flight Training System, NTEC/Industry Conference Proceedings, 19-21 Nov 1974, NAVTRAEQUIPCEN IH-240.

ABOUT THE AUTHORS

MR. JOSEPH A. PUIG is a Research Psychologist in the Human Factors Laboratory at the Naval Training Equipment Center. He is currently engaged in the evaluation of training systems and studies in the application of human factors technology to simulation. He holds a B.A. and M.S. degree from New York University and an M.A. degree in Experimental Psychology from St. John's University. He is a member of the Institute of Electrical and Electronic Engineers.

MS. SUSAN GILL is employed as an Education Specialist for the Chief of Naval Air Training. She holds a B.A. degree in English and Education, and an M.A. degree in English from the University of West Florida, in Pensacola. She is currently working on a master's degree in Education from the University of Houston.