Journal of Aviation/Aerospace Education & Research

Volume 15
Number 3 *JAAER Spring 2006*

Article 1

Spring 2006

ATC Best Practices: An Opportunity for Computer-Based Scoring

Todd P. Hubbard
thubbard@ou.edu

Scholarly Commons Citation
Hubbard, T. P. (2006). ATC Best Practices: An Opportunity for Computer-Based Scoring. *Journal of Aviation/Aerospace Education & Research, 15*(3). https://doi.org/10.15394/jaaer.2006.1487

ATC BEST PRACTICES: AN OPPORTUNITY FOR COMPUTER-BASED SCORING

Todd P. Hubbard

Abstract

This paper explains how computer-based scoring of computer-generated scenarios in air traffic was designed and developed to meet the objectives of the Federal Aviation Administration (FAA). Radar Terminal Facility (RTF) instructors from the Air Traffic Division at the FAA Academy, other Certified Professional Controllers (CPC) from Miami and Chicago, and professional educators converged in September 2002 to design, develop, and implement a course that would enhance controller effectiveness by capturing controller best practices. These best practices were synthesized from a series of intensive air traffic control simulations in a fictional airspace. Programmers from the FAA Academy worked with subject matter experts to design and develop an objective means by which these simulations could be graded. The result was a modified version of an Academy invented program called SIGNAL. The newly enhanced software met the team's expectations and looked to have future applications in initial air traffic controller training. The success of SIGNAL has also had an effect on all pretest-posttest measures for other groups within the Academy.

ATC Best Practices: Assessing a Skill-Based Program

During 2000, Jane Garvey, the Federal Aviation Administration (FAA) Administrator, was fully focused on the growing number of runway incursions. As a means to solve the problem, Garvey helped the Runway Safety Program Office draft 10 initiatives, which included an interesting provision that required all education and training programs to create the means to evaluate each program's effectiveness (Federal Aviation Administration [FAA], 2000). The 10 initiatives were disseminated throughout the FAA and soon captured the interest of training managers and staff specialists who were in the best positions to arrest the runway incursion problem. One group chartered to quell the growing number of operational deviations was the Air Traffic Team Enhancement (ATTE) Steering Committee. The group's mix of educators, air traffic controllers, and air traffic managers focused their attention on team-building strategies as a means of allaying these operational deviations.

By summer's end, members of the ATTE steering committee had been asked to determine the means by which the ATTE program could assess its effectiveness. This task was more a plea for an evaluation instrument than the invention of yet another process. The committee solicited help from the FAA Academy's Air Traffic Division at the Mike Monroney Aeronautical Center in Oklahoma City, Oklahoma, since the Academy instructors had experience with assessing air traffic training programs (J. Cope, personal communication, July 8, 2000). The Academy, in turn, solicited help from adult education professionals, training specialists, and academics from the University of Oklahoma. In October, Hubbard (2000) issued a report to the FAA Academy in which the current forms of assessment of the ATTE program were appraised. In summary, the report indicated that the program's effectiveness was only slightly assessed through its use of self-reporting surveys. One could not directly attribute a reduction or gain in runway incursions to the information gained from ATTE workshops. Without a more defining instrument with which to measure the program's efficacy, the merits of the program remained a question. As Hubbard (2000) suggested, the reason why the ATTE workshops lacked efficacy was because they were designed to separately sample cognitive, affective, or psychomotor activity. Workshop objectives did not fuse the cognitive, affective, and psychomotor domains of learning. Since 2000, the ATTE program has addressed these issues.

The requirement to measure the effectiveness of

education and training has been an ongoing initiative within the FAA and a special interest of the Superintendent of the FAA Academy in Oklahoma City. In 1999, 2001, and 2003 the FAA Academy hosted the International Aviation Training Symposium (IATS) in Oklahoma City. Hundreds of civil aviation managers, trainers, and air traffic specialists from over 40 countries shared their experiences at each of the events, validating the FAA Administrator's belief that program assessment should be the chief interest of all training managers. Training evaluation and assessment were specifically targeted in the Embry-Riddle Aeronautical University paper on *Quality Management in Aviation Training* (Hisam, 1999), in the International Civil Aviation Organization's TRAINAIR paper on *Regional Training Planning: A Cooperative Approach to Meeting Shared Objectives* (Fox, 2001), and in EUROCONTROL's paper on *E-Learning: the EUROCONTROL Experience* (Drain, 2003).

Still under the influence of international interest in training assessment and the FAA Administrator's direction through the Runway Safety Program Office, the Superintendent of the FAA Academy challenged his staff to create an assessment model that would measure knowledge and skill abilities before course start and again after the course was completed. At nearly the same time, Air Traffic Service Director (AAT-1), Bill Peacock, announced that he wanted the FAA Academy Air Traffic Division to create an Air Traffic Controller (ATC) Best Practices program, designed to capture best practices from highly skilled controllers and teach these same best practices to all other controllers. Taken together, the mandates issued by the FAA Administrator, the Air Traffic Service Director, and the FAA Academy Superintendent put pressure on the Academy's Air Traffic Division to explore all opportunities and means to measure knowledge-based and skill-based learning.

The purpose of this paper is to expose and comment on the design, development, and implementation processes that the ATC Best Practices program staff employed to make this program a success. More specifically, this paper details the creation of a computer-based scoring system that proved to be the remedy for pretest-posttest assessment of skill-based learning and provided an objective means by which the Academy could isolate and measure best practices in Certified Professional Controllers (CPC).

Program Design and Development Phase

The technical approach for setting up the *Best Practices* program was guided, in part, by the *FAA Academy Guidelines for the Development, Delivery, and Evaluation of Training* (FAA Academy. March 1998). Modifications to the guidelines, to accommodate a more flexible development process, were presented and approved by Air Traffic Division management and Mr. Peacock. Foundational objectives had been issued from headquarters in two documents: Performance Measure Results Task 1 and Critical Work Activity 1. Within the documents were nested five, broad objectives: (1) develop a tool to measure effectiveness and impact of technical training, (2) research, develop, and prototype a process for collecting information, (3) establish necessary baseline data, (4) establish improvement goals, and (5) manage to reach goals.

Working in the regulatory milieu, managers at the FAA Academy discovered that the layers of requests for precise measures of effectiveness from various levels of supervision, pointed to the development of a unifying instrument. In the next few weeks, management would select a team, devise a strategy for rapid response, and begin the process of creating a means to measure effectiveness.

Selecting the Team

Responsibility for the program was given to Mike Morrison, the supervisor for the Terminal Radar Training section (AMA-512). Morrison, in turn, put together a team of professionals that included FAA Academy air traffic controller instructors (Terminal Option), subject matter experts from the Chicago and Miami Terminal Radar Approach Control (TRACON) facilities, training specialists from Washington's FAA Air Traffic Resource Management Program, and a contract instructional systems designer. The controllers from the terminal control facilities at Chicago and Miami were handpicked by the president of the National Air Traffic Controllers Association (NATCA), John Carr. Background data were gathered that provided enough information to complete a task and skills analysis, devise a training proposal, and set up the framework for a training development plan. Periodic meetings were scheduled over the next seven months, to give the team members time to design the program, develop a prototype scenario for testing, test the scenario, and make modifications for further testing. In addition, the FAA Academy provided computer programmers to support the building of a scoring management system within the radar simulators, to capture data during the scenario events.

Developing an Instrument

The team used a rapid prototyping technique to create the tools that would measure the effectiveness of technical training. Based on direction from the program's founder and principal supporter, Mr. Peacock, the program had to measure the three ingredients of air traffic events: safety, expeditious control of aircraft, and efficient use of

https://commons.erau.edu/jaaer/vol15/iss3/1
DOI: https://doi.org/10.15394/jaaer.2006.1487

airspace. Based on the time constraints imposed by FAA higher headquarters, the logical venue for the new program was the Radar Training Facility laboratory at the FAA Academy. The most logical and most accessible medium for the project was the Academy's radar simulator.

FAA Academy Radar Training Facility (RTF) instructors, along with the subject matter experts (SMEs) from Chicago and Miami, discussed how they would assess a controller's ability, while also capturing best practices. Human-based assessment presented too many reliability challenges. The only viable non-human alternative of any note was an idea to program the simulator computer to score specific items. This solution, although created by humans, would not involve humans as observers or evaluators.

Having selected the means by which controllers would be scored, the team set out to establish baseline data for controller capacity. As the team discovered, controller capacity had to be defined systematically. Controller capacity was simply defined as the sum of the aircraft landed, minus the result of operational errors, procedural errors, and inaccurate phraseology. Operational errors are defined as "an occurrence attributable to an element of the air traffic system in which less than the applicable separation minima results between two or more aircraft, or between an aircraft and terrain or obstacles" (Department of Transportation [DOT], 2002, p. 5-1). Five types of procedural error were tracked, they were transfer of communication errors (instructed to contact Tower before Final Approach Fix), intercept altitude errors (intercept the localizer above the glide slope), intercept angle errors (excessive angle), intercept point errors (aircraft intercepted the final approach inside the approach gate), and approach clearance altitude errors (aircraft not on a published route or segment of the approach being cleared for approach).

Based on the advice of the SMEs from Miami and Chicago, the computer programming staff and RTF instructors created an aggressive simulation of ATC Level 12 traffic volume. The scenario was tested repeatedly by the RTF instructors, before being given to the ATC Level 12 members of the team. Following numerous trials, the ATC Level 12 SMEs agreed that the scenario was equivalent to the traffic level they had intended. However, two items on the planning list still needed to be determined: (1) determine the duration of each scenario, and (2) determine the 100% capacity level of an ATC Level 12 controller. After further trials the team agreed on a 30-minute session length.

A 30-minute session time was chosen for three reasons: (1) 30 minutes provided enough time to measure safety, expediency, and efficiency at the 100% intensity level; (2) 30-minute sessions allowed the instructors to run more scenarios throughout a given training day; and (3) controllers working beyond peak performance begin to fatigue at about 30 minutes, normally followed by increased operational errors and deviations, which would not be tolerated in a real situation.

The team had defined controller capacity, but they had not manipulated controller capacity by subjecting controllers to air traffic intensity levels greater than and lesser than their peak capacity (100% capacity level). A first step toward documenting changes in controller capacity was to establish the perfect solution; and the simulator's computer provided the means. However, a perfect solution had to be based on the parameters given to human participants and those parameters had not been fully expressed. This led to another period of extensive trials, which in the end resulted in the validation of a series of air traffic intensity levels.

Establishing air traffic intensity levels for ATC Level 12 controllers proved to be difficult, because it was virtually impossible to account for all the variance. As the team began their trials, they set limits to the environment to help control variance. For example, all traffic would sequence through two points on the radar map, called posts, instead of four points on the map. This limited the scan pattern for new aircraft to just two areas. Further, intermediate altitude restrictions caused by airway corridors were eliminated. This eliminated confusion when giving vectors and giving instructions for descent into the radar pattern. All of the aircraft were sequenced to a single runway, as opposed to a multiple-runway operation. To help each controller focus on arriving traffic, all departures were eliminated. To ensure that all controllers would have an equal chance of succeeding, the team created a lesson that explained the airspace and created extra scenarios to help controllers practice the new procedures before being graded.

Even the radio calls from the aircraft were controlled. Pseudo pilots are commonly used to give the air traffic controller students practice communicating with pilots. These pseudo pilots receive training before assisting in the radar lab, but they can have off-peak-performance days. This problem with consistency was alleviated by pre-selecting pilots and then using the same pilots for all trials.

Having reduced the variance in the radar environment of the lab, the team refined their definition of the 100% solution. A scenario was considered to be a 100% solution if the intensity level of air traffic equaled the controller's capacity to safely control those aircraft for 30 minutes. This meant that during the 30-minute scenario, the controller could not commit any errors. Further, the controller had to sequence the maximum number aircraft to

the outer marker and had to efficiently use the airspace. The first 100% solution was set by the SMEs from Miami and Chicago. Confidence for their 100% solution was based on trials at the 110% intensity level.

Determining any degree of difficulty above the 100% level was made possible by first establishing normative behavior among ATC Level 12 controllers. Over a period of weeks, the Miami and Chicago consulting controllers had established a consistent pattern of controller behavior at their peak output level; which prompted the team to make a judgment about this level of consistency. However, the team was reluctant to generalize the behavior to all ATC Level 12 controllers; nonetheless, they predicted similar outcomes in future trials.

As a precaution, the team created a control for their experiments. The results of the computer-generated controlled trial are shown in Table 1.

Table 1. Computer-generated Score (prototypical phase)

Aircraft Landed	Operational Errors			Procedural Errors	Standard Excess Mileage	Total Score
	High	Mod.	Low			
21	0	0	0	0	8.77	321

The computer-based controlled trial excited interest in a computer-based scoring and data-gathering instrument. During the next two months the programmers modified the Academy's simulator support software program (SIGNAL), allowing the team to observe a perfect simulator session without the influence of a pseudo pilot or an air traffic controller. After several trials, the programming was perfected to the point that the team was willing to use SIGNAL as the new computer-based scoring device. Objective data, based on ratio scale accuracy, would eliminate human error and would increase the internal and external reliability of the program.

The three parameters that were measured by the computer were safety, expediency, and efficiency. The safety component of the computer score measured operational errors. Operational errors were measured by calculations of IFR separation loss, using a high, moderate, and low scale that had been derived from the FAA Order 7210.56C (dated August 15, 2002), *Air Traffic Quality Assurance*. The calculation tables found in Chapter 6 (*Severity Index*) of that order were used to assess points for loss of standard separation and loss of separation based on wake turbulence criteria. The sum of the points was used to calculate the severity of the error.

The computer also kept track of how many aircraft crossed the outer marker inbound for a final landing, which the team viewed as a measure of expediency. The latitude-longitude of the outer marker in relation to the position of each aircraft on final approach was used to determine when the aircraft was counted as having landed. If the aircraft crossed the outer marker, the computer would tally another landing, since the scenario did not involve landing clearance coordination between the Tower controller and the Final controller.

Efficient airspace control was determined by taking the average separation between aircraft on final approach—starting when an aircraft turned on to final and ending when the aircraft crossed the outer marker. All average and excess mileage distances were based on nose to tail separation between aircraft on final. All separation distances that exceeded the minimum required, were summed as total excess mileage. Low mileages at the end of a 30-minute session indicated more efficient control of the airspace.

On a computer monitor, positioned at an observing station used by instructors at the back of the simulation room, instructors could view, in real time, how each participant was doing in relation to safety, expediency, and efficiency. Each package consisted of a number of sheets of data, starting with a cover sheet that depicted total aircraft landed and the safety score, separated by type of error. Phraseology errors were also depicted on the cover sheet, and were the only subjective component of the score. The next few pages showed the raw data breakout of all the separation errors. This allowed each instructor to further analyze the reason for each error. An event listing that recorded event time, event name, aircraft identification, aircraft type, and spatial data immediately followed the error-scoring sheets. Average and total excess mileage statistics were shown on the last page of the package.

The control mechanism having been set and having implemented a computer-based scoring system, the team was in a position to test their prediction about ATC Level 12 controller behavior. To test their prediction the team devised a plan to validate the existing 100% capacity level and to reinforce that validation with comparable testing of

https://commons.erau.edu/jaaer/vol15/iss3/1
DOI: https://doi.org/10.15394/jaaer.2006.1487

controllers from ATC Levels 8-12 facilities. The first step toward standardization of behavior required the team to find an average score for each of the three categories. Very quickly, within the first 30 days of having a computer scoring system, the Miami and Chicago controllers set benchmark scores for the three parameters.

The second step toward standardization of behavior required the team to verify the average scores. They elected to incrementally increase the intensity level of the scenario problem, in hopes that they could measure the decline of controller capacity. This procedure was based on conclusions derived by Yerkes and Dodson in their study of arousal and task complexity. Their upside down U-shaped curve indicated that the quality of performance declined as the arousal state exceeded peak performance (Telfer & Biggs, 1988). The shape of the curve indicated that the decline was steady. Having been satisfied earlier by the incredible consistency of control of this initial group of participants, the team repeatedly pushed these same controllers past their peak, to measure the difference between 100% capacity and some predicted lesser capacity at traffic volumes higher than the standard. In similar fashion to Yerkes and Dodson, it was reasoned that as the intensity level of traffic was increased, the number of errors would also increase. Further, the team expected to see a breakdown in efficiency. Since safety, efficiency, and expediency translated to numerical scores, it was easy to measure incremental differences in capacity. Therefore, it was quite possible to set a declination scale of behavior that exceeded the 100% limit. Having completed the second step toward standardization of behavior, the team began a new course of testing.

Given the low number of participants in the initial test of the 100% solution, the team expanded the standardization process to groups of instructors containing ATC Level 12 controllers, as well as controllers from air traffic facilities with less intense traffic volume. The computer-directed trial scores and the Miami-Chicago scores presented the team with a starting point for further testing. Table 2 shows the mean scores for all the categories at the 100% intensity level, using 24 participants. Range of scores was 100-281, compared to the range of 200-281 for the Miami-Chicago trials.

Table 2. Prototypical Phase Data, 100% intensity level (n=24)

Average Aircraft Landed (expeditious)	Average Operational Errors (safety)			Average Procedural Errors	Average Standard Excess Mileage (efficiency)	Average Total Score
	High	Mod.	Low	Errors		
18.7		2.1	1	1.1	14.57	184
Miami-Chicago Trials						
19.5					13.8	232

Given the data from the additional trials, it was surmised that the range of combined scores and the averages for efficient and expeditious control were more telling components of standard performance. By comparing the Miami-Chicago results with these additional trials, the team developed a sense of what might happen in the first course conduct. That hypothesis suggested that at the 100% level, one could anticipate averages at 19 aircraft (expeditious), 14.2 miles (efficiency), and 208 points (combined score).

Testing was also performed at reduced intensity levels, such as 75%, 80%, and 90% (see Table 3). These alternative scenarios were the product of the completed Miami-Chicago tests and were validated by repeated trials by RTF instructors at the Academy (75%: n=37; 80%: n=34; 90%: n=28). Members of the team commonly accepted that certified controllers who were not familiar with the airspace supporting these scenarios would not perform well until becoming more familiar with the environment. Therefore, during the remaining trials and in the first course conduct, participants would spend at least 12 hours building speed and familiarization before running the final scenario at the 100% level. The intervening levels of intensity, with established practice time at each level, provided the team with a means to measure improvement; thus meeting the fourth objective of the Administrator's list.

Table 3. Prototypical Phase Data, Varied Intensity Levels (averaged data)

Intensity Level	Aircraft Landed (expeditious)	Operational Errors (safety)			Proc.	Standard Excess Mileage (efficiency)	Total Score
		High	Mod.	Low	Errors		
75%	16.05		1	.27	1.05	12.58	178
80%	17.70		1.05	.35	1.67	14.12	199
90%	17.89		1.42	.28	1.57	13.48	191

Periodically throughout the baseline data collection process the team refined the scenarios and programmed adjustments into the computational grading performed by the computer. The refinements and adjustments influenced the internal validity of the baseline scores to the extent that a separate trial was conducted in February, just before the first ATC Level 12 first course conduct (also referred to as operational tryout). Data were collected during this separate trial. These new data were compared to the initial baseline data and both sets of values were used as referents for the first ATC Level 12 operational tryout data.

Pretest-Posttest Assessment

Four individuals were chosen to participant in the pretest-posttest trial. Each participant completed the 100% pretest scenario, practiced at the 75%, 80% and 90% levels for two days, and then completed the 100% posttest scenario. The same intermediate intensity level scenarios were used for each participant. The data were recorded and analyzed (Tables 4 & 5).

Table 4. Pretest Assessment, Preliminary Tryout

Aircraft	Operational Errors			Proc.	Stand Excess	Total Score
	High	Moderate	Low	Errors		
19		1	0	0	15.9	220
19		6	2	0	17.37	110
17		6	1	0	16.7	63

Note: One of the participants did not complete the pretest.

Table 5. Posttest Assessment, Preliminary Tryout

Aircraft	Operational Errors			Proc.	Stand Excess	Total Score
	High	Moderate	Low	Errors		
17		0	0	0	17.38	263
18		0	0	0	18.18	281
15		2	1	0	21.01	130
15		1	0	0	20.05	190

Tables 4 and 5 indicate that the data were part of a preliminary tryout.

https://commons.erau.edu/jaaer/vol15/iss3/1
DOI: https://doi.org/10.15394/jaaer.2006.1487

The following comparison of total scores indicates pretest-posttest alignment by individual.

Participant 1	220/281
Participant 2	63/130
Participant 3	110/190
Participant 4	*/263

* indicates the absence of a pretest

The intervening period between the pretest and posttest was set at two days. Rather than being set by some scientific means, the intervening time of two days was chosen for purely administrative reasons, to help cut the cost of the program. If the course were contained within three days, each participant could travel on Monday and Friday, avoiding having to pay per diem for travel over a weekend. During the three days of the course the instructors planned to run an initial assessment (pretest) of controller skill at the 100% level, introduce best practices to CPCs, allow time to integrate the new skills, and then measure the improvement on a final skill test (posttest).

From the data collected during the operational trials and operational tryout, it appeared that the team had met the Administrator's first objective: *develop a tool to measure effectiveness and impact of technical training.*

Lesson Development Phase

Managing for Change

During the time that the air traffic controllers on the team were establishing the baseline data for the program, other members on the team were creating the means to pass on best practices to the participants who would complete the program after the operational tryouts. The program was planned for a one-hour introduction (administrative details), followed by a 100% intensity level scenario (performed by each participant), followed by a three-hour block where best practices were discussed. During the three-hour block, the instructor used an active 100% scenario to illustrate when best practices should be employed. After the instructional period, the participants were given a number of practice sessions at the 75%, 80% and 90% intensity levels. After each 30-minute session, each participant was debriefed for 30 minutes. Following the debriefing of the scenario, each participant completed another practice session at the same or higher intensity level. This process continued for four hours on day one and eight hours on day two. On the third day the participants practiced at the 100% intensity level, were given a 30-minute rest, and then were given the final, 100% intensity level scenario. Following the posttest, the scores were tallied and the best performer was announced. At the end of the program, each participant was given some time to critique the design and offer suggestions for improvement, to include suggesting additional best practices.

The initial set of best practices had been derived from observations of the expert controllers from Chicago and Miami terminal approach facilities. Since these experts were selected from all terminal option controllers, the team was reasonably sure that these individuals would be capable of isolating those techniques that made them better controllers, in terms of expeditious, safe, and efficient control of aircraft. Most of the initial best practices were variations of the controller's standard tools for separation: vectors, sequencing, and speed control. A list of these practices formed the basis for the scenario debriefing sheet (see Appendix, *Scenario Debriefing Sheet*, for a complete list of best practices).

As the team discussed how the program would evolve, they came up with a solution for long-term growth that would ensure continued success. The team envisioned using the top performers as adjunct instructors for future iterations of the best practices program. In this way, the program would be staffed by a collection of the best controllers, who would in turn pass on their achievements to the rest of the controller force. The team felt that this approach would fulfill the wishes of AAT-1 and might influence the controller population in such a way that air traffic control would be safer, more efficient, or more expeditious and it would meet the Administrator's fifth evaluation objective: *manage to reach goals.*

The team became concerned about the singular focus on ATC Level 12 facilities and the omission of input from levels of traffic below the ATC 12 level. To manage the differences between ATC Levels, the best practices design and development team planned to establish new baselines for each level. The process for establishing the baseline would be largely the same as was used in the prototype program. However, each baseline would be supported by scenarios that sampled the kind of traffic volume and the appropriate complexity level indicative of any specific ATC Level. As in the ATC Level 12 course, the top performers for each ATC Level course would be asked to join the adjunct staff for future classes of controllers.

Overcoming Acceptable Risks

There were a number of issues that the team worked through, but did not fully resolve during the design and development process. One threat to internal validity was the lack of a quantifiable set of behaviors that would define the average certified professional controller. The team had reviewed the knowledge and performance requirements for a terminal option, certified controller (Ammerman, Fairhurst, Hostetler, & Jones, 1989; Broach & Manning, 1997), but the literature review did not present the means to capture performance. The team subsequently chose to

sample knowledge, attitude, and performance simultaneously during the scenario by using the computer-based scoring system. The data set the boundaries for high and low performance.

Another known risk to the project was the pretest-posttest assessment strategy. It is commonly reported that a pretest can have an affect on the posttest, depending on the time interval between the tests (Wiersma, 2000). The time interval between tests was at least 12 hours, which in itself would suggest that the pretest would influence the posttest; but that was not the case in the best practices program. During the 12-hour interval each participant performed at least five other scenarios. The scenarios were based on the same airspace, but the amount of traffic and the rate of traffic were different for each scenario. Had the scenarios been similar, one would hazard a guess that the posttest was compromised. However, what the team found during the baseline trials and the operational tryouts was that the intervening sessions seemed to diminish the memory of the pretest.

The problem of learning regression did occur in nearly all the participants. Certified controllers from several terminal option facilities demonstrated a pronounced decrease in controller effectiveness after taking the pretest. The 100% intensity pretest shocked the participants to the degree that when given scenarios at 75% and 80% of peak intensity, the controllers did poorly. Compare the data in Table 7 with the data in Table 6. Notice that the effectiveness scores (aircraft landed) between the pretest and the intermediate intensity level scenarios has decreased. The safety (error rate) scores also showed an overall decrease from the pretest.

Table 6. ATC Level 12 First Course Conduct Pretest-Posttest Comparison

Aircraft	Operational Errors			Proc.	Stand Excess Mileage	Total Score	
	High	Moderate	Low	Errors			
Pretest							
19		1	0	0	12.15	250	Person 1
16		4	2	0	12.2	86	Person 2
19		2	3	0	12.89	140	Person 3
20		2	2	0	9.94	200	Person 4
Posttest							
19		0	0	0	15.08	300	Person 1
19		0	0	1	12.6	260	Person 2
18		0	0	2	14.72	206	Person 3
19		0	0	2	8.98	240	Person 4

Table 7. ATC Level 12 First Course Conduct (Intermediate intensity levels)

ATC 12 Op Tryout (intensity level)	Efficiency (excess mileage)	Effectiveness (aircraft landed)	Safety (number of errors)	Trials
75%	11.36	16.33	3.11	6
80%	14.91	16.5	2.99	6
90%	13.93	16.71	2.98	7
100%	12.99	18.33	4.46	27

https://commons.erau.edu/jaaer/vol15/iss3/1
DOI: https://doi.org/10.15394/jaaer.2006.1487

Discussion

In a regulated air traffic environment, new processes and procedures are often challenged by those who believe that nothing has changed or by those who lack imagination. During the *Best Practices* project, new instructional development processes were challenged, but the team countered with an imagined process that provided for simultaneous design and development. All of the team members were allowed to engage freely. Management kept the group centered by reminding everyone of the objectives, but management did not dictate how each person would perceive the project, nor did they limit each person's involvement. The members of the team that perceived intuitively *felt* the flow of the project and allowed the project to *talk back*. Those team members that perceived more concretely followed the established procedures set by regulation or precedent.

What had not been intentionally thought out before the project began was how people would relate to each other. However, from a crew resource management point of view, persons involved in the development of this project behaved in ways that exemplified the very best in effective communication, decision-making, risk management, workload management, and situational awareness. If any other group would attempt this project, they should select persons who already function well on teams. Perhaps a brief description of the team members would help illustrate this point.

The People

Objectives laid down by Mr. Peacock provided a central focus for the team. Having thus established the broader boundaries of the project, each team member concentrated on exploiting his or her own gifts. It is more fitting to describe personal contributions as gifts than to call them skills, at least in this case; because persons on the team did not solely function from their acknowledged skill set, but rather included unheralded skills that became visible as they were needed.

The manager of the project could be described as a gifted negotiator and coach. He used his gifts to help each team member see where he or she could contribute to the end goal. He also knew how to preserve the best from each member and showed each member how he or she could knit his or her effort into the fabric of the final product. The RTF instructors selected for this project demonstrated great self-control. Instead of fighting for control, each instructor channeled his or her energy into perfecting the scenario designs. This was particularly evident when the ATC Level 12 controllers from Miami and Chicago visited the Academy. Any personal bravado, heard during the trial period, was mitigated by over-powering professionalism.

In a project that premiered the art of the air traffic controller, one might think that anyone without those credentials would be overlooked or overwhelmed by those having the skill. However, this was not the case. Credit for the fusion of skills should be given to the manager of the team. He extended the effectiveness of the team by letting parts of the team work independently, convening the whole group to discuss new features in the design and development of the project.

Rapid Prototyping

Rapid prototyping proved to be a very useful process during the development of this project. It allowed sub-groups to work independently on specific elements of the design. It also invited experts to engage and disengage for finite periods of time, thus freeing them to teach in the classroom or to work on other projects after their job within the project was completed. Most importantly, it allowed the team to shift directions when parts of the design did not work well.

Objective Scoring

Perhaps the most beneficial outcome of the project was the creation of an objective scoring system for the air traffic control simulator. SIGNAL software provided the instructors a common point of reference for assessing safety, efficiency, and expediency. At the writing of this paper, the Academy is validating an enhanced version of SIGNAL to assess students in the initial air traffic control course.

Historical scoring routines for air traffic control lab sessions had relied on the experience and judgment of the instructor in assessing student behavior. To put it simply, hard graders graded hard and easy graders graded easy. The dichotomy in grading practices among instructors left the students at a disadvantage. Remedies for the perceived or real difference in grading standards had been attempted over the years, but those efforts had not always been successful. With the advent of Air Traffic Controller Best Practices, management finally found an objective means by which students or CPCs could be graded.

Pretest-Posttest Assessment

The computerized grading scheme that the programmers had created for *Best Practices* had a prospective effect on the initial air traffic controller training, while also providing a real time solution for the assessment of changes in learning during project development. Pretest-posttest schemes have, for decades, been successfully applied to assess a change in knowledge during air traffic control training. The interval scale data gained from the paper and pencil tests provided a precise measure of learning of all the course objectives.

Paper and pencil testing has worked well for knowledge-based learning, but has not proven effective in assessing a change in skilled performance, particularly skills that require the person to operate on and respond to a machine. Perhaps the reason why skill-based learning has resisted the pretest-posttest assessment routine, is because no one expects a person to have any pre-course skill. Why measure pre-course performance if the reason for the course

is to teach the skill? Paper and pencil testing is also the wrong medium for the assessment of controller skill.

Computer scoring, such as that accomplished by SIGNAL during *Best Practices*, is the remedy for skill-based assessment, or at least is a great start. Whereas pretest-posttest measures of skill were inadequate before *Best Practices*, they have since become adequate and relevant. Perhaps that success is partly due to the ratio scale data now available with SIGNAL, which by comparison diminishes the value of interval scale data obtained through the paper and pencil means. That success is also due in part to the context in which SIGNAL was used. The persons being scored in the validated *Best Practices* course were going to be fully qualified controllers who worked at ATC Level 12 facilities. They would be persons with a great deal of self-esteem and a great deal of proven talent in their skill. Had the scoring been any less objective—say with Academy RTF instructors—the participants would have cried foul and the whole program would have dissolved under the weight of bad press. One must keep in mind that these participants were also bargaining unit members. Therefore, it was imperative that the scoring be objective. However, this type of scoring has limited potential if used elsewhere for air traffic controller training. This is precisely why the Academy is undergoing a validation of an enhanced version of the original idea.

Before leaving this subject, it is useful to note that computer scoring does not diminish the need for subjective intervention by expert air traffic controllers. This issue was raised during the *Best Practices* trials. If a controller intentionally reduces the separation between aircraft on final to some distance below that required by FAA Order 7110.65, he or she is in violation of the Order and can be sanctioned for the safety error—in a real world situation. However, this apparently unsafe act is not always unsafe, even though it is in violation of the Order. It is a matter of control. Just as there are those who believe that a little compression (2.7 miles instead of 3 miles) cannot hurt if the situation on final approach was controlled, there are others who see no gray in the Order. This point has been so highly contended over the years that the FAA has published memoranda that allow some facilities to run aircraft closer than the established minimums.

The computer scoring system can measure distances between aircraft, but that system lacks the power to understand why or how. Expert controllers can determine why and how. Any grading system in the future will need to incorporate the human element.

Effects of Learning Regression

Computer scoring of each scenario allowed the team to track shifts in learning. For example, the team noted that most of the fully qualified controllers would perform worse on their first, 100% intensity level scenario. Using 321 points as the perfect computer score, the team saw scores in the 100s on the first, 100% scenario. It was common for these same individuals to remain in the 100s even when given scenarios at the 75% and 85% levels. However, after six or so additional scenarios these same participants began to show some improvement. By the end of the trial period, these controllers routinely scored in the mid to high 200s.

Based on the evidence of dozens of trials with Academy instructors, the *Best Practices* developers believed that any one person's attitude toward the *Best Practices* scenarios would be positive and welcoming. As a whole, the projected view was that all or most of the future participants would have a similar reaction to the training as did the Academy instructors. A representative from Chicago terminal radar control forecasted that ATC Level 12 controllers would have little trouble adapting to the scenarios and would do very well, perhaps even becoming bored at times. By the end of the first course conduct, thoughts that the project would improve controller skills, while giving a boost to the ego of each participant, were frustrated by an unforeseen reaction to perceived failure.

No one on the *Best Practices* team was prepared for the reactions of the fully certified ATC Level 12 controllers. The scores on the first 100% scenario (pretest) were 250, 86, 140, and 200. Scores in the 200s were not exceptional, but they were well within the range anticipated by the team. Scores below 200 were not new; but the team had anticipated a better showing, even for the pretest scenario. The reader will recall that the range of scores for the Miami-Chicago trials was 200-281, while the range for other participants during the trial period was 100-281.

Perhaps the most shocked were the team representatives from Miami and Chicago. They were profoundly surprised by the participants' failure to readily adapt to new airspace. SMEs from Chicago and Miami had insisted that ATC Level 12 controllers could control aircraft anywhere, any time. The team now knows that this announcement might have been made to identify ATC Level 12 controllers as superior to other controllers. However, this notion was not supported by the first group of participants. The participants were equally surprised by the results of their efforts. Perhaps they too had the same impression as the Miami and Chicago controllers.

Surprise led to self-doubt for some. On repeated trials, the poorer performers stumbled on the less intense scenarios. By the end of day two of the course, one participant was starting to show improvement, another was maintaining an average to good score on all trials, a third participant was trapped in continual poor performance, and the fourth participant was getting consistently better on each trial. Since controllers are largely ego-driven, the effect of having an ace performer and a poor performer was esteem-building to the good performer while devastating to the poor performer. Going into the third and final day of the course,

https://commons.erau.edu/jaaer/vol15/iss3/1
DOI: https://doi.org/10.15394/jaaer.2006.1487

the participants started to vocalize their feelings. One person said that he was embarrassed by his performance and could not believe he was doing so poorly. Another person was so embarrassed that he apologized to the instructors of the course and his fellow participants, even during the final critique.

Final Results

The pre and post scores, when grouped together, indicated that the course was a success (250/300, 86/260, 140/206, 200/240). The person with the large excursion—from 86 on the first 100% scenario to 260 on the final 100% scenario—showed the most improvement, which would not have been predicted by the poor showing during the intervening trials. His scores remained low up to the last few trials. It was clear that this person had some psychomotor breakthrough moment, where everything became clear. The individual starting with 200 and ending with 240 had performed consistently throughout the three days. Surprisingly, this individual had the lowest excess mileage score (8.98 miles). He landed as many aircraft as the person scoring 300, but he lost points with procedural errors. The highest scoring individual secured his top score by not making any errors, but had the most excess mileage of the group (15.08 miles).

ATC Best Practices was designed to enhance air traffic controller skills. The program measured three categories: safety, efficiency, and expediency. Safety scores improved when the controller reduced the number of errors per session. Expediency scores improved when the controller landed more aircraft during a session. And finally, efficiency scores improved when the controller reduced the excess mileage between aircraft on final approach for landing.

In most of the cases, the participants did not show marked improvement in all three categories. Table 8 shows what improved and what did not improve. The expediency scores, when comparing pretest and posttest results by individual, showed one gain, two losses, and one even set of scores. The safety scores showed a marked change; whereas all participants made moderate or low operational errors in the pretest, none of the participants made operational errors on the posttest. Efficiency of air traffic control from pretest to posttest for these individuals was counter-intuitive. One might think that there is a link between number of aircraft landed, excessive mileage, and total score, if operational errors were the same for all participants. For example, if a person were to land 20 aircraft during a 30-minute session, it might seem logical that high effectiveness scores would be associated with low excess mileage scores. This is not a proper analysis of these scores. On three posttest scores individuals landed 19 aircraft. None of the three had any operational errors. However, the excess mileage among the three was very different (19/15.08, 19/12.6, 19/8.98). If we compare these data with final scores, we see how troublesome it would be to predict a final result if the prediction were based on aircraft landed and excess mileage (19/15.08/300, 19/12.6/260, 19/8.98/240). Based on the data, it seems that procedural errors made the difference between a great score and an average score.

Table 8. ATC Level 12 Outcomes

Aircraft	Operational Errors			Proc.	Stand Excess Mileage	Total Score	
	High	Moderate	Low	Errors			
Pretest							
19		1	0	0	12.15	250	Person 1
16		4	2	0	12.2	86	Person 2
19		2	3	0	12.89	140	Person 3
20		2	2	0	9.94	200	Person 4
Posttest							
19		0	0	0	15.08	300	Person 1
19		0	0	1	12.6	260	Person 2
18		0	0	2	14.72	206	Person 3
19		0	0	2	8.98	240	Person 4

Conclusions and Recommendations

Air traffic controllers are highly competitive and take every opportunity to establish themselves as the best, in everything. From sizes of houses, to prices of cars, to longest period without an operational error or deviation, to handling the most traffic at an ATC Level 12 facility, controllers want to be the best. Despite each person's desire to be the best, the FAA has not had an objective means by which it could measure the best. *ATC Best Practices* was developed to do two things: (1) capture best practices from among air traffic controllers, and (2) enhance controller skills by helping each person use these best practices to improve their ability. Finding a means to measure the best occupied the time of more than a dozen individuals over the course of a year. When they finished, they had a decisive tool to measure air traffic control and they had a means by which they could capture best practices from among the FAA's best controllers.

Notwithstanding the need to meet Mr. Peacock's objectives for the program, it was also important to build a program that satisfied the five objectives of the Administrator's policy on training. Without any formula or other project to copy, the *ATC Best Practices* team created their own prototype, tested the prototype, validated the prototype, and used it to conduct their first course. They not only created the means to measure controller performance, they developed a strategy to capture best practices and use those best practices to enhance the rest of the controller force.

Despite the huge successes, there were casualties. The team had overlooked the effects of competition on fully certified, well experienced, controllers. To be fair, the team had not overlooked the effect of competition—that controllers thrived on it, but it had not included in that analysis any discussion of what would happen if a controller did poorly and reacted emotionally to that poor performance. While meeting all of the other goals and objectives, the team had ruled out any notion that controllers would expose a *thin skin*.

On the final day of the first course, one of the ATC Level12 controllers apologized for his poor performance. He publicly announced his feelings to his fellow participants, the instructors of the course, and the controller-observers. His apology was emotional, suggesting that he had been emotionally wounded. Several days after the participants had returned to their facilities, this same person e-mailed the instructors and apologized again. What had not been a prominent factor during the first course conduct had suddenly become a factor as the team analyzed course outcomes.

There are a number of recommendations for those wishing to assess best practices among air traffic controllers. First, it is important that each ATC facility level have its own *best practices* program. What works for ATC Level 12 might not work for ATC Levels below 12. Secondly, since there are differences between controllers from different ATC Level 12 facilities, more trials should be run on persons from more diverse ATC Level 12 facilities. With only three field controllers, one from Miami and two from Chicago, it was easy to tailor the course toward the influences of those two areas of control. Had the team expanded their focus beyond these two areas, they might have prepared a more representative scenario for a broader assortment of ATC Level 12 participants. Thirdly, it might have been a good idea to make more time for discussions about best practices. Time had been allotted for a brief discussion of best practices on the first day and some more time had been allotted for a brief discussion on the last day. However, this might not be enough time to capture tips to enhance controller effectiveness.

Lastly, any future effort in this area should address the effect of failure in a simulated environment. Persons are affected by failure in different ways. Some persons do not see any difference between simulation and reality. Others value the simulator and the real world equally. The FAA will need to evaluate its stand on this issue. Is the Academy that right place to call into question a person's ability, especially when that person's ability has been dully certified? There are mechanisms for this type of evaluation at the facility. The FAA Academy is not trying to decertify a controller, based on his or her performance in this course. Instead, the Academy wishes to enhance a controller's ability. If a controller's ability is not enhanced by this course, the course manager should be finding out why.

As is often true, too few studies have been done in this area. Computer scoring of radar control problems is relatively new at the FAA Academy and much more data needs to be collected before the instrument is perfected. Sample size during the first course conduct was too small to derive accurate results, despite the rigorous procedures used to establish a standard. More candidates will be needed before the results provide meaningful input.

Every effort is being made to further the research in computer-generated scoring at the Academy. It is clear that skill-based assessment can be adequately measured by objective grading practices such as SIGNAL, but there will always be a place for the subjective opinions of experienced controllers.

At the writing of this report on Best Practices the FAA staffers in Washington are planning to revive the

https://commons.erau.edu/jaaer/vol15/iss3/1
DOI: https://doi.org/10.15394/jaaer.2006.1487

program and complete studies to include facilities with traffic counts below the ATC Level 12 level (M. Morrison, personal communication, June 14, 2005).

Todd Hubbard Ed.D., Lt. Col. USAF (ret) has been a pilot since 1974, flying more than 3300 hours in the KC-135, T-37, T-38, U-2R, U-2CT, TR-1A, F-16, RF-4 and an assortment of general aviation aircraft. He has been the principal instructor for subject areas such as Crew Resource Management, Research Methods, Aviation Law, Flightdeck Environmental Issues, and Contemporary Topics in Aviation for Southeastern Oklahoma State University and the University of Oklahoma. He also provides Crew Resource Management instruction for FAA Aviation Safety Inspectors through a contract with the Regulatory Standards Division at the Mike Monroney Aeronautical Center in Oklahoma City. Since 1997 he has been an instructional systems designer on contract with the University of Oklahoma (OU), providing new and revised courses for air traffic controllers at the FAA Academy. The *Best Practices* program was only one of many assignments completed during his employment with OU. He was the first Editor-in-Chief and creator of the FAA's *International Journal of Applied Aviation Studies* and continues to support the journal as a consulting editor. In 2005 he was appointed as an Associate Professor at Oklahoma State University and holds the Clarence E. Page Endowed Chair. Among his many duties at OSU, he is the principal doctoral dissertation coach and principal research professor for the aviation and space education program. He has written scholarly papers about the challenges of inhibition training among Attention Deficit Hyperactivity Disorder individuals and ab initio student pilots, air traffic controllers' dimensional perception anomalies, gender identity issues in aviation research, potentially dangerous behaviors among pilots, and the root causes of dangerous decisions within aviation organizations. His is currently on contract to co-author a text on aviation mental health.

References

Ammerman, H. L., Fairhurst, W. S., Hostetler, C. M., & Jones, G. W. (1989). *FAA air traffic control task knowledge requirements, Volume II: TRACON controllers* (DTF-A01-85-Y-010304). Colorado Spring, CO: CTA Incorporated.

Broach, D., & Manning. C. A. (July, 1997). *Review of air traffic controller selection: An international perspective.* (DOT/FAA/AM-97/15). Oklahoma City, OK: Federal Aviation Administration Civil Aeromedical Institute.

Department of Transportation. (2002). *Air traffic quality assurance* (FAA Order 7210.56C, dated 8/15/02). Washington, D.C.: Author.

Drain, A. (2003). E-learning: The EUROCONTROL experience. *Proceedings of the International Aviation Training Symposium, FAA Academy, Oklahoma City, Oklahoma* [CD].

FAA Academy. (March, 1998). *FAA academy guidelines for the development, delivery, and evaluation of training.* Oklahoma City, Oklahoma: FAA Academy.

Federal Aviation Administration. (2000). *Ten initiatives for reducing runway incursions.* Washington, D.C.: Runway Safety Program Office.

Fox, M. (2001). Regional training planning: A cooperative approach to meeting shared objectives. *Proceedings of the International Aviation Training Symposium, FAA Academy, Oklahoma City, Oklahoma* [CD].

Hisam, T. (1999). Quality management in aviation training. *Proceedings of the International Aviation Training Symposium, FAA Academy, Oklahoma City, Oklahoma,* 231-240a.

Hubbard, T. (2000). *Air traffic team enhancement workgroup evaluation instrument.* Unpublished manuscript.

Telfer, R., & Biggs, J. (1988). *The psychology of flight training.* Ames, IA: Iowa State University Press.

Wiersma, W. (2000). *Research methods in education* (7th ed.). Needham Heights, MA: Allyn and Bacon.

https://commons.erau.edu/jaaer/vol15/iss3/1
DOI: https://doi.org/10.15394/jaaer.2006.1487

Appendix

ATC Best Practices **Scenario Debriefing Sheet**

Scenario (circle one): 75% 80% 90% 100%

Vectoring
- ○ Shotgun to base from NE gate Time _____ Heading_____
- ○ Flow to airport fix for downwind from NE gate
- ○ Flow to airport fix for downwind from NW gate
- ○ Fly over fix to left downwind (Runway 28)
- ○ Consistent downwind ground track
- ○ Double downwind
- ○ Vector from NE flow to NW flow
- ○ Dual downwind

Sequencing
- ○ NE flow to downwind with observable spacing pattern
- ○ NW flow to downwind with observable spacing pattern
- ○ Front-loading at beginning of push
- ○ Builds gaps for wake turbulence
- ○ Builds gaps for .65 spacing requirements
- ○ Consistent turns to base and final **Base turn mileage**_____
- ○ Straight-in sequence control **Heading off localizer**_____ See speed control.
- ○ Scan pattern (area, sweep, hub and spoke)
- ○ Tie points

Speed Control
- ○ Routinely slows aircraft at NE gate **Time**_____
- ○ Routinely slows aircraft at NW gate **250kts 210kts Time**_____
- ○ Routinely slows aircraft on downwind **250kts 210kts 190kts 170kts Other**_____
- ○ Routinely slows aircraft on final **210kts 190kts 170kts Other**_____
- ○ Slows individual aircraft as the need arises **A/C ID**_____ **Position**_____ **Time**_____
- ○ Straight-in speed control **Reduce to**_____**kts Increase to**_____**kts**

Priority of Duties
- ○ Established a recognizable pattern of control between gates, downwind, base turn, and final
- ○ Took early control of aircraft at NW gate to manage flow
- ○ Took early control of aircraft at NE gate to manage flow
- ○ Concentrated on base turns at a designated mileage **Mileage**_____
- ○ Concentrated on consistent downwind ground track
- ○ Concentrated on speed control at the gate or certain points

Errors _____ **Aircraft Landed** _____ **Total Score** _____

Acknowledgements

The *ATC Best Practices* project was a success because of collaboration among its creators. The project was well managed through the efforts of Ned Reese, Dick Pollock, and Mike Morrison. A special thanks to Mike Morrison for selecting Ken Sivley, Clint Windom, Kelly Fives, and Alan Zeleznik to be the RTF air traffic control instructors on the project. Their professionalism helped the team reach consensus. They also managed the dozens of practice trials performed by the tireless group of FAA Academy air traffic controllers, too numerous to mention here. Another group that aided the project immeasurably was the certified professional controllers from Chicago and Miami. Duane Semcken (Miami), Craig Arruda (Chicago), and Al Qualiardi (Chicago) set the limits of the scenarios, provided the initial list of best practices, and evaluated the first course conduct. James Garmon modified SIGNAL and created a useful computer-based scoring device. His resourcefulness and skill solved the scoring dilemma and provided the team with an excellent performance instrument. Lisa Leckrone, the team's Washington staffer, handled all the negotiations for field support; controlled the budget for the project, and ensured that the project was not slowed by lack of support from Washington or the FAA Regions. Todd Hubbard set up the data collection methodology, analyzed the data, kept the meeting minutes, and provided educational advice when necessary.

https://commons.erau.edu/jaaer/vol15/iss3/1
DOI: https://doi.org/10.15394/jaaer.2006.1487