Initial Evaluation of the Operational Assessment Program

Cristina L. Byrne
Linda G. Pierce
Melissa S. Beben

Civil Aerospace Medical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

January 2017

Final Report

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents thereof.

_____

This publication and all Office of Aerospace Medicine technical reports are available in full-text from the Civil Aerospace Medical Institute's publications website: http://www.faa.gov/go/oamtechreports

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-17/3	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Initial Evaluation of the Operational Assessment Program		5. Report Date January 2017
		6. Performing Organization Code
7. Author(s) Byrne CL, Pierce LG, Beben MS		8. Performing Organization Report No.
9. Performing Organization Name and Address FAA Civil Aerospace Medical Institute P.O. Box 25082 Oklahoma City, OK 73125		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency name and Address Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591		13. Type of Report and Period Covered
		14. Sponsoring Agency Code

15. Supplemental Notes

Work was accomplished under approved task AM-B-HRR-523

16. Abstract

The Operational Assessment Program (OAP) was developed to help Level 10 and above terminal radar approach control (TRACON) facilities promote Certified Professional Controllers In-Training into their facilities. The OAP consists of two hurdles: a job knowledge test and a skills assessment. The OAP was validated using a content validation approach (AIR, 2011) and meets the Equal Employment Opportunity Commission standards for use as a promotion/selection tool (EEOC, 1978). However, the content validation approach does not provide the Federal Aviation Administration with information on the effectiveness of the processes used to implement the OAP and the reactions of TRACON personnel who participated in executing the programs, both of which are necessary to identifying necessary improvements in future implementation cycles. This study collected data from critical facility personnel involved in the implementation of the OAP at their facilities to determine perceptions of the OAP and its component processes. The results indicate that the program was perceived to be implemented in a fair and effective manner, but several important changes and improvements were recommended by respondents. These results and participants' recommendations are presented and discussed in detail in this report.

17. Key Words Operational Assessment Program, CPC-IT, TRACON, Transfer, Promotion	18. Distribution Statement Document is available to the public through the Internet: http://www.faa.gov/go/oamtechreports/		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 24	22. Price

Form DOT F 1700.7 (8-72)   Reproduction of completed page authorized

ACKNOWLEDGMENTS

The opinions expressed are those of the authors alone, and do not necessarily reflect those of the Federal Aviation Administration, the Department of Transportation, or Federal government of the United States.

Please address questions or comments to Linda G. Pierce, NAS Human Factors Safety Research Laboratory (AAM-520), P.O. Box 25082, Oklahoma City, OK 73125. email: linda.pierce@faa.gov

Contents

INITIAL EVALUATION OF THE OPERATIONAL ASSESSMENT PROGRAM

INITIAL EVALUATION OF THE OPERATIONAL ASSESSMENT PROGRAM

INTRODUCTION

The FAA's Air Traffic Organization (ATO) developed the Operational Assessment Program (OAP) in order to increase the certification rates of Certified Professional Controllers-In Training (CPC-ITs) at Level 10 or higher terminal radar approach control (TRACON) facilities. Although previously certified, CPC-ITs must re-enter training and certify at the new facility. Overall, 80–90% of trainees at TRACON facilities succeed in training. However, at high level TRACONs the success rate for both CPC-ITs and new hires can be as low as 50%. These numbers are somewhat difficult to pin down due to the dynamic nature of these data and trainees continuously moving in and out of facility training programs (i.e., there is always a portion of trainees that are currently in progress which can impact exact success rate percentages at any given time). Though not exact and always changing, these numbers do seem to indicate and support the purpose behind the development of the OAP: that these extremely busy, high level TRACONs need a better system for selecting transfers. A substantial number of CPC-ITs in these higher level facilities are unable to become certified at their new facility (i.e., after training, they don't demonstrate adequate controller skills to receive position signoffs for certification at the new facility and thus cannot be authorized to independently control air traffic). These training failures and relatively low certification rates as compared to other facilities increase the burden on the employee, facility, and the ATO.

The FAA's Technical Requirements and Forecasting Group (formerly the Management Services Technical Workforce Services Team), with the contract support of industrial and organizational psychologists (I/O psychologists) from American Institutes for Research (AIR), along with air traffic control (ATC) subject matter experts (SMEs), were responsible for developing the OAP in accordance with Equal Employment Opportunity Commission (EEOC) standards (EEOC, 1978), as well as Federal and FAA human resource policies. The SMEs represented the three pilot test locations (Atlanta, Chicago, and Southern California TRACONs), as well as other large TRACON facilities. Researchers from the Civil Aerospace Medical Institute (CAMI) were enlisted to assess perceptions of fairness and effectiveness of the OAP among TRACON personnel who participated in the implementation of this new transfer/promotion process.

The SMEs provided guidance and input throughout each phase of development. Union and management perspectives were represented on the SME panels from each of the three key sites involved. Their work resulted in the decision to design the OAP using two methods of assessment: a job knowledge test and a scenario-based skills assessment. This multiple-hurdle approach was selected to assure that applicants would have the necessary background radar knowledge before continuing on to the more resource-demanding skills assessment.

OAP DEVELOPMENT

The OAP was evaluated using a content validation strategy (AIR, 2011) based on two prior TRACON air traffic controller (ATC) job analyses (Alexander, Ammerman, Fairhurst, Hostetler, & Jones, 1989; Krokos, Baker, Norris, & Smith, 2007). The requisite knowledge requirements were extracted from these job analyses and used as the basis for developing the content areas used for the exam. Approximately 400 exam items taken from basic-level radar courses at the FAA Academy in Oklahoma City were used to

create an initial item pool for the development of the knowledge test. Several SME panels reviewed and revised these items with several iterations resulting in 291 items classified into 27 knowledge content areas. The 291 items were then separated into two forms with 69 items appearing on both tests. The two forms were pilot-tested using 211 experienced TRACON controllers. This process provided the data needed to set the standards and cut scores for each exam. The knowledge exams were found to be equivalent in terms of both content area representation and difficulty. The passing score was determined during a standard setting meeting attended by air traffic SMEs and I/O psychologists from AIR. The resulting knowledge exams were administered through the FAA's electronic Learning Management System (eLMS). Selecting officials at facilities using the OAP received the results of these exams, notified applicants of their results, and scheduled passing applicants for the skills assessments portion of the OAP (AIR, 2011).

Air traffic SMEs from the three key sites and I/O psychologists from AIR worked together to develop the OAP scenario-based skills assessment. Thirty-five SMEs were interviewed regarding the 43 air traffic control subactivities described in prior job analyses. These interviews and workshops resulted in the development of five performance dimensions identified as critical to successful CPC performance. These performance dimensions are: Communication, Working Speed, Planning and Projection, Prioritization, and Speed Control (AIR, 2011). Six-point behaviorally anchored rating scales were developed to provide benchmarks for use in assessing each of the five dimensions. SMEs that were currently certified at Level 12 TRACON facilities helped to design the procedures used during the skills assessment. Furthermore, SMEs participated in the development of the behaviorally-anchored rating scales used to assess applicant performance on the scenarios.

Frame-of-Reference training was conducted at the three OAP pilot facilities to familiarize raters with the performance dimensions and ensured an acceptable level of agreement between raters. During training, pseudo applicants (e.g., current trainees or other facility personnel) completed portions of the scenario-based skills assessment, while raters observed their performance and practiced making ratings. After observing and making their initial ratings, raters participated in facilitated group discussions to help improve agreement and inter-rater reliability, as well as to help reduce potential bias, errors, and prejudice inherent in any rating process.

Following completion of rater training, all three facilities began using the OAP to make actual transfer/promotion decisions. Two facilities completed one staffing cycle using the OAP, and the third facility had an open and continuous bid over several cycles before administering the surveys used in this evaluation.

EVALUATION

The purpose of this evaluation was to assess participant perceptions of the Operational Assessment Program components to identify potential improvements to the process. This evaluation was conducted after implementation of the OAP at the Chicago TRACON (C90), Southern California TRACON (SCT), and Atlanta TRACON (A80). An interview/survey approach was used to gather information from those closely involved in OAP implementation. Three types of surveys were developed for use in interviewing respondents based on their role in OAP implementation. The first survey was designed to capture the perspective of those trained to rate applicants for CPC-IT positions during a scenario-based skills assessment. Respondents included OAP raters, as well as FAA training personnel. The second survey was de-

signed for use with facility personnel charged with making the CPC-IT selection decisions. Both the "rater" and "selecting official" surveys addressed all aspects of OAP implementation. The third survey addressed the OAP more generally and was used to gather information from facility personnel who either performed a more tangential role in the implementation of the OAP or had substantially less time available for an interview. Respondents to the "general" survey included, for example, air traffic managers (if not the selecting official) and union facility representatives. The general survey was a shortened version of the other two surveys and asked for general, broad-based opinions of the OAP.

Each survey contained four basic question formats. The first type of question asked the respondents to provide a rating on a 1 to 7 scale, where 1 indicated "not at all" and 7 indicated "to a great extent."

The second type of question asked the respondents to give "yes" or "no" responses. Respondents were allowed to provide comments or qualifying statements to both of these types of questions. The third type of question was open-ended only. The fourth type of question asked the respondent to give an exact number (e.g., how many practice scenarios did applicants complete?). The data provided in the tables throughout this report include the mean (or average) of all scaled questions, the standard deviation (or variation) around the mean, the frequency counts of yes/no type responses, and the mean of all exact number questions. Follow-up or open-ended comments are summarized in the text following each table. These comments are broken down by survey type (i.e., raters, selecting officials, or general). Additionally, this report is organized by content areas addressed in the surveys. In the tables, the column labeled "N" indicates the number of people that answered the question. Sometimes this number is fewer than the number of people that were asked the question, because, in some cases, the respondent did not know or could not remember the information requested. In other cases, the question was not applicable, or the respondent did not feel they had the relevant information to answer.

Participants

A total of 29 interviews were conducted at C90, SCT, and A80. Eighteen of the respondents had been trained to be OAP raters, three were selecting officials, three were FAA training personnel, and five had general knowledge of OAP implementation (three facility union representatives, one FAA training manager, and one facility manager).

Rater Survey. There were a total of 21 respondents to the rater survey. Six raters from each facility were interviewed. Additionally, one member of the C90 training department and two members of the SCT training department were interviewed using this survey.

Selecting Official Survey. The selecting official at each facility was interviewed. Two selecting officials were district managers and one was a staff manager.

General Survey. The Union Facility Representatives (FACREPs) were interviewed at all three facilities. In addition, one member of a training department and one facility manager were interviewed using the general survey.

Demographic Information

Information was gathered on the background of the raters, including how long they had worked and been a CPC at their current facility, and whether they were a supervisor or bargaining unit employee. Selecting officials were queried about the staffing process, including certification rates for CPC-ITs prior to implementation of the OAP, their expectations for CPC-IT certification rates using the OAP, and if they

had used any other processes to eliminate applicants before participation in the OAP. They were also asked how many controllers they intended to select for the year and what sources, in addition to the OAP, had been or were currently being used to staff ATCSs for their facilities. Additionally, all respondents were asked by what process raters had been selected to participate in the OAP (Table 1).

Table 1. Demographics

		M	SD	Yes	No	N
1	What was your developmental CPC-IT certification rate before starting the OAP? **(%)**	49.67	21.46			3
2	What do you expect your CPC-IT certification rate to be using the OAP? **(%)**	78.33	2.89			3
3	How many people do you intend to hire using the OAP in 2012?	20.33	16.44			3
4	Did you use any process to eliminate applicants before participation in the OAP?			3	0	3
4a	How well do you think it worked for limiting applicant pool size?	7.00	0.00			3
5	How long have you been at your current facility? **(years)**	12.24	7.39			21
6	How long did it take you to reach CPC at this facility? **(months)**	10.45	6.37			21
7	Are you a bargaining unit employee or are you a supervisor?			13 (BUE)	8 (S)	21

Note. Questions 1–4a were asked of three selecting officials; Questions 5–7 were asked of 18 raters and three training personnel.

Rater Survey. Most raters were also trainers and indicated that they had volunteered to participate in the OAP. At SCT, at least one rater was selected to represent each geographical region within the TRACON. The number of years the raters had been at the facility varied between 3 and 31. Time to reach CPC at the facility varied from six months to 1.5 years.

Selecting Official Survey. Certification rates for new hires prior to implementation of the OAP had been less than 50% at both C90 and SCT for all staffing and about 60% at A80. The SCT certification rate for CPC-ITs had risen to approximately 64% due to more reliance on Employee Relocation Requests (ERRs) instead of general public hires. All selecting officials expected that using the OAP would improve certification rates of CPC-ITs to about 80%. The OAP applicant pool was restricted by requiring applicants to have radar experience. Going forward, the plan is for C90 to use the OAP only to staff CPC-ITs , while SCT and A80 plans to staff using both the OAP and CPC-IT ERRs, and no one expects to receive new hires in the near term. SCT will also staff properly vetted VRAs (i.e., former military controllers hired under the Veterans Recruitment Appointment). At the time these surveys and interviews were administered, C90 expected to staff 39 controllers, SCT 20 controllers, and A80 eight controllers. Selecting

officials indicated that selecting raters was a collaborative process between the training department and the union.

General Survey. When asked how raters had been selected to participate in the OAP, respondents to the general survey said that, for the most part, raters were selected for their experience with training or the training lab simulation capabilities. These respondents also indicated that solicitation of rater volunteers was done through collaboration between the union and management.

Knowledge Exam

Taking and passing the knowledge exam was a precondition for participating in the OAP skills assessment. Raters were only asked if they knew how each applicant scored on the knowledge exam. Selecting officials were asked several questions regarding how the knowledge exam scores were reported to them, as well as their opinion of the importance of the knowledge exam in the OAP (Table 2).

Table 2. Knowledge Exam

		M(1–7)	SD	Yes	No	N
8	How well do you think the OAP knowledge exam allows applicants to demonstrate the knowledge you believe is needed to work as a controller at your TRACON?	3.67	2.52			3
9	How important were the knowledge exam scores in selecting applicants to participate in the skills assessment?	2.33	2.31			3
10	Did you personally know how each applicant scored on the knowledge exam before they participated in the skills assessment?			2	19	21

Note. Questions 8–9 were asked of three selecting officials; Question 10 was asked of 18 raters and three training personnel.

Rater Survey. No raters at the facilities knew the actual knowledge exam scores prior to assessing applicants. They all knew, however, that OAP applicants attending the skills assessment had passed the knowledge exam, thus, the two "yes" endorsements.

Selecting Official Survey. Similarly, selecting officials were only told if an applicant did or did not pass the knowledge exam. The selecting officials stated that so few people failed the knowledge exam that they did not see it as a useful component of the OAP. They suggested either making the test more difficult, including questions that go beyond basic radar knowledge, or stopping the use of the exam altogether. One official also suggested that it only be used if an applicant did not have previous radar experience, but that use would require eliminating radar as a qualifying requirement.

Skills Assessment

Applicants passing the knowledge exam were invited to participate in the second phase of the OAP. The skills assessment was held at the facilities and was preceded by classroom training and practice scenarios to familiarize the applicants with procedures used during the skills assessment and the local simulation equipment. The skills assessment included completing three graded scenarios using the five per-

formance dimensions. The scenarios were developed by the individual facilities based on guidance provided in the notice (Notice # JO 3330.70, Operational Assessment Program, see Appendix A) and by the Technical Workforce Services Team (the original notice is provided as an appendix in this report, but there is also a revised version Notice # JO 3330.71 that is accessible via www.FAA.gov under notices). Raters and selecting officials were asked a series of questions about how the scenarios for the skills assessment were developed and conducted, how well applicants were prepared for the skills assessment, and the perceived utility of the skills assessment in choosing applicants that would likely succeed (Table 3).

Table 3. Skills Assessment

		M(1–7)	SD	Yes	No	N
11	How well did the scenarios capture the skills needed to work as a controller at your TRACON?	6.66	0.50			23
12	How difficult was it for your employees (you) to develop the scenarios for the skills assessment following the OAP guidelines?	2.71	1.41			14
13	Would you change any of the guidelines?			0	3	3
14	What amount of effort was required to adequately develop the scenarios?	5.67	1.53			3
15	How important are customized scenarios as a feature of the OAP?	6.00	1.73			3
16	Do you think the use of generic airspace in the OAP would be appropriate for assessing applicants for specific facilities?			2	1	3
17	Were you involved in the development of the scenarios used in the skills assessment?			8	13	21
18	Approximately how many hours of familiarization did you provide the applicants before the practice scenarios? (**hours**)	5.74	3.54			17
19	How many practice scenarios did you allow applicants to complete? (**number**)	5.80	0.95			20
20	How well do you believe the practice scenarios at the facility prepared the applicants for the skills assessment?	6.62	0.67			21
21	How well did the OAP skills assessment allow the applicants to demonstrate their ability to work as a controller in your facility?	6.90	0.30			21

Note. Questions 11–12 were asked of three selecting officials, 18 raters, and three training personnel; Questions 13–16 were asked of three selecting officials; Questions 17–21 were asked of 18 raters and three training personnel.

Rater Survey. Most of the raters surveyed at C90 and A80 had not participated in the development of the scenarios, while most of the raters surveyed at SCT did participate, to some extent, in scenario development. Greater participation at SCT was to ensure that specific challenges within each geographical region in the TRACON were represented in the scenarios. Training personnel interviewed at all facilities had been involved in scenario development and stated that the scenarios were not particularly difficult to build within the guidelines prescribed in the notice. However, some SCT raters reported that it was difficult to comply with the Notice JO 3330.70 guidelines when attempting to represent highly variable terrain and to include visual flight rules traffic. In the end, all raters and training personnel reported that the scenarios captured the skills needed to work as a controller in that facility. In addition, most raters and training personnel reported that the six practice scenarios prepared applicants fairly well for the skills assessment and that the skills assessment allowed applicants to demonstrate their ability to work as a controller. One concern expressed was the need to build more complexity into one or two of the practice scenarios to ensure that applicants had an opportunity to practice all types of graded scenarios.

Selecting Official Survey. The selecting officials reported that the scenarios were not difficult to develop following the OAP guidelines and that the guidelines should not be changed. They did report, however, that the effort required to create the scenarios was fairly high in that a significant amount of time and number of personnel had to be allocated to the task.

The way in which the scores on the graded scenarios was reported to and used by the selecting officials varied. At one facility, raters met with the selecting official to discuss the skills assessment, applicant performance, and their recommendation to either select or not select each applicant. The scores on the graded scenarios were provided to the selecting official only after the selections were made to check for consistency with the staffing decisions. At the other two facilities, the training department representative provided a spreadsheet containing information on applicant performance to the selecting official. The selecting officials used performance averages for selection in a top-down manner, with one selecting official receiving a recommended soft cut score from the raters as a general suggestion for who to hire (i.e., a score the raters believed was indicative of high performance, but they would not necessarily exclude applicants that scored just below this score).

The selecting officials preferred that the assessments be conducted at their respective facilities, rather than at a centralized location, so that the facility personnel could meet with the applicant and the applicant could evaluate their own fit with the facility. However, one selecting official was open to using a centralized location, as long as raters would be sent from each facility to help rate applicants. The selecting officials also differed on the extent to which they believed generic airspace could or should be used for assessments. Two selecting officials believed that generic airspace could be used as long as the critical issues of the facilities were addressed by the scenarios (e.g., focus on arrivals). The other selecting official, however, believed it was important for the scenarios to capture the unique aspects of each facility by using actual facility-specific airspace rather than generic airspace.

Familiarization Packets

Information prescribed in Notice JO 3330.70 was prepared by the training department at the facilities and sent to each of the OAP applicants to familiarize them with facility-specific airspace before their arrival. The packet also contained descriptive information about the facility and housing opportunities in the area. Respondents were asked to rate the extent to which they perceived that the information was, in fact, reviewed by the applicants prior to the skills assessment, and whether or not the familiarization packets

provided adequate preparation for the skills assessment. Selecting officials were asked primarily about the amount of resources and time that were required to develop the familiarization packets, as well as any feedback that they received from applicants (Table 4).

Table 4. Familiarization Packets

		M(1–7)	SD	Yes	No	N
22	What amount of effort or resources was put into developing the familiarization packets?	6.00	1.73			3
23	Did you receive any feedback from applicants on the familiarization packets?			1	2	3
24	Did you include any additional information other than what was required in the checklist of familiarization packet contents?			8	6	14
25	What proportion of applicants do you believe reviewed the familiarization packet before they arrived? **(1-none to 7-all)**	5.25	1.52			20
26	How well do you think applicants reviewed the familiarization packet before they arrived?	5.36	1.61			21
27	How helpful do you think the familiarization packets were to the applicants in preparing them for the process?	6.37	0.83			19
28	Was there any information that should be included in the familiarization packets in the future?			4	14	18

Note. Questions 22–23 were asked of three selecting officials; Question 24 was asked of three selecting officials, 18 raters, and three training personnel; Questions 25–28 were asked of 18 raters and three training personnel.

Rater Survey. The raters and training personnel thought that most of the applicants had reviewed the familiarization packets and that the information was useful in preparing applicants for the skills assessment. Most believed that the information packets were fairly complete. However, several mentioned that some additional clarification on facility-specific information might be useful to applicants. There was also a suggestion that a demonstration of how to achieve the traffic pattern could (not necessarily should) be included in the packet to better inform those applicants who had limited experience with airport traffic patterns.

Selecting Official Survey. The selecting officials reported that a lot of effort was expended by the training department to prepare and send the familiarization packets to the applicants, and they also stated that they believed that work would need to continue to ensure that the packets meet the needs of the applicants.

Rater Training

The raters were trained to evaluate applicant performance using the performance dimensions developed for the OAP. This training was initially conducted by the Technical Workforce Services Team. Questions were asked to determine how well the training had prepared raters to evaluate applicant performance with a focus on the usability and utility of the performance dimensions.

Rater Survey. Overall, the rater training was seen as being highly effective (Table 5). A number of the raters indicated that it was difficult initially to get their ratings in agreement with each other. Some raters were seen as too focused on trying to help the applicant develop as a controller, while some had difficulty discriminating between performance levels in a consistent manner. Rater training was seen as useful in helping raters to align their ratings and, in some cases, was used to eliminate potential raters who were unable to conform to other raters or the guidelines. One member of the training department recommended that a refresher briefing be provided on the first day of each new skills assessment to go over the responsibilities and rules of rating. After training, all raters believed that they were well prepared to rate applicant performance.

Table 5. Rater Training

		M(1–7)	SD	Yes	No	N
29	How well were your raters (you) prepared to assess the skills of the OAP applicants?	6.83	0.34			24
30	How adequate were the performance dimensions for assessing the skills needed to work in your facility?	6.18	0.86			22
31	Should any of the currently assessed performance dimensions not be assessed?			0	21	21
32	Were there any performance dimensions that were not assessed that should have been?			2	20	22
33	How effective was rater training?	6.62	0.67			21
34	To what extent were you able to rate applicant performance using the definitions provided for each of the performance dimensions?	6.29	0.72			21
35	In general, how difficult were the performance dimensions to use in rating applicant performance?	1.90	1.26			21
36	To what extent were the performance dimensions defined well enough to allow you to distinguish between applicant skill levels?	6.05	1.12			21

		M(1–7)	SD			N
37	To what extent was the rater training on biases useful in making your ratings?	6.15	1.18			20

The performance dimensions were seen as very adequate for rating performance, although there was a suggestion to weight the dimensions based on their operational importance. Also, at one facility, some raters believed that there was too much overlap between some of the dimensions. Finally, the raters recommended that the facilities be allowed to customize some of the behavioral examples under the performance dimensions to be more facility/scenario-specific.

The final area discussed was rater training on biases and common rater errors. This training was seen as fairly effective and important. One recommendation was that the biases be listed on the rating sheet as a reminder. The extent to which training on biases was being included in new rater training was unclear. No respondents mentioned specific details on the training materials that were used in subsequent training sessions.

Selecting Official Survey. Raters were believed to be well prepared to do the assessments. At two facilities, it was reported that the performance dimensions were well thought out and pertinent. The remaining facility preferred to withhold judging their adequacy of the performance dimensions until observing the performance of those who were selected for a CPC-IT position.

Rating Process

The rating process section of the interview dealt with how the ratings were conducted, including the use of one or two raters, assignment of raters to applicants, the behavior of the raters toward the applicants, and the extent to which the raters participated in the selection decision (Table 6).

Table 6. Rating Process

		M(1–7)	SD	Yes	No	N
38	Was prior personal knowledge between raters and applicants considered in the process of determining which raters would evaluate which applicants?			12	8	20
39	Did you participate in the selection decision in any way other than rating applicant performance?			8	11	19
40	Did you have one or two raters per graded scenario?			17 (ONE)	7 (TWO)	24
	If you used one rater:					
41	Would you have preferred having two raters?			1	16	17
42	Were any problems encountered due to the use of one rater?			0	17	17

#	Question					
43	To what extent do you believe that having two raters was beneficial as compared to having one rater?	6.86	0.38			7
44	Were any problems encountered due to the use of two raters?			0	7	7
45	Did the same raters work together throughout the OAP?			4	2	6
46	Were ratings first made independently or together?			6 (Indep.)	0 (Together)	6
47	Did you notice any instances of raters saying or doing anything inappropriate (toward an applicant, while making ratings, afterwards, etc.)?			0	2	2

Note. Questions 38–39 were asked of 18 raters and three training personnel; Questions 40 was asked of three selecting officials, 18 raters, and three training personnel; Questions 41–42 were asked of two selecting officials, 12 raters, and three training personnel; Questions 43–44 were asked of one selecting official and six raters; Question 45–46 were asked of six raters; Question 47 was asked of three selecting officials.

Rater Survey. One rater was used to assess practice and graded scenarios at two of the three facilities. At the third facility, two raters were used during the graded scenarios. At all three facilities, raters were rotated around so that they got a chance to see multiple applicants, and applicants were viewed by multiple raters. At two facilities, the ratings were made using an over-the-shoulder process to observe the applicants. One of these facilities used two raters. At the other facility, the applicant's performance on each scenario was recorded for replay so that other raters could observe recorded sessions that they had not rated to have a better overall assessment of each applicant's performance. Raters at this facility would discuss applicant performance as a group to ensure that the best applicants were recommended to the selecting official. At the third facility, a single rater sat with a remote pilot operator and observed the performance of applicants on the pilot's display in real time. The remote pilot operators were also trained raters. In some cases, the official rater consulted with the remote pilot operator on what was observed. Raters rotated from applicant to applicant in one direction, and remote pilot operators rotated the opposite direction such that raters and pilots were paired differently across applicants for each scenario. Although multiple people viewed applicant performance and some discussion was allowed among trained raters at both facilities using a single rater process, it was ultimately the responsibility of the officially assigned rater to determine applicant scores. Neither facility commented that having one rater per applicant was a problem, and both preferred having one official rater per applicant, per scenario, than having two official raters. Using two raters per applicant, per scenario was also seen as an effective process. Raters commented that working with a second rater ensured that all pertinent applicant behavior was observed; having two raters was seen as a way to ensure rating consistency. The raters discussed their ratings and reached consensus before submitting final scores for each applicant

Training personnel at one facility also participated by observing during the test scenarios as a quality control measure but saw no need to intervene. One facility made no attempt to ensure that raters were not scheduled to observe applicants they knew personally. Raters were not scheduled to watch applicants they knew personally at the other facilities.

Two facilities compiled and submitted the rater scores to the selecting official before the selections were made. One of these facilities only involved the raters in the selection decision through submission of their scores on the skills assessment. At the other facility, raters developed a soft cut score and generally recommended selecting those with scores above this point but also included a few that fell slightly below this point. The third facility involved the raters more in the selection process. The raters at this facility met with the selecting official to discuss the performance of each applicant and to make a final recommendation of "select" or "not select." The selecting official did not see the scores until after the selections had been made.

Selecting Official Survey. Selecting officials at two facilities reported that they used one official rater per scenario and did not experience any problems using one rater. The only issue encountered was the need to allow one applicant to redo a scenario because the remote pilot operator was not performing to standard and had to be replaced. At the third facility, two raters were used per scenario. This process was also seen as effective by the selecting official.

Feedback

There were several opportunities to provide feedback to applicants during the skills assessment process. The feedback procedures prescribed by the notice included giving applicants group level feedback after each practice scenario. Raters and selecting officials were asked to describe the type of feedback applicants were given about their skills assessment performance, if any. Questions were also asked about the process through which this information was given to raters, if any occurred. And, selecting officials were asked how applicants were given feedback on the final selection decision (Table 7).

Table 7. Feedback

		M(1–7)	SD	Yes	No	N
48	Were raters provided with any information about how the selection decisions were made or how their scores were used in the decision?			0	2	2
49	To what extent were applicants allowed to ask questions about the job, the facility, or the selection process during the OAP?	6.79	0.49			19
50	To what extent do you think the group level feedback provided to the applicants during the practice scenarios was effective?	5.90	1.48			20

Note. Question 48 was asked of three selecting officials; Question 49 was asked of three selecting officials, 18 raters, and three training personnel; Question 50 was asked of 18 raters and three training personnel.

Rater Survey. The raters participated only in the group-level feedback sessions by providing notes, comments, and suggestions to the person in charge of the training. The training personnel then delivered this feedback to all the applicants simultaneously. It was the general impression of the raters that the applicants did better on the scenarios after they received this feedback. It was suggested that going forward, a standard list of suggestions could be generated for use in group level feedback, given the common nature of the mistakes made by the applicants. It was also suggested that raters should be able to give instant feedback to the individual during the practice scenarios and then provide the information *again* in the group feedback session so that everyone could benefit from the suggestions. Raters at all facilities said that no feedback was provided during the graded scenarios, but that they would have liked to provide feedback to the applicants on their performance on the graded scenarios. Additionally, raters believed that applicants were allowed to ask questions freely, except during graded scenario runs.

Selecting Official Survey. In general, selecting officials would like to provide feedback to applicants on their performance at the end of the assessment. Although they recognized that this was not a training session, it was still the consensus that the applicants should receive some type of feedback on their performance. As it is now, applicants learn the selection decision through the Aviator system. Applicants can call the selecting official for more information, but most do not. The selecting officials believe that performance feedback would be useful for those planning to reapply in the future.

OAP Processes

Facility personnel were asked to assess the extent to which the OAP identified applicants with the skills needed to work in their facility and its ability to assess the likelihood that those applicants selected would succeed. They were also asked about the fairness of the OAP and if they would change it in any way. Furthermore, they were asked about any instances of equipment failure, applicant withdrawal, or anything else that might have occurred (Table 8).

Rater Survey. All raters indicated that the OAP was an effective process and that those selected using the OAP were highly likely to succeed. There was even a comment that using the OAP to select applicants should decrease time to certification and that all applicants should be able to certify in less than a year. The raters also believed that the OAP was administered fairly, with everyone being treated equally and given the same chance as everyone else to succeed. Although a few minor equipment malfunctions had occurred, mostly during practice runs, no applicant's performance was affected. The only surprise to some raters was a perceived lack of effort or poor attitudes displayed by a few applicants. These applicants were seen as wasting the time of the raters and the money of the facility. A number of suggestions for modifying the OAP were made by the raters. Several raters suggested slight revisions to or fine tuning of the scenarios. One rater even suggested eliminating the departure scenarios and using arrival or combined arrival and departure scenarios instead. This rater viewed it as a missed opportunity to view the applicants' performance in a more challenging situation. Raters also suggested providing more performance feedback to the applicants at the end of the OAP. Another suggestion involved limiting those who were required and/or allowed to participate in the OAP and that applicants from Level 12 TRACONs did not need to participate in the OAP. It was thought that there should be additional rules for screening applicants based on legal limitations, for example, due to mandatory retirement age, and that such rules should be developed and used to limit the applicant pool. There was also a recommendation to include personality, work attitudes, or work behavior assessments, either before or during the applicants' time at the facility.

Table 8. OAP Processes

		M(1–7)	SD	Yes	No	N
51	To what extent do you think that the OAP is an accurate assessment of the skills needed to work at your facility?	6.50	0.71			2
52	Did you receive any complaints from applicants about the OAP?			2	6	8
53	How effective do you think the OAP was in identifying candidates who, if selected, would succeed on the job?	6.88	0.39			29
54	Do you think the OAP was administered fairly?			29	0	29
55	Did anything else unique or unusual happen during the process?			7	17	24
56	Were any applicants unable to complete a test scenario for any reason (e.g., equipment failure, lab problem, or applicant performance)?			4	20	24
57	Did any applicants withdraw from the selection process once they started the skills assessment?			0	24	24
58	Would you change the OAP in any way?			19	5	24
59	To what extent did you feel all applicants were treated the same during the skills assessment?	6.87	0.44			23
60	Did you experience any situations that you felt were necessary to discuss with other raters to ensure that raters remained consistent in their ratings?			11	10	21

Note. Question 51 was asked of three selecting officials; Question 52 was asked of three selecting officials, three facility union representatives, one training manager, and one facility manager; Questions 53–54 were asked of three selecting officials, 18 raters, three training personnel, four facility union representatives, and one facility manager; Questions 55–59 were asked of three selecting officials, 18 raters, and three training personnel; Question 60 was asked of 18 raters and three training personnel.

Selecting Official Survey. The selecting officials at two facilities believed that the right skills were being assessed for the job and that those selected would be successful. At the other facility, the selecting official was not sure and wanted to wait and see OAP selectees performed during subsequent training. The selecting officials believed that the OAP had been administered fairly and that any problems encountered during the administration of the OAP were managed appropriately without having a negative effect on any applicants' opportunity to succeed.

Recommendations

One recommendation was that the OAP be changed to allow for feedback to be given to the applicants at the end of the process. Another recommendation was that the knowledge test be eliminated as long as the applicants were required to have radar experience. Selecting officials suggested that the knowledge exam would be more appropriate for use with applicants with no radar experience (i.e., tower-only experience). Finally, there was some concern expressed at one facility about the timeframe for staffing using the OAP. They would like to start using the OAP earlier in the fiscal year.

General Survey. The FACREPs indicated that they had received a few complaints. Specifically, a complaint heard at one facility was that applicants wanted more feedback at the end of the process. The FACREP was not sure who should give the feedback but believed that it should be provided. There were also some comments about the length of the assessment, with some indicating that the classroom training was too long and some indicating that the OAP was too short. One FACREP said that "those who do well do not complain and that those who do not do well sometimes complain." Overall, the FACREPs reported that the OAP was a fair and equitable process.

Selection Decision

Selecting officials were asked about their decision-making process for selecting CPC-ITs. Specifically, they were asked the extent to which the components of the OAP were used to inform selection decisions, the use of or need for additional information, and if any other information influenced their decision (Table 9).

Table 9. Selection Decisions

		M(1–7)	SD	Yes	No	N
61	How useful did you find the knowledge exam results in making your selection decisions?	1.67	1.15			3
62	How useful did you find the skills assessment scores in making your selection decisions?	7.00	0.00			3
63	Was there information from the OAP you would have liked, but did not have access to, in making your selection decisions?			0	3	3

Note. Questions 61–63 were asked of three selecting officials.

Selecting Official Survey. In making the selection decision, the perception of the selecting officials was that the knowledge test was not useful. However, the skills assessment was seen as highly useful. In addition to the OAP, the selecting official at one facility talked to each applicant's current facility managers/supervisors. This selecting official did not see applicants' scores on the skills assessment prior to making the selections. Rather, a meeting was held with the raters to discuss the applicants and to get their recommendation to select or not select an applicant. The selecting officials at the other facilities reviewed the OAP scores on the skills assessment and the applicants' previous ATC experience, in particular their radar experience.

Documentation

One measure of OAP success will be determined by how well those selected using the OAP perform during training. Do they achieve certification and, if so, in what length of time? To assess the utility of the OAP, it will be necessary to document and maintain applicant scores and performance in training. Selecting officials will be responsible for ensuring that the processes necessary to document and maintain the data are established. To do so will require resources. These issues were addressed with the selecting official survey (Table 10).

Table 10. Documentation

		M(1–7)	SD	Yes	No	N
64	To what extent do you have the resources needed to document and maintain applicant data?	7.00	0.00			3
65	Is there anything that could be done to better help you record and document this process?			0	3	3
Note. Questions 64–65 were asked of three selecting officials.						

Selecting Official Survey. Although there is a cost, all three facilities have the resources to document and maintain the applicants' scores and demographic information. The training departments at the facilities are responsible for this portion of the process and the records are maintained primarily through the use of spreadsheets.

Comparisons

Survey respondents were asked to make two types of comparisons. First, they were asked to indicate how the OAP applicants compared with previous applicants or new hires and if the applicants performed as expected on the skills assessment. Respondents were also asked to compare the use of the OAP for selecting in relationship to other staffing processes used by the FAA (Table 11). Finally, they were asked if the OAP should be extended to other high level TRACONs, and if it should be implemented at one centralized location to support controller staffing at all high level TRACONs.

Rater Survey. Some raters had no expectations regarding how well each applicant would perform, but a few raters were surprised by how some of the OAP applicants performed based on their current facility. Some applicants were expected to do better than they actually did, and some performed much better than expected, based on current facility assignment. The raters saw this as evidence that it is hard to judge how well an applicant will do based on where they are currently assigned, thus, the need for this type of assessment.

Table 11. Comparisons

		M(1–7)	SD	Yes	No	N
66	How does this process for hiring compare with other hiring procedures you have experienced (or helped to administer) in the FAA?					
66a	Is it more or less fair?	6.57	0.79			29
66b	Is it more or less effective?	6.61	0.70			28
67	To what extent do you think that the OAP should be extended to other TRACONs?	6.00	1.48			7
68	Would you be interested in having the OAP implemented at one location to support hiring of new controllers at all high level TRACONs?			2	6	8
69	To what extent did applicants perform as well as you expected they would?	4.25	1.12			20

Note. Questions 66a–b were asked of three selecting officials, 18 raters, three training personnel, three facility union representatives, one training manager, and one facility manager; Questions 67–68 were asked of three selecting officials, four facility union representatives, and one facility manager; Question 69 was asked of 18 raters and three training personnel.

Generally, the OAP applicants were seen by the raters as having a much higher likelihood of success than new hires. The OAP as a staffing process was seen as being more fair and effective than other staffing procedures (e.g., reviewing applications and résumés from the ERR list) in the FAA. One advantage of the OAP is that the applicants are actively looking to progress their careers and are seeking the challenge of a high level TRACON, rather than direct new hires with little choice in facility placement options who may or may not want to begin at this level impacting motivation and confidence during training. Another advantage seen by the raters is that the selecting officials are able to assess actual applicant performance on a job sample task, rather than having to rely on a paper review of an applicant's experience and qualifications. However, some suggested that the ultimate test of effectiveness is yet to be determined. The issue will be how many OAP applicants certify and in what amount of time.

Selecting Official Survey. The selecting officials thought the applicants' qualifications seemed to be about the same as those of other experienced applicants they receive, although they indicated that there were clear differences when compared with general public hires. The selecting officials indicated that the OAP allows for an assessment of who among the applicants is highly qualified, although there will be a need to wait and see how well the applicants do during training. They perceived the OAP to be more fair and effective than the previous processes because the facility gets to see the applicant perform, not just review an application. They also thought it was better for the applicants because they get to see the facility and the local area, and thy have a better understanding of what will be expected of them before accepting an offer.

In regard to expanding the OAP, the selecting official at C90 believed that Notice JO 3330.70 should indicate that the ATO can implement the OAP wherever needed, in such places as Detroit or New York. The selecting officials at SCT and A80 also believed the OAP should be implemented where needed, based on staffing needs. However, the selecting official at SCT did not believe the OAP is needed at SCT. He believed he could get the people he needs using the ERR list and the current paper review of applicants' experience. In general, selecting officials did not recommend having the OAP implemented in one centralized location to support staffing of new controllers at all high level TRACONs. Conducting the skills assessment portion of the OAP at the selecting facility allows the applicant to come to the facility and observe operations.

General Survey. The respondents thought that comparing OAP applicants with previous applicants, especially those with no radar experience, was a difficult comparison to make. The OAP applicants perform significantly better *because* they have radar experience. The OAP has also broadened the applicant pool by allowing applicants from lower level facilities to participate – giving a chance to those who would not normally be selected only using a paper review of their experience at high level facilities. The OAP was seen as being more fair and more effective than other staffing procedures (i.e., reviewing applicants from the ERR list). The only fairness concern expressed was the requirement that applicants have radar experience to take the knowledge test. It was proposed that applicants lacking experience, but with sufficient experience to pass the knowledge test, should be allowed to participate in the OAP. In addition, it was recommended that the OAP be extended, if needed, to all high level TRACONs, starting with those facilities with the greatest need based on success rates of trainees.

Of the respondents, only one indicated that the OAP should be centrally located. The other two preferred a more customized approach with each facility allowed to refine the OAP to meet the needs of the facility. They thought that having the OAP at the facility allowed the applicant a more realistic preview of what was expected on the job, the workload, the people, and the local area. It also allowed the facility to see the applicant perform, which is a key benefit of the OAP in comparison to using the ERR list. Arguments for centralization of the OAP were to better standardize and control implementation of the process, including the training of raters and the building of scenarios. If centralized, the possibility that facilities could send raters to participate in the OAP skills assessment was proposed. Additionally, the scenarios could use a generic airspace but be tailored to the issues that participating facilities face. They could be designed as a collaborative effort among the high level facilities.

Summary

Implementation of the OAP at the Atlanta, Chicago, and Southern California TRACONs was perceived to have gone very well by the vast majority of those closely involved in the process. Most respondents indicated that the OAP would likely improve the selection of CPC-ITs, which would be seen with those selected succeeding at a higher rate and needing less time for certification than CPC-ITs selected without the OAP. The skills assessment was seen as a strength of the OAP. Respondents said that the scenarios were a sufficient test of the skills needed by a candidate and that the performance dimensions allowed raters to distinguish between different levels of applicant performance. However, some respondents did recommend that the process for implementing the skills assessment be modified to include immediate feedback to the candidates on their performance after the graded scenarios. If immediate feedback is not possible, given potential legal restrictions and/or issues of perceived fairness, it might be best to provide some form of written feedback from the facilities after selection decisions have been made. Some respondents also recommended that the performance dimension descriptors be revised and that they

be allowed to customize the descriptors to be site-and-scenario-specific exemplars of performance. The knowledge test was not seen as a valuable component of the OAP, except for as a way to ensure that each candidate had the required radar experience which could also be determined by examining previous facility experiences and reference calls. The knowledge test did not provide a way to distinguish levels of experience or expertise in radar. It simply helped to ensure a basic level of understanding and knowledge. One suggestion was that the knowledge test only be used, if applicants without radar experience were to be included in the future and only be required of those without reported radar experience.

Regarding rater training, it was unclear to what extent new raters were being trained to the same standard as those trained initially. A more structured approach for new rater training will be needed to ensure that all raters are trained to the same standard. This will allow all new raters to equitably assess performance and allow for the selection of the most promising applicants. Standardizing new rater training will be especially important if, as preferred by most respondents, the OAP continues to be conducted on-site rather than in a centralized location. The ability of the TRACONs to maintain a well-trained cadre of raters is critical to implementation of the OAP. Refresher training will also be an important component of maintaining the program.

Also important will be the documentation of the OAP records. Respondents indicated that although the OAP is resource-intensive, they do have the resources needed to document and maintain the OAP data. These data will be needed if the impact of the OAP is to be determined. It is also important, given the resources involved, that only the most promising applicants be selected to participate in the skills assessment. In a few cases, respondents were surprised to find that some applicants seemed to display disinterested or combative attitudes. In addition, some applicants were nearing mandatory retirement age, and it was unlikely that, even if selected, they would have sufficient time to transfer and certify at a new facility—even with a one-year waiver. These applicants were seen as a drain on facility OAP resources, as they probably should have been excluded as facility transfer candidates on the basis of the limited time remaining in their CPC careers. Applicants also varied in how open they were to instruction during the practice scenarios. To improve the OAP, it may be necessary to broaden and expand the criteria used to determine OAP eligibility.

This initial evaluation is only the first step in a comprehensive effort to ensure that the OAP is implemented in a valid and fair manner. An additional assessment from the perspective of the OAP applicants is currently in progress. Furthermore, once applicants have transferred to their new facilities, data on their training performance should be collected and analyzed. When enough data are collected, a statistical analysis of the fairness and validity of the OAP will be completed.

In sum, given the data collected to date, the OAP is perceived to have been successfully implemented with relatively minor issues at the three facilities evaluated. Recommended changes for improvement include:

1. Design a legal, standardized way to deliver more substantial feedback to the applicants after they complete the OAP, particularly for those not selected.

2. Allow for the customization of the behavioral anchors on performance dimensions so that facilities can tailor the behaviors that raters are to observe during the OAP scenarios.

3. Increase the breadth and/or difficulty of the knowledge test so that it discriminates among high and low performing applicants, or only use the knowledge test when applicants have no radar experience.

4. Design a structured rater training program with refresher training and implement the program for the training of future raters.

5. Implement standardized documentation and reporting procedures across OAP facilities, so that the fairness and validity of the OAP can be assessed.

6. Include a set of limiting, as well as exemption factors for participation in the OAP. For example, one limitation that would be appropriate would be to exclude applicants who are approaching retirement age, unless they have applied for a waiver or have previously worked at a Level 12 TRACON.

7. Include other measures (e.g., work attitudes, work behaviors, personality assessment) in the OAP. This could potentially be a structured interview administered after the skills assessment.

CONCLUSION AND NEXT STEPS

The content validation of the OAP was consistent with industry standards, and the use of job analysis procedures to establish content validity and job relatedness have been accepted by the courts (AIR, 2011; EEOC, 1978). This is especially true when simulations and work samples are used to evaluate applicants. As a consequence, the OAP is a defensible job placement selection tool as currently applied.

One area in which the initial evaluation revealed to be problematic is the use of the knowledge exam to identify the applicants most likely to succeed during the skills assessment. The OAP knowledge exam cut score was set based on pilot tests using over 203 air traffic controllers. The exam working group initially expected that the knowledge exam would screen out 30% of the applicants, but this expectation was based on the misperceptions that OAP facilities would allow applicants with no radar experience to apply via the OAP. However, the pass rate was approximately 99% for the 150 applicants tested to date because the OAP facilities later decided to limit the applicant pool to those with radar experience. With a pass rate this high, the knowledge exam lacks effectiveness in differentiating or screening applicants before they can progress to the rated performance scenarios.

Several options are being considered regarding the knowledge exam. One OAP facility has recommended eliminating the exam, while the other two facilities would like to see the exam cut score raised. Program administrators have recommended eliminating 25% of the easiest questions and re-administering the exam.

The preliminary feedback on the rating process was positive from both raters and administrators. One issue, in regards to rater training, is whether or not it should be implemented on a recurring basis. Rater proficiency remains a concern, and stakeholders are aware that fair and unbiased ratings are crucial to the program's success. Further discussion of this issue will take place before any changes are made to the OAP. Additionally, the performances dimensions of Communication, Working Speed, Planning and Projection, Prioritizing, and Speed Control are working well for OAP raters, but some respondents suggested combining dimensions. Currently, any discussions concerning changes in the OAP rating dimensions are preliminary, and no immediate changes are expected.

The comments from training support managers regarding the performance of selected applicants have, thus far, been favorable. However, until a sufficient number of selected applicants actually, complete training and achieve certification at their new facilities, there is insufficient data to verify that the OAP is successful. Procedures are now in place to collect the data needed to complete a full assessment and evaluation of the OAP. These data will be used to solidify the validity of the Operational Assessment Program as a transfer/promotion tool.

REFERENCES

Alexander, J. R., Ammerman, H. L., Fairhurst, W. S., Hostetler, C. M., & Jones, G. W. (1989). *FAA Air Traffic Control operations concepts Volume VIII: TRACON controllers* (DOT/FAA/AP87/01). Atlantic City International Airport, NJ: DOT/FAA Technical Center.

American Institutes for Research. (2011). *TRACON Operational Assessment Validation.* (Draft report prepared under FAA contract DTFAWA-09-A-80027 Appendix H). Oklahoma City, OK: Federal Aviation Administration Management Service's Technical Workforce Hiring Team (AJG-P33).

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). Uniform guidelines on employee selection procedures. *Federal Register, 43*(166), 38290–39315.

Krokos, K. J., Baker, D. P., Norris, D. G., & Smith, M. A. (2007). *Development of performance standards for Air Traffic Control Specialists* (FAA technical report submitted under Grant # 99-G-048). Washington, DC: American Institutes for Research.

U.S. Department of Transportation, Federal Aviation Administration, Air Traffic Organization Policy. (2014). *Operational Assessment Program (OAP)* (Notice # JO 3330.70). Retrieved from http://www.faa.gov/documentLibrary/media/Notice/OAP_3330.70.pdf