



**Federal Aviation
Administration**

DOT/FAA/AM-08/13
Office of Aerospace Medicine
Washington, DC 20591

Screening Air Traffic Control Specialists for Psychopathology Using the Minnesota Multiphasic Personality Inventory-2

Raymond E. King
David J. Schroeder
Carol A. Manning
Paul D. Retzlaff
Clara A. Williams

Civil Aerospace Medical Institute
Oklahoma City, OK 73125

June 2008

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 Web site:
www.faa.gov/library/reports/medical/oamtechreports/index.cfm

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-08/13	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Screening Air Traffic Control Specialists for Psychopathology Using the Minnesota Multiphasic Personality Inventory-2		5. Report Date June 2008	
		6. Performing Organization Code	
7. Author(s) King RE, Schroeder DJ, Manning CA, Retzlaff PD, Williams CA		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-HRR-523			
16. Abstract The purpose of this paper is to model and document the use of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) as a psychological screening tool for conditionally selected FAA Air Traffic Control Specialists (ATCSs). A sample of 1,014 ATCSs in training voluntarily completed the MMPI-2 as part of a research program. Those data are used to estimate the number of future candidates that will be referred for follow-up psychological evaluations, given varying MMPI-2 scale cut-scores. At the individual scale level, Scale 1 (<i>Hypochondriasis</i>) had the lowest percentage of subjects identified across all cut scores, while Scale 9 (<i>Hypomania</i>) had the highest. Looking at participants with one or more scales above the cut-scores, about 15% had one or more scales at or above 65T, but only about 2% had one or more scales at or above 80T. A final algorithm of 70T or above on scales 1, 2, 3, 4, 6, 7, and 8, as well as 75T on scale 9, was selected. The identification rates are discussed in terms of impact on follow-up psychological evaluation referral. Initial cut scores are recommended.			
17. Key Words Personnel Screening, Personnel Selection, Psychological Testing		18. Distribution Statement Document is available to the public through the Defense Technical Information Center, Ft. Belvoir, VA 22060; and the National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 16	22. Price

CONTENTS

THE MMPI-2	2
Scales	2
METHOD	3
RESULTS	4
Cut Score Decision Process.	6
Cut Score Decision.	7
Applicant Level Reliability	8
DISCUSSION AND RECOMMENDATIONS.	9
REFERENCES	10
APPENDIX A: Excerpt from FAA Order 3930.3A	A1

SCREENING AIR TRAFFIC CONTROL SPECIALISTS FOR PSYCHOPATHOLOGY USING THE MINNESOTA MULTIPHASIC PERSONALITY INVENTORY-2

Selecting among applicants for safety-sensitive jobs such as air traffic control specialist from pools of promising individuals can be a difficult task because applicants and organizations usually have conflicting goals. Applicants are attempting to look their best to increase their chances of being hired and are likely to minimize any mental health issues. At the same time, organizations attempt to hire the most qualified and fit individuals, those who are most likely to successfully complete their training and become effective employees. Organizations face two challenges when hiring. The first task is to “select in” those applicants who have the positive attributes that are required. Select-in methods determine who is best suited for completing the complex tasks associated with safe and efficient performance. The second task is to “select out” those applicants with negative qualities that would pose a safety risk or otherwise make success difficult due to their limited adaptability. Neither approach alone is sufficient. Select-out criteria often eliminate applicants with a psychiatric diagnosis (as currently defined in the *Diagnostic and Statistical Manual of Mental Disorder*, DSM-IV TR; APA, 2000) suggesting a lack of fitness. The result here is the identification of a small, probably impaired subset of the candidate pool. Aviation occupations require the highest standards of psychological suitability and fitness to ensure the safety of the public.

Due to the high stakes involved, Butcher (2002) issued a call for research into the mental health of commercial pilots and advocated the use of modern psychometric instruments. He made no mention, however, of other vital members of the aviation team such as air traffic control specialists (ATCSs). To ensure aviation safety, personnel in both occupations need to be alert, attentive, and ready to respond immediately to critical events. Although the Federal Aviation Administration (FAA) certifies the medical fitness of pilots and ATCSs, commercial aviators typically work for airlines or other private employers that bear the responsibility for selecting qualified applicants. The vast majority of ATCSs (commonly termed “air traffic controllers”) are employees of the FAA. Thus, the FAA is responsible for developing the appropriate selection tools and conducting the selection screening.

In an effort to ensure the emotional health (fitness) of the ATCS workforce, the FAA has used the 16 Personality Factor (16 PF) test since 1965 during the medical assessment (select out) process. This procedure is part of the

Controller Health Program which was initiated by FAA Order 9430.2 and now is outlined in FAA Order 3930.3A (see Appendix A). A “case identifier” scale, composed mostly of anxiety items from the 1967-1968 edition, has evolved over the years (Convey, 1984). The 38-item scale displays acceptable reliability (the 18 items from Form A had a Cronbach alpha of .71, and the 38-item full scale had a Cronbach alpha of .85; King, Retzlaff, Detwiler, Schroeder, & Broach, 2003). Unfortunately, its clinical utility has been limited. Historically, only a very small percentage of job candidates have been identified by this scale. Dollar, Broach, and Schroeder (2003) indicated that the 16PF is somewhat effective in predicting who will go on to retire on disability but posited that other factors must be at play in determining disability retirements. It is possible that some individuals who would have gone on to retire on disability were screened out with the 16PF, leading to a potential restriction in range.

There was interest in comparing the 16PF with other psychological tests as early as 1971, when Smith gauged the item ambiguity (the degree to which an item elicits multiple interpretation) of the 16PF to the Minnesota Multiphasic Personality Inventory (MMPI). Smith found the items of the MMPI to be less ambiguous than those of the 16PF. While both tests have been substantially revised since 1971, the FAA continues to use the 1967/1968 edition of the 16PF. In 1996, Schwarzkopf, Buckley, and Pace urged replacement of the FAA’s 16PF procedure due to declining scientific interest, its sole focus on anxiety symptoms, and its “fakeability.” In their paper, written under contract to the FAA, they urged consideration of the MMPI-2. Indeed, the point of the current paper is to explore the feasibility of using the MMPI-2 for this *initial* screening of ATCS candidates to better identify those harboring symptoms suggestive of emotional instability and requiring additional assessment.

Graham (1990) noted that the MMPI-2 has been used in two ways in selection. It can be used to screen for psychopathology, and it can be used to predict the quality of an applicant’s job performance by matching personal characteristics to job requirements. Graham wrote that police and nuclear power plant operators are best selected by eliminating persons with very elevated scores on one or more of the clinical scales, obviously following the former approach. Graham asserted that screening applicants for psychopathology is most justified

when considering individuals for occupations involving susceptibility to stress, personal risk, and personal responsibility. He specifically delineated “air traffic controller” (p. 197) among such sensitive occupations. Lowman (1989) provides useful information regarding some of the difficulties and concerns when conducting pre-employment screening for psychopathology, focusing mainly on the occupation of nuclear power plant workers. Westefeld and Maples (1998) reported on matching applicants, mostly in police departments, on the basis of the MMPI-2 to successful occupational incumbents.

The major concerns about the future psychological functioning of ATCSs, who are young at the time of entry into the workforce (by law, they must be younger than 31 years of age), include mood disorders and other DSM-IV-TR (2000) axis I disorders (termed “neurosis,” and “psychosis” in FAA Order 3930.3A; see Appendix A), in addition to personality disorders. Hammen (2001) summarized the epidemiological research on depression and cited varying lifetime prevalence, and concluded that prevalence estimates are influenced by demographic factors (with women and those with lower income and education levels having higher rates), as well as the method of assessment. McNally, Malcarne, and Hansdottir (2001) called for increased longitudinal research on the spectrum of anxiety disorders to better understand these conditions and their development during the course of a lifetime. Harvey (2001) explored the relatively rare prevalence (approximately 1% of the population) of schizophrenia and noted that it is particularly deleterious to occupational success when it develops in late adolescence/early adulthood. Finally, Geiger and Crick (2001) considered personality disorders, by definition an *enduring* pattern of maladaptation, and found that the diagnosis is typically not reliable; hence, prevalence estimates are problematic. Overall, the prevalence of psychopathology in the pool of candidates for ATCS positions is likely to be relatively low, particularly at the time the candidates are assessed, but the potential consequences cannot be discounted.

There is therefore a growing interest in enhancing the FAA’s ability to identify those ATCS candidates who possess psychological symptoms that could compromise aviation safety. The assessment of such psychological symptoms is currently being conceptualized as a two-tier process. The first step will be the testing of all candidates that have been tentatively offered employment in compliance with the Americans with Disabilities Act (1991). This will entail the use of the MMPI-2. Those that score above certain levels would then be evaluated more thoroughly by a licensed psychologist. This approach is consistent with FAA Order 3930.3A, which specifically prohibits medical disqualification on the basis of a single

psychological test (FAA, 1980). This study was designed to explore the feasibility of utilizing the MMPI-2 to replace the 16PF as the initial screen.

THE MMPI-2

Scales

There are 13 main scales on the MMPI-2. The first three are “validity” scales and used to determine the test-taking “style” of the client. The other 10 are the main “clinical” scales and focus on various psychopathologies. The scales have names, but these names are typically not used in the profession because they are quite archaic. Instead, clinical psychologists refer to the scales by their letters or number. Hence, a patient scoring high on the fourth scale would not be said to have scored high on “Psychopathic Deviate” but to have scored high on “PD” or “scale 4.”

Validity Scales

L (“Lie”). Elevation reflects a deliberate attempt of the individual to present him/herself in a positive light: Denial of minor flaws/weaknesses that most individuals would admit. Excessive elevations on this scale (relatively rare) render the profile invalid and hence uninterpretable.

F (“Fake Bad”). Used to detect atypical ways of responding. Scored responses are agreed to by few normal adults. High scores may reflect an invalid test profile due to malingering. **Not expected in a job applicant;** more likely in a person seeking the benefits of a patient (compensation, avoidance of jail).

K (“Fake Good”). A more subtle index of individuals who are trying to present themselves in a positive light. Tends to be elevated as education and socioeconomic status increases (and when used in an employment selection setting); denies psychopathology. Used to correct several of the clinical scales (Scales 1, 4, 7, 8, & 9) by adding various amounts of *K* to get a more accurate reading of their functioning.

Clinical Scales

Scale 1 – Hypochondriasis. High scores reflect individuals who have an excessive number of vague nonspecific complaints and body concerns (gastrointestinal distress, fatigue, pain, and general weakness).

Scale 2 – Depression. Individuals with high scores often reflect depressive symptoms (depressed, blue, unhappy) and are generally lacking in self-confidence.

Scale 3 – Hysteria. Symptoms involve two dimensions. One reflects a general denial of physical health and includes rather specific somatic complaints. The other group involves a general denial of psychological or emotional problems.

Scale 4 – Psychopathic Deviate. High scores reflect difficulty in incorporating the values and standards of society; may involve asocial or antisocial behaviors, impulsiveness, and need for immediate gratification. May be a bit elevated in younger test takers as a normal function of late adolescence.

Scale 5 – Masculinity-Femininity. A legacy bi-directional scale (in other words, both low and high scores have meaning) that is not considered a clinical scale. Reflects interest and not sexual orientation. As it does not indicate psychopathology, it is not suggested for use in the medical screening of ATCSs.

Scale 6 – Paranoia. High scores reflect individuals with disturbed thinking, ideas of reference, suspiciousness, hostility, and paranoia.

Scale 7 – Psychasthenia. High scores reflect individuals experiencing a great deal of psychological turmoil and discomfort. They tend to be anxious, tense, and agitated. They are worrisome individuals that have difficulty concentrating.

Scale 8 – Schizophrenia. High scores are reflective of bizarre mentation, delusions, and possible hallucinations. Confused thinking, poor judgment, and alienation are common.

Scale 9 – Hypomania. High scores are suggestive of overactivity, poor impulse control, irritability, and possible aggressive outbursts.

Scale 0 – Social Introversion. Not a clinical scale. Individuals with high scores tend to be introverted, while low scorers are extroverted. As it does not indicate psychopathology, it is not suggested for use in the medical screening of ATCSs.

T Score Conversion. The number of items endorsed in the keyed direction on each scale (the “raw score”) are converted to T scores by using the published norms. These are standard scores with a mean of 50 and a standard deviation of 10. Thus, a T score of 70 means that the individual scored two standard deviations above the mean of the population on which the test was normed for the scale in question. An elevation of two standard deviations may be better understood as the 95th percentile, meaning higher than 95% of the population. So, the higher the T score values, the more items the person taking the test endorsed, suggesting a relatively greater presence of traits consistent with a psychiatric disorder.

METHOD

The MMPI-2 was administered to a cohort of recently hired ATCSs during their first days of training at the FAA Academy in Oklahoma City, OK. These 794 male and 220 female fledgling air traffic controllers had all been hired as ATCSs and thus had been interviewed at an air traffic control facility and had passed a medical examination, which included passing the existing 16 PF case-identifier-procedure hurdle. All voluntarily agreed

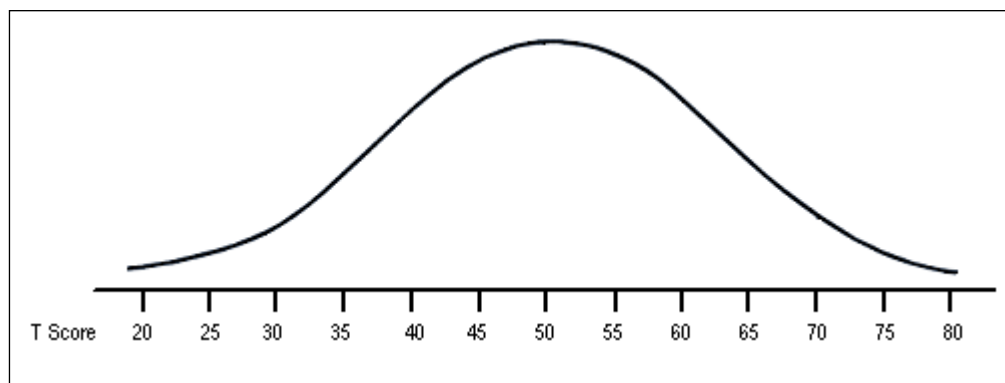


Figure 1: T-scores and the normal distribution

to participate in the present study. All participants had at least 12 years of education, as a high school diploma is a pre-requisite to be hired as an ATCS. Many participants had several years of college, as most were hired under the College Training Initiative (CTI) program, which is hosted by 13 four-year and community colleges throughout the United States. Indeed, 544 of the 1,014 participants indicated that they had more than 12 years of education.¹

Non-gender norms were used, as this research is intended to support personnel selection and gender-specific norms are prohibited by the Civil Rights Act of 1991. A consideration of gender was also unnecessary due to the lack of inclusion of scale 5 (Masculinity-Femininity), which specifically requires attention to the test taker's gender. Non-K corrected Clinical Scales are also reported here, as *K* corrections tend to be elevated in applicant populations (due to the tendency for positive impression management – also known as “faking good”). Table 1 presents *K*-corrected clinical scales to allow comparison to the data published by Butcher (1994).

RESULTS

Table 1 provides the means and standard deviations for the *K*-corrected MMPI-2 scores for the 1,014 ATC applicants; Figure 2 presents this information graphically for ease of comparison. Again, *T* scores represent the norms, with a mean of 50 and a standard deviation of 10. As this sample's data diverge from these, differences become apparent. In terms of validity scales, both *L* and *K* are about 7 points higher. This elevation is almost one

standard deviation and approaches a significant elevation. It is apparent that both groups, air traffic control specialists and pilot applicants, have placed themselves in a positive light and deny pathology.

Nevertheless, the clinical scales are remarkably similar to the general population normative group published in the MMPI-2 manual (Butcher, Graham, Ben-Porath, Tellegen Dahlstrom, & Kaemmer, 2001). Here, clinical scales do not vary as much as 3 points from the norm. It would be expected that the sample would minimize pathology in the clinical scales, but they do not. Minimization of these scales would entail means down around 40, not at the mean of 50. The additive *K* corrections, however (.5 for scale 1, .4 for scale 4, 1 for scale 7, 1 for scale 8, and .2 for scale 9), may inflate some of the clinical scales to an extent.

Table 1 also provides the means and standard deviations for the pilots from Butcher's (1994) work. While the normative sample used to establish the *T* scores can serve as a general control group, Butcher's pilots can be considered an aerospace-specific control group of individuals of approximately the same age, although half a generation apart. This pilot sample data differs from that of ATCSs in other ways as well. While Butcher tested only men, the current sample is about 20% female. Also, Butcher's data were collected before applicants were offered employment. As such, the current comparison is less than ideal. That being noted, Butcher's pilots scored at about the same level on the *L* validity scale as our sample. His pilots also scored about 16 points above the mean on the *K* scale

Table 1. *K*-corrected means and standard deviations for air traffic and pilot samples.

	Air Traffic Mean (<i>sd</i>)	Pilots Mean (<i>sd</i>)	t-test	<i>p</i>
<i>Validity Scales</i>				
<i>L</i>	57.5 (11.7)	57.4 (11.7)	0.149	0.559
<i>F</i>	46.9 (9.4)	40.4 (3.1)	19.676	0.001
<i>K</i>	57.2 (9.3)	65.7 (6.4)	20.089	0.001
<i>Clinical Scales</i>				
<i>1</i>	50.3 (7.5)	48.3 (4.5)	6.268	0.001
<i>2</i>	47.1 (7.6)	42.9 (9.9)	7.920	0.001
<i>3</i>	48.9 (7.5)	52.3 (10.1)	6.326	0.001
<i>4</i>	50.4 (7.9)	49.3 (6.0)	2.899	0.002
<i>6</i>	48.3 (8.7)	47.8 (5.7)	1.295	0.098
<i>7</i>	47.9 (8.0)	48.4 (4.7)	1.483	0.931
<i>8</i>	49.2 (8.4)	47.8 (4.6)	4.075	0.001
<i>9</i>	52.6 (9.5)	46.0 (5.5)	16.592	0.001

¹ Due to a misunderstanding of the demographics portion of the data collection, participants did not consistently report their total years of education.

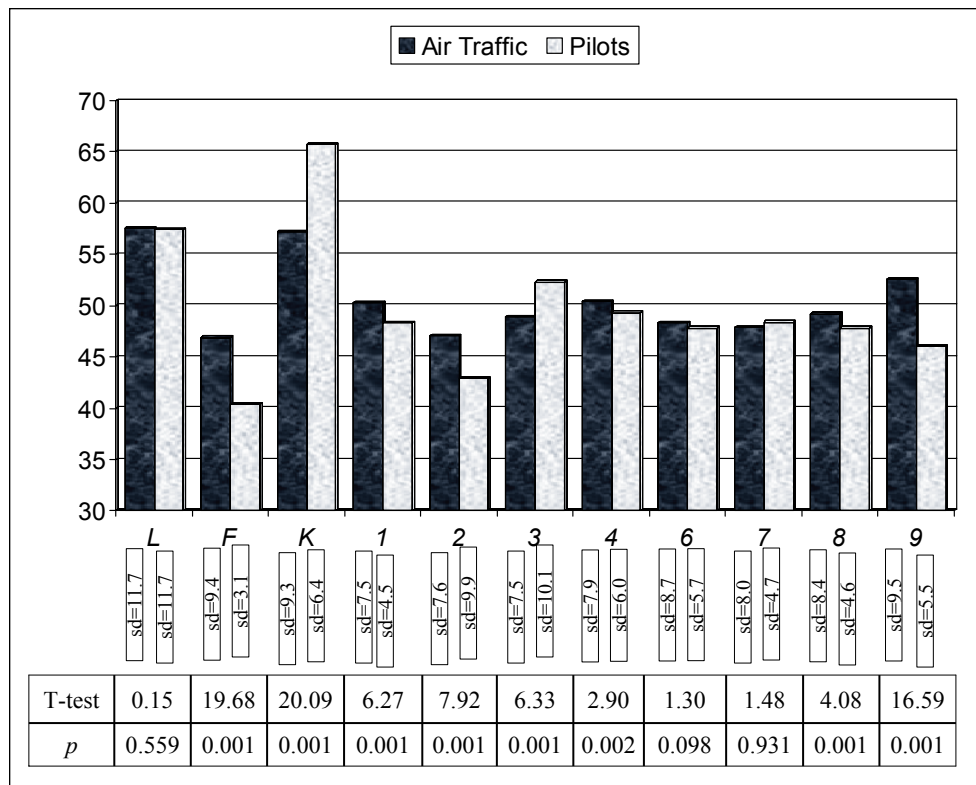


Figure 2. *K*-corrected means and standard deviations for air traffic control specialist (research) and pilot (applicant) samples.

($p < 0.001$). His sample appears to be more defensive than the ATCS sample. It is possible that the validity scales would be more similar if the groups were responding in more analogous settings, rather than research participants (ATCSs) and job applicants (pilots).

Looking at differences on the clinical scales between ATCSs and pilots, there are more similarities than differences. Most scales are within a point or two across the samples. The only apparent differences are on 2 and 9, with both scales being higher in the ATCS sample ($p < 0.001$).

Table 2 provides the means and standard deviations using non-*K*-corrected norms. Here it is noted that scales 7 and 8 are particularly lower than with *K*-correction, compared to the norms. This sample endorsed fewer items that reflect anxiety or disordered thinking than the general population normative group.

Table 2. Non-*K*-corrected means and standard deviations for ATCS sample.

Scale	Mean T (sd)
1	45.7 (8.3)
2	47.1 (7.6)
3	48.9 (7.5)
4	47.5 (7.6)
6	48.3 (8.7)
7	43.0 (8.4)
8	44.1 (9.2)
9	50.8 (9.2)

Table 3. Non- *K*- corrected and non-gender norms: percentages above cut-scores for ATCS sample.

Scale	≥ 65T (%)	≥ 70T (%)	≥ 75T (%)	≥ 80T (%)
1	1.0	0.4	0.2	0.1
2	1.7	0.4	0.2	0.1
3	2.0	0.8	0.3	0.1
4	1.7	0.7	0.3	0.1
6	3.4	2.0	1.0	0.3
7	1.8	0.8	0.3	0.1
8	2.7	2.0	0.8	0.6
9	9.3	4.2	1.6	1.1
1 or more	14.6	7.4	3.2	2.0
2 or more	3.7	2.0	1.0	0.4

In using the MMPI-2 as a screen for ATCS applicants, a practitioner's interest would not be to directly compare the applicant to a group average. Rather, it would be to compare an applicant to established cut scores on the various clinical scales. Table 3, therefore, provides the percentage of participants scoring at or above potential cut-scores of 65T, 70T, 75T, and 80T. At the 65T cut-score level, only about 2-3% of participants are elevated on a given scale. The exception is on scale 9, "*Hypomania*," where about 9% fall at or above this cut-score. These elevations are more likely indicative of this young sample's generally high energy level than reflective of a high number of individuals with possible mood disorders.

As a number of participants scored high on more than one scale, the percentage of participants identified by the test in general is not a simple total of the percentages for each scale. Summing across subjects with one or more scales at or above 65T, we identified about 15% of the participants. Participants with two or more elevated scales represented only about 4% of the sample.

At a less stringent 70T cut-score level, only about half as many participants were identified. Indeed, less than 1% were at or above that level for most scales, with the particular exception of 9, which was at about 4%. About 7% had one or more scales elevated, and only 2% had two or more scales elevated.

The percentage above 75T continued to drop. Very few are identified by individual scales alone. About 3% had one or more high scales, and only 1% had two or more high scales.

Using a cut score of 80T results in very few applicants being identified. Here only 2% had one or more scales highly elevated, and only 0.4% had two or more. Most individual scales were identifying only 0.1%. Differences across the various cut scores point to one of the concerns associated with use of a screening tool. With lower criteria a clinician is likely to identify a higher percentage of individuals who, in fact, would not represent a risk.

In turn, at higher cut scores, a clinician is going to miss a number of individuals who are likely to prove to be at increased risk.

The most relevant statistic here is the percentage of participants with one or more scales at or above the cut-scores. If the 65T cut-score is used, then the assumption is that some 15% of the ATCS applicants have significant psychopathology. This seems high for a relatively high functioning group. After all, this group has demonstrated a high degree of functioning by getting hired, either by demonstrated ability or by passing a rigorous examination. If the 70T cut-score is used, then a psychopathology prevalence rate of about 7% is suggested. This number seems far more consistent with the probable psychopathology rate of this group of people. A cut score of 75T results in about 3% being identified. Finally, a cut score of 80T appears to be so high that only 2% are identified.

Cut Score Decision Process

There are many factors that contribute to the establishment of cut scores. First, the best way to make decisions of this type is to collect research data before hand and compare them with actual multi-year outcomes. Unfortunately, this project has moved forward more quickly than that. As such, a rigorous research program should be delineated very quickly to allow for cut score changes in the future that are data-based.

In the meantime, there are probably two over-arching factors. The first is a practical issue and that is the number of applicants referred for follow-up, in-person evaluations. The second is the a priori probability of the presence of psychopathology within this population.

If applicants are to be referred for follow-up evaluation, the number of referrals must be sufficient to warrant a program, yet low enough to avoid huge costs on the high end. It is probably not worth having a program if fewer than 30 applicants are referred each year. Conversely, at about \$1000 per evaluation, things get expensive and

difficult to manage if there are more than 100 or so per year. If there are 1500 applicants per year, this logic suggests that cut scores resulting in between 2% and 7% would be a good target from a programmatic perspective.

The second approach is the epidemiological approach. Here an a priori estimate of psychopathology in this group is approximated, and the cut scores are set to that level. Some groups have more pathology and some less. In an inpatient psychiatric facility, nearly 100% have some sort of significant pathology. In an outpatient setting, there are many clients who are seeking assistance for such things as marital problems and child problems, so the prevalence of significant psychopathology is probably quite low. In any case, individuals with severe psychopathology are not likely to pursue a career as an ATCS.

The *lifetime* prevalence of mental disorders in adults aged 18 to 54 was 29.4% between 1990 and 1992 and 30.5% between 2001 and 2003 (Kessler et al., 2005). The question that remains unanswered is the likelihood of an employee *developing* mental illness over the course of adulthood; a question that can only be answered by longitudinal research (Reifler, 2006). Use of the MMPI-2, or any other selection instrument, cannot completely eliminate the risk of an employee developing a mental illness over the course of a career, but it can help identify who is *currently* suffering from psychiatric symptoms and is a big step to the longitudinal research envisioned by the Controller Workforce Plan (FAA, 1965).

In samples such as air traffic control specialist candidates, where there are several prior screening hurdles and a certain self-selection process, it is doubtful that the level of true psychopathology is much over 5%. There is probably at least 1% who are experiencing some degree of psychological discomfort, but it is hard to believe that as many as one in 10 would meet criteria to be diagnosed as having a “mental illness.”

Both of these approaches converge on a cut score solution that is very similar. A program should identify and refer between 2% and 7% of applicants. This cut point is programmatically and clinically logical and is likely to strike a balance of false positives and false negatives.

In deciding upon cut scores, there are two possible approaches. The first is to identify different cut scores for different scales. The second is to select a single cut score for all scales. The allure of choosing differing cut scores is that some of the anomalies in the table can be “smoothed out.” For example, Scale 9 seems to pick up more participants than would be clinically suggested. Alternatively, a single cut score for all scales would be the most parsimonious solution but the least sensitive.

Then the question becomes: What cut scores should be used? Here the percentages in the “1 or more” row are the relevant data. If a cut score of 65 is used, 14.6%

would be referred. This rate is probably too high. If a cut score of 80 is used, only 2.0% would be referred. This rate seems too low.

So the discussion should center around the use of either 70 or 75 (or a combination). With 1500 applicants, a cut score of 70 would result in about 100 (7.4%) being sent for follow-up evaluations. This rate is not unreasonable. It also has the benefit of resulting in enough follow-up evaluations for scale-specific outcome purposes.

A cut score of 75 would result in 3.2%, or about 40 of 1500 applicants. This rate would be less costly and easier to manage initially.

It should be remembered that as the program matures and data are collected, the data may support future adjustments in the cut scores. As such, cut scores should not be “fixed in stone.”

Cut Score Decision

MMPI-2 cut scores of 70 appear to best meet the needs of the agency, with the exception of scale 9 where a 75 is more reasonable. Hence, the decision model is to use a cut score of 70 and above for scales 1, 2, 3, 4, 6, 7, and 8. Since scale 9 appears to “over classify” the controllers, that scale will use a cut score of 75 and above. With these cut scores, any applicant with one or more scales above the cut score will be referred for second-tier psychological assessment.

This algorithm will identify 4.9% of applicants. This outcome is well within the parameters suggested. This number is low enough that second-tier costs will not become too great, the management of candidates will not become unruly, and there will not be unacceptably long periods of time during which the medical status of candidates is undetermined.

Looking at the individual scale identifications in Table 4, the percentage found above each cut score is reasonable. The sum of these individual percentages does not result in the overall total of 4.9%, as a number of participants score high on more than one scale.

Table 4. Non-K-corrected and non-gender norms: Percentages above cut-scores for ATCS sample.

Scale	≥ 70T (%)	≥ 75T (%)
1	0.4	n/a
2	0.4	n/a
3	0.8	n/a
4	0.7	n/a
6	2.0	n/a
7	0.8	n/a
8	2.0	n/a
9	n/a	1.6

Applicant Level Reliability

With those applicants identified, the individual records can be examined to determine if the high scores are a function of some non-clinical response set such as random responding. Table 5 provides *F* and *VRIN* for each of the identified male cases, with Table 6 showing the variables for women.

F is considered to be a gross indicator of “faking bad.” This group of participants, however, had no reason to “fake bad,” and so it is probably more indicative of random response to some number of items when high. For males, nine of the 42 identified cases had *F* scale scores above 80. None of the women had an *F* above 80.

Table 5. Male research participants who exceeded clinical cutoffs ($T \geq 70$ on scales 1, 2, 3, 4, 6, 7, 8 and $T \geq 75$ on scale 9) using non-gendered, non-*K*-corrected norms.

Participant Number	<i>F</i>	<i>VRIN</i>	1	2	3	4	6	7	8	9	Code	Mean Elevation
1M	55	61F	50	52	48	47	75	52	47	68	6	55
2M	73	57F	42	45	32	45	46	57	67	77	9	51
3M	107	84T*	61	37	41	53	60	57	73	60	*	*
4M	45	54T	55	35	44	58	46	54	59	85	9	55
5M	67	46F	48	45	53	64	71	40	43	48	6	52
6M	120	73T	55	43	37	58	64	71	91	100	987	65
7M	61	46T	63	62	48	78	71	75	73	60	4786	66
8M	67	61	45	35	39	58	34	49	55	82	9	50
9M	85	73T	57	50	46	64	64	56	76	62	8	59
10M	58	38T	48	56	72	55	56	45	51	54	3	55
11M	73	61	63	58	46	71	82	65	74	68	684	66
12M	58	61F	70	76	62	51	60	57	70	52	218	62
13M	107	80T	76	62	67	67	75	82	90	91	98716	76
14M	120	92T*	70	60	53	43	71	71	80	65	*	*
15M	39	54T	45	43	44	49	71	49	55	60	6	52
16M	64	61T	61	58	50	64	82	53	67	71	6	63
17M	67	57	57	60	60	69	49	74	70	62	7	63
18M	64	54T	66	74	64	87	49	72	67	62	427	68
19M	39	50T	42	31	37	51	36	49	56	77	9	47
20M	64	69F	66	58	62	58	78	57	65	48	6	62
21M	76	42T	65	50	37	78	71	69	76	68	486	64
22M	101	92T*	74	68	62	55	71	67	82	80	*	*
23M	89	54T	57	58	42	58	67	53	72	88	98	62
24M	55	42	72	56	64	62	56	58	62	62	1	62
25M	51	42T	45	37	42	47	36	48	60	88	9	50
26M	55	57F	53	70	48	53	53	56	52	46	2	54
27M	73	65F	63	68	81	64	89	65	74	77	6398	73
28M	51	50	50	43	55	49	71	48	52	50	6	52
29M	55	42F	63	52	74	51	60	42	48	60	3	56
30M	64	50	45	39	41	62	42	53	63	85	9	54
31M	82	61T	86	62	79	69	71	67	72	60	1386	71
32M	73	61T	59	70	57	64	75	78	84	44	8762	66
33M	48	50T	42	37	37	55	56	53	60	85	9	53
34M	64	61T	57	52	50	74	53	56	60	88	94	61
35M	61	50T	68	72	57	45	53	71	70	42	278	60
36M	61	65T	61	58	46	67	75	68	73	77	968	66
37M	42	50	48	43	41	55	46	46	48	77	9	51
38M	64	54F	61	66	55	64	67	64	73	60	8	64
39M	55	50	48	48	41	55	71	53	51	62	6	54
40M	51	54T	57	70	50	60	67	56	64	60	2	61
41M	85	34T	55	66	48	74	75	69	80	74	864	68
42M	70	50	55	62	37	71	71	60	59	57	46	59

* Likely invalid due to *VRIN* > 80T (random responding).

Table 6. Female research participants who exceeded clinical cutoffs ($T \geq 70$ on scales 1, 2, 3, 4, 6, 7, 8 and $T \geq 75$ on scale 9) using non-gendered, non-K-corrected norms.

Participant Number	<i>F</i>	<i>VRIN</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	Code	Mean Elevation
1F	51	50F	57	33	44	45	46	57	60	80	9	53
2F	61	50F	63	70	55	62	49	63	56	52	2	59
3F	48	42T	48	60	64	62	75	53	43	46	6	56
4F	65	66T	78	62	60	60	53	64	67	52	1	62
5F	51	46T	66	80	57	64	49	58	60	46	2	60
6F	61	58T	70	39	57	60	53	53	62	68	1	58
7F	58	50F	48	72	48	69	53	63	51	46	2	56
8F	68	42	50	48	33	47	36	58	70	68	8	51

VRIN, a relatively recently introduced scale, is considered to be a more sensitive indicator of random responding (Graham, 2006). In *VRIN*, very similar items across the test are paired, and the similarity of responses is quantified. Here only three of the 42 males had *VRIN* scores above 80. Again, none of the eight identified women had high *VRIN* scores.

While not all of the high scores are due to random responding, it is not likely that actual applicants taking the test will choose to randomly respond, as their future employment depends upon it. As such, it is probable that the current percentage of identified individuals will “shrink” slightly as the MMPI-2 is used with actual applicants. So, the 4.9% in this sample is likely the upper bounds of the percentage identified when the program is initiated. This rate of identification would be at an acceptable level.

DISCUSSION AND RECOMMENDATIONS

The purpose of this paper was to model the behavior of the MMPI-2 in ATCS applicants for initial psychological test screening. The data suggest that the ATCS participants in this study were remarkably close to the published norms as well as similar to the pilots reported by Butcher (1994). Overall, the vast majority of the participants in this study rendered profiles solidly within normal limits, when compared with the normative sample. It should also be noted that elevated MMPI-2 scales do not necessarily indicate the presence of a disqualifying medical condition, as situational circumstances can result in the temporary elevation of clinical scales without a concomitant presence of a psychiatric condition. Sorting these matters out is the function of the second-tier assessment.

Cut scores of 65, 70, 75, and 80 were applied to the dataset, and the resulting numbers of identified participants were examined for clinical consistency and

programmatic need. Therefore, we recommend using initial cut scores of 70 and above for scales 1, 2, 3, 4, 6, 7, and 8, as well as a cut score of 75 and above for scale 9. These cut scores resulted in no more than 4.9% of the sample being identified as requiring further psychological evaluation.

This approximately 5% appears to be relatively well represented across the eight MMPI-2 scales that are being used. Individual subject analysis suggests that some proportion of participants took the task less than seriously and responded randomly. Actual ATCS applicants, however, will likely be more responsive to the testing situation; thus we expect an identification rate of about 4% to 4.5%. This rate seems acceptable from both a clinical and an administrative perspective.

With between 1200 and 2000 applicants being evaluated per year over the next 10 years, this rate should result in between 48 and 90 follow-up, second-tier evaluations per year. This flow should be sufficient to justify the program but not so great as to require large budgetary requirements and an unmanageable flow of candidates awaiting medical clearance.

A series of research studies are needed to refine the cut score algorithm as applicants go through the hiring and training process. Thus, we offer the following recommendations:

- 1) The applicant data, when the MMPI-2 is used with candidates with a tentative offer of employment, should be immediately analyzed to ensure that the percentages above the cut scores are not vastly different from the research sample. If the “yields” begin to differ, then the cut scores should be reconsidered.
- 2) The MMPI-2 scales should be compared to the data from second-tier assessments and clinician recommendations as they become available. Differential cut scores could be used, for example, if it is shown that some scales tend to predict negative clinician recommendations better than others.

- 3) The scores should also be compared with ATCS training outcomes. While there are many reasons for poor training outcomes, some amount of the outcome is certainly related to psychological functioning.
- 4) The scores should be compared with dismissals and psychological problems that are reported to medical authorities, as well as other outcome measures. While it will take several years to accumulate this type of data on individuals who have taken the MMPI-2, it will be instructive due to the very considerable costs that may be avoided in the future.

REFERENCES

- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders*, 4th Edition Text Revision. Washington, DC: APA.
- Americans With Disabilities Act of 1990, Pub. L. No. 101-336, §2, 104 Stat. 328 (1991). (www.ada.gov/pubs/ada.htm)
- Butcher JN (1994). Psychological assessment of pilots with the MMPI-2. *Journal of Personality Assessment*, 62, 31-44.
- Butcher JN (2002). Assessing pilots with the “wrong stuff”: A call for research on emotional health factors in commercial aviation. *International Journal of Selection and Assessment*, 10, 168-84.
- Butcher JN, Graham JR, Ben-Porath YS, Tellegen A, Dahlstrom WG, & Kaemmer B (2001). *MMPI-2 manual for administering, scoring, and interpreting*. Revised Edition, Minneapolis, MN: NCS Pearson.
- Civil Rights Act of 1991, Public Law No 102-166 (1991). (www.eeoc.gov/policy/cra91.html)
- Convey JJ (1984). Personality assessment of ATC applicants. In SB Sells (Ed.), *Selection of air traffic controllers* (pp. 323-51). (DOT/FAA/AM-84/2). Washington, DC: FAA Office of Aviation Medicine.
- Dollar C, Broach D, & Schroeder D (2003). *Personality characteristics of air traffic control specialists as predictors of disability retirement*. (DOT/FAA/AM-03/14). Washington, DC: FAA Office of Aerospace Medicine.
- Federal Aviation Administration (1965, October 15). *FAA Order 9430.2 – Establishment of health program for agency air traffic control specialist employees*. Washington, DC: U.S. Government Printing Office.
- Federal Aviation Administration (1980, May 9). *FAA Order 3930.3A - Air traffic control specialist health program*. Washington, DC: U.S. Government Printing Office.
- Geiger TC & Crick NR (2001). A developmental psychopathology perspective on vulnerability to personality disorder. In RE Ingram & JM Price (Eds.), *Vulnerability to psychopathology* (pp. 57 – 102). New York: The Guilford Press.
- Graham JR (1990). *MMPI-2: Assessing personality and psychopathology*, 2nd Edition. New York: Oxford University Press.
- Graham JR (2006). *MMPI-2: Assessing personality and psychopathology*, 4th Edition. New York: Oxford University Press.
- Hammen C (2001). Vulnerability to depression in adulthood. In RE Ingram & JM Price (Eds.), *Vulnerability to psychopathology* (pp. 226-57). New York: The Guilford Press.
- Harvey PD (2001). Vulnerability to schizophrenia in adulthood. In RE Ingram & JM Price (Eds.), *Vulnerability to psychopathology* (pp. 355-81). New York: The Guilford Press.
- Kessler RC, Demler O, Frank RG, Olfson M, Pincus HA, Walters EE, Wang P, Wells KB, & Zaslavsky AM (2005). Prevalence and treatment of mental disorders, 1990 to 2003. *The New England Journal of Medicine*, 352, 2515-23.
- King RE, Retzlaff PD, Detwiler CA, Schroeder DJ, & Broach D (2003). *Use of personality assessment measures in the selection of air traffic control specialists*. (DOT/FAA/AM-03/20). Washington, DC: FAA Office of Aerospace Medicine.
- Lowman RL (1989). *Pre-employment screening for psychopathology: A guide to professional practice*. Sarasota, FL: Professional Resource Exchange, Inc.
- McNally RJ, Malcarne VL, & Hansdottir I (2001). Vulnerability to anxiety disorder across the lifespan. In RE Ingram & JM Price (Eds.), *Vulnerability to psychopathology* (pp. 322- 25). New York: The Guilford Press.
- Reifler CB (2006). Epidemiological aspects of college mental health. *Journal of American College Health Association*, 54, 372-6.
- Schwarzkopf A, Buckley MR, & Pace, TM (1996). *16PF screen evaluation* (DTFA-02-93-D-93088). Unpublished manuscript, University of Oklahoma.

Smith RC (1971). *Personality assessment in aviation: An analysis of the item ambiguity characteristics of the 16PF and MMPI* (DOT/FAA/AM-71/35). Washington, DC: FAA Office of Aviation Medicine.

Trites DK (1961). *Problems in air traffic management: I. Longitudinal prediction of effectiveness of air traffic controllers* (DOT/FAA/AM-1). Washington, DC: FAA Office of Aviation Medicine.

Westefeld JS & Maples M (1998). The MMPI-2 and vocational assessment: A brief report. *Journal of Career Assessment*, 6, 107-13.

APPENDIX A

Excerpt from FAA Order 3930.3A

CHAPTER 4. INITIAL HIRE

40. GENERAL. The medical standards for initial employment prescribed under the Physical Requirements section of the Office of Personnel Management (OPM) Qualification Standards for the Air Traffic Control Series 2152 shall be applied to all applicants for initial ATCS employment.

41. PSYCHOLOGICAL TESTING. A comprehensive psychological test battery shall be administered to all ATCS applicants at the time of their preemployment interview or during the interval between the interview and the medical examination. The administration of this test battery shall be the responsibility of local facility chiefs; test materials shall be provided by the Flight Surgeons. Completed answer sheets shall be sealed by the applicant, and transmitted by the facility to the Federal Air Surgeon: Attention AAM500. TEST RESULTS SHALL NOT BE THE SOLE BASIS FOR REJECTION OF AN ATCS

APPLICANT. Personnel who conduct this testing shall instruct applicants to refrain from discussing the content of tests with other applicants.

42. SUPPLEMENTAL MEDICAL INFORMATION. The Flight Surgeon shall obtain and evaluate applicable military and Veterans Administration medical records through established regional procedures. If military medical records are not received within 120 days following request, the Flight Surgeon may grant conditional medical clearance pending receipt and review of the additional medical

g. Psychiatric.

The applicant must have no established medical history or clinical diagnosis of any of the following:

- (1) A psychosis;
- (2) A neurosis;
- (3) Any personality disorder or mental disorder that the Federal Air Surgeon determines clearly indicates a potential hazard to safety in the Air Traffic Control System. The determinations will be based on the medical case history (including past social, and occupational adjustment) supported by clinical psychologists and board certified psychiatrists, including such psychological tests as may be required as a part of medical evaluation as the Federal Air Surgeon may prescribe.

h. Substance Dependency.

A history, review of all available records, clinical and laboratory examination will be utilized to determine the presence or absence of substance dependency, including alcohol, narcotic, and nonnarcotic drugs. Wherever clinically indicated, the applicant must demonstrate an absence of these on thorough psychiatric evaluation, including any clinical or psychological tests required as part of the medical evaluation.

