Screening Air Traffic Control Specialists with the MMPI-2: Two New Scales to Increase Predictive Utility

Roger L. Greene[1]
David S. Nichols[1]
Chris M. Front[2]
Raymond E. King[2]

[1]Clinical Psychologist, Retired

[2]Office of Aerospace Medicine Medical Specialties Division

February 2021

Final Report

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-21/03	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Screening Air Traffic Control Specialists With the MMPI-2: Two New Scales to Increase Predictive Utility		5. Report Date February 2021
		6. Performing Organization Code
7. Author(s) Roger Greene,[1] David Nichols,[1] Chris Front,[2] Raymond King[2]		8. Performing Organization Report No.
9. Performing Organization Name and Address [1]Clinical Psychologist, Retired [2]Office of Aerospace Medicine Medical Specialties Division Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591		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 Final Report, April 2020 – February 2021
		14. Sponsoring Agency Code

15. Supplemental Notes

16. Abstract

The FAA began using the Minnesota Multiphasic Personality Inventory – 2 (MMPI-2) for the psychological screening of Air Traffic Control Specialists (ATCSs) in January 2008 after a research effort at the Civil Aerospace Medical Institute (CAMI) involving 1,014 participants. Subsequently, norms were compiled for the first 5,500 ATCS candidates who took the MMPI-2 as a screening test for operational purposes. This report examines the operational use of the MMPI-2 from 2008 to 2020. During that time, the FAA administered the MMPI-2 to a total of 20,385 ATCS candidates. The present study had two goals: (1) To re-calculate MMPI-2 operational norms with the larger ATCS candidate sample in order to discover whether any updates to those norms were needed, and (2) to empirically examine the MMPI-2 scales that have most frequently resulted in referrals for further psychological assessment with the aim of reducing the number of such referrals by improving the predictive utility of the MMPI-2 screenings. A reduction in false positive findings at the time of screening, without increasing false negatives, would reduce the number of referrals for complete psychological assessments, thereby maintaining aviation safety while contributing to a more efficient and cost-effective overall medical clearance process for ATCS candidates. The operational norms remain virtually unchanged. The resulting scales, *FAA-L* and *FAA-RCTY*, will serve to identify those applicants whose pattern of item endorsement warrants a closer review and will be adjusted as operational experience with them accumulates.

17. Key Words Personnel Screening, Personnel Selection, Psychological Testing	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 39	22. Price

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

ACKNOWLEDGMENTS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

EXECUTIVE SUMMARY

The FAA began using the Minnesota Multiphasic Personality Inventory – 2 (MMPI-2) for the psychological screening of Air Traffic Control Specialists (ATCSs) in January 2008 after a research effort at the Civil Aerospace Medical Institute (CAMI) involving 1,014 participants (King, Schroeder, Manning, Retzlaff, & Williams, 2008). Subsequently, norms were compiled for the first 5,500 ATCS candidates who took the MMPI-2 as a screening test for operational purposes. This report examines the operational use of the MMPI-2 from 2008 to 2020. During that time, the FAA administered the MMPI-2 to a total of 20,385 ATCS candidates. The present study had two goals: (1) To re-calculate MMPI-2 operational norms with the larger ATCS candidate sample in order to discover whether any updates to those norms were needed and (2) To empirically examine the MMPI-2 scales that have most frequently resulted in referrals for further psychological assessment with the aim of reducing the number of such referrals by improving the predictive utility of the MMPI-2 screenings. A reduction in false positive findings at the time of screening, without increasing false negatives, would reduce the number of referrals for complete psychological assessments, thereby maintaining aviation safety while contributing to a more efficient and cost-effective overall medical clearance process for ATCS candidates. The operational norms remain virtually unchanged. The resulting scales, *FAA-L* and *FAA-RCTY*, will serve to identify those applicants whose pattern of item endorsement warrants a closer review and will be adjusted as operational experience with them accumulates.

Screening Air Traffic Control Specialists with the MMPI-2: Two New Scales to Increase Predictive Utility

INTRODUCTION

The FAA began using the Minnesota Multiphasic Personality Inventory – 2 (MMPI-2) for the psychological screening of Air Traffic Control Specialists (ATCSs) in January 2008. The MMPI-2 was selected as the most appropriate screening measure based on decades of evidence of its efficacy with pilots (Butcher, 1994) and findings from an investigation conducted with Federal Aviation Administration (FAA) ATCS employees (King, Schroeder, Manning, Retzlaff, & Williams, 2008). The operational use of the MMPI-2 with ATCS applicants was codified in FAA Order 8500.4 (2019a). This report examines the operational use of the MMPI-2 from 2008 to 2020. During that time, the FAA administered the MMPI-2 to a total of 20,385 ATCS candidates. In an internal study conducted in 2011, Paul Retzlaff, Chris Front, and Ray King calculated normative scores for the first 5,500 ATCS candidates to take the MMPI-2 as a screening test. Those norms were later published by Butcher, Front, and Ones (2018) in their discussion of the many issues involved in assessing psychopathology in high-risk populations, including ATCS. The present study had two goals: (1) To re-calculate MMPI-2 norms with the larger ATCS candidate sample in order to discover whether any updates to the norms established by Retzlaff and colleagues were needed and (2) To empirically examine the MMPI-2 scales that have most frequently resulted in referrals for further psychological assessment with the aim of reducing the number of such referrals by improving the predictive utility of the MMPI-2 screenings. A reduction in false positive findings at the time of screening, without increasing false negatives, would reduce the number of referrals for complete psychological assessments, thereby maintaining aviation safety while contributing to a more efficient and cost-effective overall medical clearance process for ATCS candidates.

THE FAA PSYCHOLOGICAL SCREENING PROCESS

The psychological selection process for ATCSs employed by the FAA is divided into two phases. The select-in phase ensures a good fit between the applicant's knowledge, skills, abilities, and aptitudes and the ATCS job duties. The select-out phase identifies any disqualifying psychological conditions that would pose a hazard in the conduct of ATCS duties. Requirements imposed by the Americans with Disabilities Act (ADA, 1991) dictate that medical examinations cannot occur until after a conditional job offer has been made. Moreover, any psychological testing or other assessment designed to identify psychopathology is considered a medical procedure under the ADA.

After the applicant has completed the application process and received a conditional job offer, the select-out procedures can begin. An aviation medical examination is conducted to ensure that the applicant meets the medical qualifications specified by the FAA. The psychological screening aspect of the medical examination—referred to as the "Tier 1" assessment—is conducted using the MMPI-2 per FAA Order 8500.4A (2019a). A Licensed Clinical Psychologist in the Office of Aerospace Medicine at FAA Headquarters in Washington, DC, reviews the resulting Tier 1 MMPI-2 profiles. Those candidates who produce valid MMPI-2 profiles with no scores of concern are designated as having "Cleared" the Tier 1 psychological screening and move forward in the overall medical clearance process. If no other disqualifying medical conditions are identified, the Flight Surgeon responsible for the medical examination of the applicant medically clears the applicant for continuation in the hiring process. Those candidates who

produce an invalid Tier 1 MMPI-2 profile or who produce a valid profile with scores of concern are designated as having "Not Cleared" the Tier 1 screening assessment. Their medical clearance process is paused, and the candidate is referred for a complete psychological assessment (i.e., a "Tier 2" assessment).

The Tier 2 psychological assessments are performed by licensed clinical psychologist consultants with expertise in personality assessment who have been trained by the FAA Clinical Psychologists. The Tier 2 assessments include a review of records (including the Tier 1 MMPI-2 score report and any other pertinent records, such as military records, police and court records, medical records, and so forth), a thorough psychosocial/clinical interview, and mental status examination. The Tier 2 assessment also involves psychological testing that must include, at minimum, a re-administration of the MMPI-2 and administration of a Personality Assessment Inventory (PAI; Morey et al., 2007) and/or Millon Multiaxial Clinical Inventory-IV (MCMI-IV; FAA, 2019b), as well as any other tests that the Tier 2 psychologist considers to be clinically indicated. The completed Tier 2 assessments are reviewed by one of the FAA Clinical Psychologists to determine whether a disqualifying condition in accordance with FAA Order 3930.3C has been identified. If no disqualifying condition is identified, then the candidate is designated "Qualified" from a psychological standpoint; that finding is communicated to the Flight Surgeon with medical clearance authority and the medical clearance proceeds. If, however, a disqualifying psychological condition is identified, then the candidate is designated "Disqualified" with regard to their psychological status; that finding is communicated to the Flight Surgeon, and the medical clearance is halted. Candidates who are medically disqualified may, if they choose, appeal to the Federal Air Surgeon. An appeal to the Federal Air Surgeon is a separate process that is outside the scope of this paper. Figure 1 presents a graphic summary of this process.
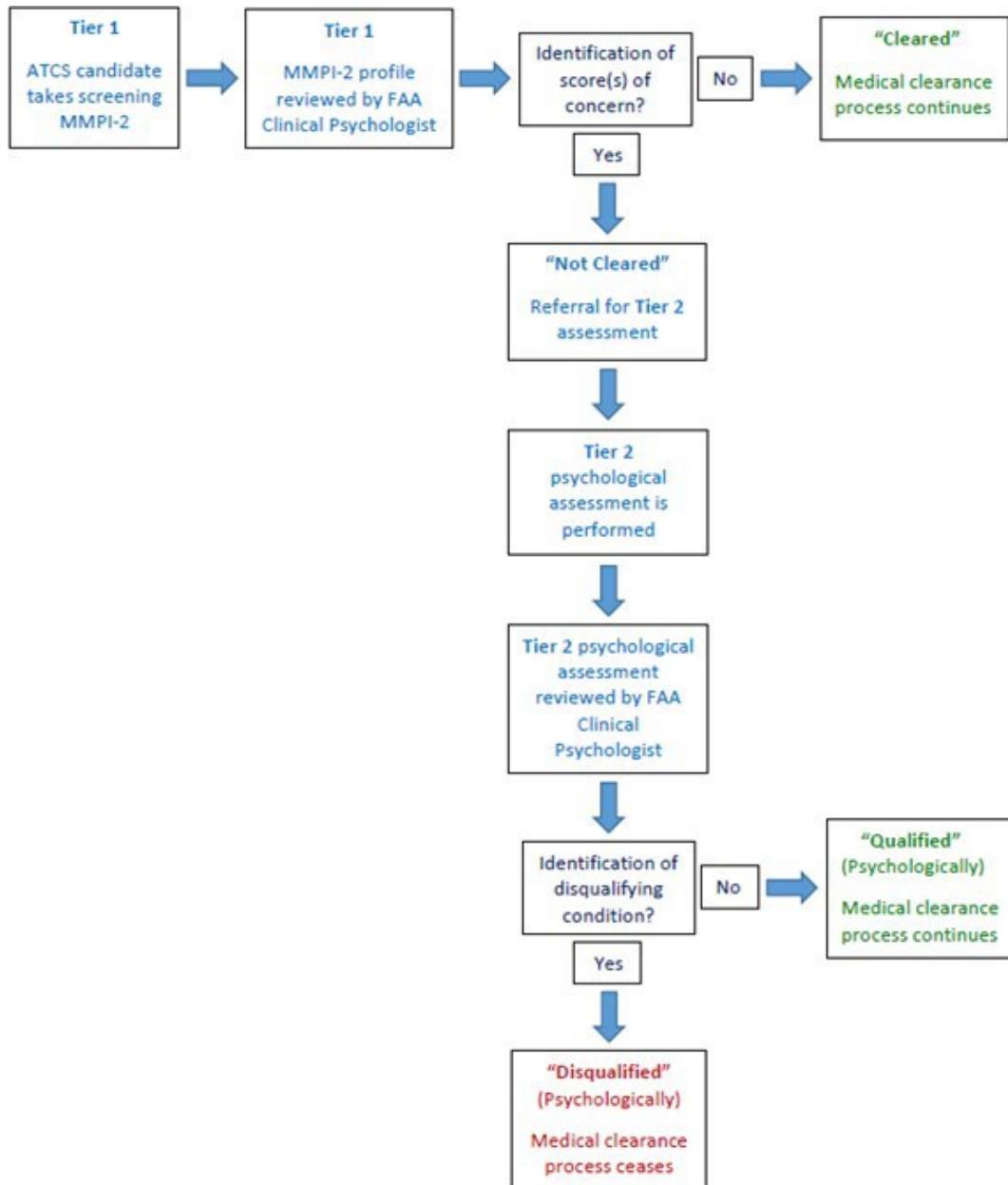
Note. ATCS = Air Traffic Control Specialists; FAA = Federal Aviation Administration; MMPI-2 = revised Minnesota Multiphasic Personality Inventory.

Figure 1. Psychological Screening Component of Medical Clearance Process for FAA ATSC Candidates

Table 1 illustrates the performance of the total pool of tested ATCS candidates, along with the findings of Retzlaff et al. (2011). There have been no significant changes in the mean scores across this entire period. Most of the differences between mean scores on the MMPI-2 scales (**d** score) are less than ±.10, which is one-tenth of a standard deviation. A **d** score of at least .30 is used as the criterion for whether a difference is of statistical significance, and a **d** score of .50 is used to suggest a difference large enough

to be of clinical importance (e.g., five T-points on the MMPI-2). The differences between means of nongender and gender T-scores in the 2020 sample also are $\leq \pm .10$ and are not included in Table 1. Nongender T-scores were used in all analyses as required by federal guidelines, as outlined first by Title VII of the Civil Rights Act of 1964.

In personnel screening settings, some of the validity scales (e.g., *Lie [L]*, *Correction [K], and Superlative [S]*) are commonly elevated to T-scores around 60, and such elevations are, therefore, to be expected with ATCS candidates. These elevations reflect a complex interplay between factors such as positive impression management, defensiveness, and the legitimate expression of better-than-average coping resources (Front, 2020; Williams & King, 2010). However, these factors mean that the K-correction of the clinical scales should *not* be used in personnel screening settings. Instead, the non-K-corrected versions of the clinical scales are more appropriate.

The Psychiatric *F* scale (*Fp*) was developed to identify unfavorable self-descriptions in clinical settings. However, the *Fp* scale contains four items overlapping the *L* scale. The *L* scale is used to identify self-favorable self-descriptions, and these items are endorsed frequently in personnel screening settings. Consequently, elevations on the *Fp* scale need to be interpreted cautiously when the *L* scale is elevated by more than five or six items because such an *Fp* elevation may reflect either a favorable or an unfavorable self-description. This faux pas reflects the blind use of a methodology to identify items for a new scale without adequate review of the actual item content before finalizing the scale.

Table 1. Descriptive Statistics for ATCS Candidates 2011 and 2020

	2011 Gender T-Scores (N = 5,500)			2020 Gender T-Scores (N = 20,385)		2020 vs 2009		2020 Nongender T-Scores (N = 20,385)	
Scale	**M**	**SD**	**Scale**	**M**	**SD**	**d**	**Scale**	**M**	**SD**
Validity Scales									
VRIN	40.10	7.70	VRIN	40.37	8.24	-.03	VRIN	40.27	8.47
TRIN	53.20	4.60	TRIN	53.29	4.71	-.02	TRIN	53.38	4.84
F	43.30	5.80	F	43.59	5.92	-.05	F	44.37	6.04
FB	43.60	3.40	FB	43.72	3.72	-.03	FB	43.69	3.67
Fp	48.90	8.00	FP	48.44	8.11	.06	Fp	49.35	8.08
FBS	45.90	7.50	FBS	44.44	7.69	.19	FBS	43.18	7.13
L	58.10	12.90	L	57.93	13.07	.01	L	58.20	13.38
K	60.00	8.50	K	59.31	8.77	.08	K	59.60	8.91
S	63.20	9.90	S	60.22	11.02	.27	S	60.14	11.18
Clinical Scales (K-corrected)									
Hs	47.30	5.90	Hs	47.06	6.09	.04	Hs	46.45	5.90
D	45.10	6.30	D	45.89	6.60	-.12	D	44.80	6.30
Hy	48.50	6.70	Hy	47.62	6.82	.13	Hy	46.74	6.56
Pd	50.00	7.20	Pd	49.26	7.04	.11	Pd	49.44	7.18
Mf-F	63.00	9.10	Mf-F	62.46	9.62	.06			
Mf-M	40.20	7.30	Mf-M	40.57	7.73	-.05			
Pa	47.10	7.40	Pa	46.98	7.51	.02	Pa	46.87	7.47
Pt	46.80	6.00	Pt	47.09	6.36	-.05	Pt	46.41	6.30
Sc	47.30	5.90	Sc	47.54	6.10	-.04	Sc	47.74	6.12
Ma	50.80	8.60	Ma	50.79	8.44	.00	Ma	51.46	8.47
Si	40.60	7.20	Si	42.06	8.17	-.18	Si	41.54	7.93
Clinical Scales (NonK-corrected)									
Hs	40.00	6.10	Hs	40.19	6.47	-.03	Hs	39.80	6.37
Pd	45.90	6.80	Pd	45.56	6.80	.05	Pd	46.11	6.57
Pt	39.90	6.80	Pt	40.68	7.30	-.11	Pt	40.16	7.11
Sc	40.20	7.10	Sc	40.98	7.40	-.10	Sc	41.00	7.41
Ma	48.60	8.40	Ma	48.76	8.23	-.02	Ma	49.37	8.40
Restructured Clinical Scales									
RCd	40.40	5.50	RCd	40.69	5.93	-.05	RCd	40.67	5.54
RC1	41.60	5.70	RC1	41.83	6.00	-.04	RC1	41.14	6.05

RC2	40.20	6.30	RC2	40.84	6.85	-.09	RC2	40.66	6.98
RC3	45.00	8.90	RC3	45.28	8.67	-.03	RC3	45.52	8.97
RC4	43.20	7.60	RC4	42.99	7.49	.03	RC4	44.36	7.79
RC6	44.80	7.10	RC6	45.00	7.25	-.03	RC6	45.97	6.84
RC7	39.50	7.10	RC7	39.85	7.39	-.05	RC7	39.44	6.96
RC8	45.10	7.30	RC8	45.15	7.51	-.01	RC8	45.24	7.41
RC9	44.80	8.50	RC9	44.27	7.91	.07	RC9	44.95	8.35

Content Scales

ANX	42.10	6.60	ANX	42.42	6.87	-.05	ANX	41.70	6.81
FRS	45.20	8.00	FRS	45.52	8.51	-.04	FRS	43.70	7.51
OBS	39.70	7.10	OBS	39.85	7.58	-.02	OBS	39.80	7.27
DEP	40.70	5.80	DEP	40.82	6.08	-.02	DEP	40.14	5.92
HEA	41.60	6.50	HEA	41.85	6.71	-.04	HEA	41.58	6.38
BIZ	44.90	6.90	BIZ	45.01	7.07	-.02	BIZ	45.44	7.37
ANG	40.20	6.60	ANG	39.90	6.61	.04	ANG	39.93	6.53
CYN	43.80	8.30	CYN	44.18	8.31	-.05	CYN	44.60	8.51
ASP	45.20	8.70	ASP	45.59	8.60	-.04	ASP	46.91	8.91
TPA	41.70	6.80	TPA	40.89	7.34	.11	TPA	41.27	7.61
LSE	40.10	6.20	LSE	40.87	6.92	-.11	LSE	40.48	6.63
SOD	42.50	8.30	SOD	44.32	9.67	-.19	SOD	44.23	9.77
FAM	42.70	7.30	FAM	42.47	7.24	.03	FAM	41.82	7.04
WRK	38.40	6.00	WRK	38.87	6.52	-.07	WRK	38.26	6.42
TRT	40.00	5.70	TRT	40.10	6.09	-.02	TRT	40.00	5.94

Supplementary Scales

A	39.80	5.20	A	40.16	5.78	-.06	A	39.92	5.48
R	51.50	7.80	R	50.85	8.39	.08	R	50.00	8.82
Es	58.60	5.80	Es	55.55	8.61	.35	Es	56.85	8.02
Do	53.50	6.40	Do	50.59	7.44	.39	Do	50.77	7.61
Re	55.50	8.70	Re	52.98	10.52	.24	Re	52.25	10.83
Mt	40.10	5.60	Mt	40.38	5.91	-.05	Mt	40.00	5.76
PK	41.60	5.90	PK	42.23	6.39	-.10	PK	42.11	6.25
MDS	43.30	6.50	MDS	43.32	6.41	.00	MDS	43.08	6.70
Ho	43.60	8.50	Ho	43.56	8.65	.00	Ho	44.06	8.79
O-H	60.30	9.00	O-H	57.97	10.15	.23	O-H	56.92	10.10
MAC-R	49.50	7.50	MAC-R	48.15	7.72	.17	MAC-R	49.79	8.14

AAS	43.50	6.30	AAS	43.36	6.35	.02	AAS	44.79	6.33
APS	43.90	9.20	APS	42.75	9.47	.12	APS	42.87	9.50
GM	59.60	6.90	GM	55.36	11.36	.37	GM		
GF	48.60	8.10	GF	46.83	9.58	.18	GF		
PSY-5 Scales									
AGGR	48.30	6.60	AGGR	46.15	6.94	.31	AGGR	47.31	7.04
PSYC	42.20	7.30	PSYCH	42.38	7.56	-.02	PSYCH	42.60	7.54
DISC	48.70	8.80	DISC	47.26	9.01	.16	DISC	50.10	8.96
NEGE	39.50	7.00	NEGE	39.55	7.21	-.01	NEGE	38.51	7.17
INTR	42.90	7.40	INTR	43.74	8.01	-.10	INTR	43.92	8.25
Harris-Lingoes Subscales									
D1	41.50	5.20	D1	42.16	5.95	-.11	D1	41.76	5.68
D2	49.70	7.80	D2	50.52	7.96	-.10	D2	49.93	8.01
D3	47.40	7.40	D3	47.89	7.77	-.06	D3	47.14	7.47
D4	41.20	4.60	D4	41.66	5.14	-.09	D4	42.15	4.64
D5	42.00	4.90	D5	42.53	5.54	-.10	D5	41.95	5.20
Hy1	56.80	7.20	Hy1	55.89	7.83	.12	Hy1	56.50	8.14
Hy2	56.60	9.20	Hy2	55.72	9.52	.09	Hy2	55.72	9.58
Hy3	41.40	4.90	Hy3	41.40	5.15	.00	Hy3	41.80	4.65
Hy4	41.30	4.70	Hy4	41.41	4.99	-.02	Hy4	41.11	4.47
Hy5	52.00	8.70	Hy5	51.56	8.71	.05	Hy5	51.10	8.81
Pd1	47.30	6.50	Pd1	46.73	6.42	.09	Pd1	46.32	6.36
Pd2	49.60	8.20	Pd2	49.49	8.19	.01	Pd2	51.70	8.46
Pd3	56.50	7.10	Pd3	55.44	7.64	.14	Pd3	55.95	7.42
Pd4	44.90	7.20	Pd4	45.09	7.49	-.03	Pd4	44.55	7.62
Pd5	42.80	6.90	Pd5	42.87	6.99	-.01	Pd5	43.10	6.86
Pa1	47.20	6.60	Pa1	47.74	6.97	-.08	Pa1	47.69	7.22
Pa2	43.00	6.90	Pa2	42.70	6.94	.04	Pa2	42.06	6.77
Pa3	55.30	9.50	Pa3	54.67	9.57	.07	Pa3	54.44	9.62
Sc1	44.10	6.50	Sc1	44.86	7.02	-.11	Sc1	44.03	6.95
Sc2	46.50	5.90	Sc2	46.63	5.82	-.02	Sc2	46.16	5.56
Sc3	44.50	4.40	Sc3	44.96	5.04	-.09	Sc3	45.10	5.43
Sc4	43.40	4.50	Sc4	43.65	4.72	-.05	Sc4	43.65	4.72
Sc5	45.80	6.70	Sc5	46.53	7.00	-.10	Sc5	46.43	6.74

Sc6	45.10	5.90	Sc6	45.34	6.14	-.04	Sc6	44.98	5.91
Ma1	45.80	6.70	Ma1	50.69	9.05	-.54	Ma1	52.01	9.14
Ma2	46.50	8.80	Ma2	46.74	8.53	-.03	Ma2	46.83	8.83
Ma3	56.30	8.90	Ma3	55.44	9.34	.09	Ma3	56.38	9.70
Ma4	47.40	8.80	Ma4	47.23	8.74	.02	Ma4	47.22	8.56
Si1	43.40	7.80	Si1	44.90	8.62	-.17	Si1	44.48	8.34
Si2	44.90	8.10	Si2	46.41	9.15	-.17	Si2	47.03	9.31
Si3	41.40	6.90	Si3	42.01	7.16	-.08	Si3	42.00	7.10

Content Component Scales

			FRS1	45.90	6.01		FRS1	45.10	5.48
			FRS2	46.55	9.34		FRS2	44.29	8.45
			DEP1	43.45	5.07		DEP1	43.49	5.01
			DEP2	44.08	5.03		DEP2	43.20	4.53
			DEP3	42.74	4.89		DEP3	42.93	4.91
			DEP4	45.11	1.59		DEP4	46.11	1.53
			HEA1	46.33	5.89		HEA1	46.20	5.21
			HEA2	43.31	5.59		HEA2	43.10	5.01
			HEA3	47.82	7.16		HEA3	48.77	7.06
			BIZ1	45.69	4.76		BIZ1	45.69	4.78
			BIZ2	45.70	7.20		BIZ2	46.00	7.35
			ANG1	42.39	5.64		ANG1	42.69	6.09
			ANG2	39.56	6.45		ANG2	39.52	6.25
			CYN1	45.31	9.47		CYN1	45.93	9.49
			CYN2	42.55	8.10		CYN2	43.58	8.24
			ASP1	45.45	9.32		ASP1	46.90	9.42
			ASP2	46.37	8.89		ASP2	43.38	18.10
			TPA1	39.96	6.87		TPA1	36.54	14.63
			TPA2	41.35	8.20		TPA2	37.74	15.56
			LSE1	42.03	5.68		LSE1	42.97	5.53
			LSE2	44.48	6.56		LSE2	44.10	6.49
			SOD1	45.48	9.22		SOD1	47.00	9.74
			SOD2	44.14	8.54		SOD2	44.31	8.43
			FAM1	42.11	7.64		FAM1	42.39	7.54
			FAM2	46.86	6.75		FAM2	47.85	6.64
			TRT1	42.90	3.48		TRT1	43.07	3.33
			TRT2	40.96	6.33		TRT2	42.21	6.61

Note. Abbreviations for all scale names are presented on pages 5-6.

Table 2 summarizes the Tier 1 MMPI-2 scales that were elevated beyond the cutting score for each scale (i.e., the scale elevation[s] that resulted in the candidate being "Not Cleared" after Tier 1 and referred for a Tier 2 assessment). Some self-favorability response distortion is expected in anyone who is applying for a personnel position. The validity scales *L*, *K*, and *S* are measures of response distortion of the self-favorability flavor. Nearly one-third of these applicants exceeded the cutting scores that are expected in a personnel screening situation on these three scales. Another set of scales that assess psychopathology (*Scale 9 [Ma]*: 28.92%; *MacAndrews Alcoholism Scale, Revised [MAC-R]*: 26.23%; Scale 7 *[Pt]*: 9.07%; Scale 4 *[Pd]*: 8.58%; Addiction Potential Scale *[APS]*: 8.58%; Scale 6 *[Pa]*: 8.33%) were elevated in 8% to 30% of these applicants. These two sets of scales, positive impression management and possible psychopathology, were the primary focus of this report.

Table 2. Summary of Failed Outcome Criteria From Tier 1 Screening

Failed Criterion	%	Qualified	%	Disqualified	%
(N = 3,556)		(n = 3,183)		(n = 408)	
L	1484	41.73	1379	43.32	135	33.09
K	1089	30.62	1019	32.01	100	24.51
S	1114	31.33	1048	32.92	96	23.53
F	60	1.69	81	2.54	9	2.21
Fp	89	2.50	108	3.39	11	2.70
FB	21	.59	50	1.57	1	.25
TRIN	55	1.55	81	2.54	4	.98
VRIN	12	.34	42	1.32	0	.00
Hs	41	1.15	64	2.01	7	1.72
D	103	2.90	114	3.58	19	4.66
Hy	31	.87	58	1.82	3	.74
Pd	174	4.89	169	5.31	35	8.58
Pa	153	4.30	149	4.68	34	8.33
Pt	153	4.30	146	4.59	37	9.07
Sc	156	4.39	156	4.90	30	7.35
Ma	506	14.23	418	13.13	118	28.92
AAS	89	2.50	99	3.11	20	4.90
APS	220	6.19	215	6.75	35	8.58
MAC-R	411	11.56	334	10.49	107	26.23
Other	148	4.16	141	4.43	38	9.31

Note. Abbreviations for all scale names are presented on pages 5-6.

CREATION OF FAA NONGENDER T-SCORES FOR FAA SCALES

At the time of this study, a total of 20,385 applicants had been psychologically screened with the MMPI-2. A total of 3,119 women and 17,266 men were in the pool of applicants. To attain an equal number of men and women in the FAA Nongender normative sample, 3,119 men were randomly selected from the larger pool of men. The resulting sample of 6,238 individuals was used to create the weights and T-scores for the FAA Nongender normative sample scales.

Base Rates

The "Disqualified" sample (**N** = 380) is much smaller than the Tier 2 sample (**N** = 3,314; 11.47%) and the Total sample (**N** = 18.041; 2.10%). Predicting that all Tier 2 Applicants will be "Qualified" will be accurate for 88.53% of these applicants. The task of identifying the 11.47% who will ultimately be "Disqualified" (on Tier 2) is daunting, at best. However, the task of identifying those applicants who do not *need* to be referred for a Tier 2 evaluation is much easier. The following sections will describe the development of two new scales which will refine the accurate identification, at Tier 1, of those ATCS candidates who should be referred for a Tier 2 assessment while also more accurately identifying those who do not need such referral. The first of the two scales, the *FAA L* scale, is expected to reduce the number of unnecessary Tier 2 referrals while the second scale, the *FAA Reactivity* scale, will assist in identifying those candidates who do, in fact, need more careful scrutiny in the form of a Tier 2 assessment.

Positive Impression Management: L Scale

The *Lie* (*L*) scale is composed of 15 items that were selected on a rational basis to identify persons who are deliberately trying to avoid answering the MMPI frankly and honestly (Dahlstrom et al., 1972; Ben-Porath & Tellegen, 2008). The scale assesses attitudes and practices that are culturally laudable but found only in the most conscientious persons. The content areas within the *L* scale deny minor moral failings/lapses, personal dishonesties and denial of aggression, bad thoughts, and weakness of character. The *L* scale has been well substantiated as a measure of positive impression management that is used routinely in personnel selection.

Table 3 provides the distribution of Nongender *T* and Z-scores on the *L* scale in all FAA samples of applicants. Figures 2 and 3 illustrate this same information for T-scores and Z-scores, respectively. The scores on the *L* scale increase precipitously at a T-score of 75 (Z-score of 2.50) in the Qualified and Disqualified samples because that was the cutting score on the *L* scale at Tier 1. Interestingly, the Disqualified sample shows less of an increase than the Qualified samples. Slightly over 50% of the Qualified sample (53.25%) have T-scores of 76 or higher (Z-scores of 2.51 or higher), as compared with 35.27% of the Disqualified sample. This may be because a subset of the ultimately Disqualified sample produced valid profiles (i.e., low or moderate elevations on *L*, *K*, and *S*), combined with elevations of one or more clinical and/or substance use scales at Tier 1. Therefore, the elevation of L for the sample as a whole was attenuated. The Qualified sample thus appears to be working harder to make a positive impression than the Disqualified sample, but that may not be the case. This result appears to suggest that individuals who are identified at Tier 1 due to an elevated *L* scale will ultimately be found aeromedically qualified. However, it is also clear that certain examinees who elevate the *L* scale are found to have disqualifying psychopathology. Hence, elevations of the *L* scale cannot be dismissed without risking a false-negative finding.

Table 3. Distribution of Nongender T and Z-Scores on the L Scale in all FAA Samples of Applicants

Sample					Nongender · T-Score					
	30-40	**41-50**	**51-55**	**56-60**	**61-65**	**66-70**	**71-75**	**76-80**	**81-85**	**86+**
Normative	7.74%	24.22%	15.64%	12.68%	11.06%	14.95%	4.81%	3.58%	2.76%	2.57%
Total	7.16	23.22	14.32	12.86	10.76	15.81	5.50	4.30	3.04	3.02
Tier 2	12.72	19.74	6.33	4.16	2.28	2.45	1.13	20.06	14.91	16.21
Qualified	13.03	18.22	6.17	3.72	1.99	2.51	1.11	20.99	15.31	16.95
Disqualified	10.33	31.49	7.56	7.56	4.53	2.01	1.26	12.85	11.84	10.58
Z-score	**-2.00~-1.00**	**-.99~.00**	**.01~.50**	**.51~1.00**	**1.01-1.50**	**1.51-2.00**	**2.01-2.50**	**2.51-3.00**	**3.01-3.50**	**3.51+**

Note. FAA = Federal Aviation Administration.
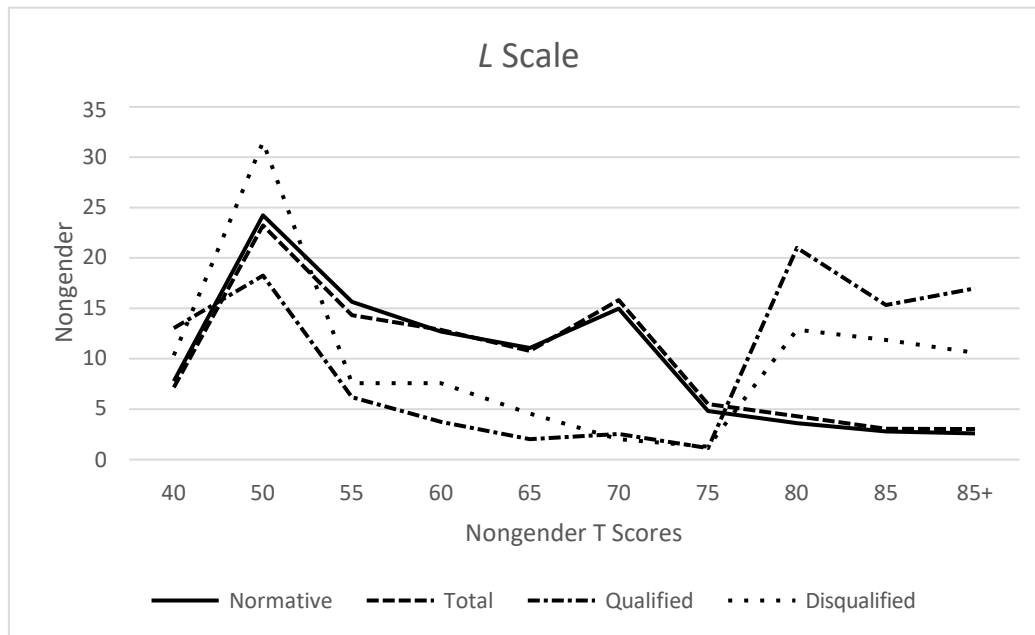
Note. FAA = Federal Aviation Administration.

Figure 2. Distribution of Nongender T-Scores on the L Scale for All FAA Samples
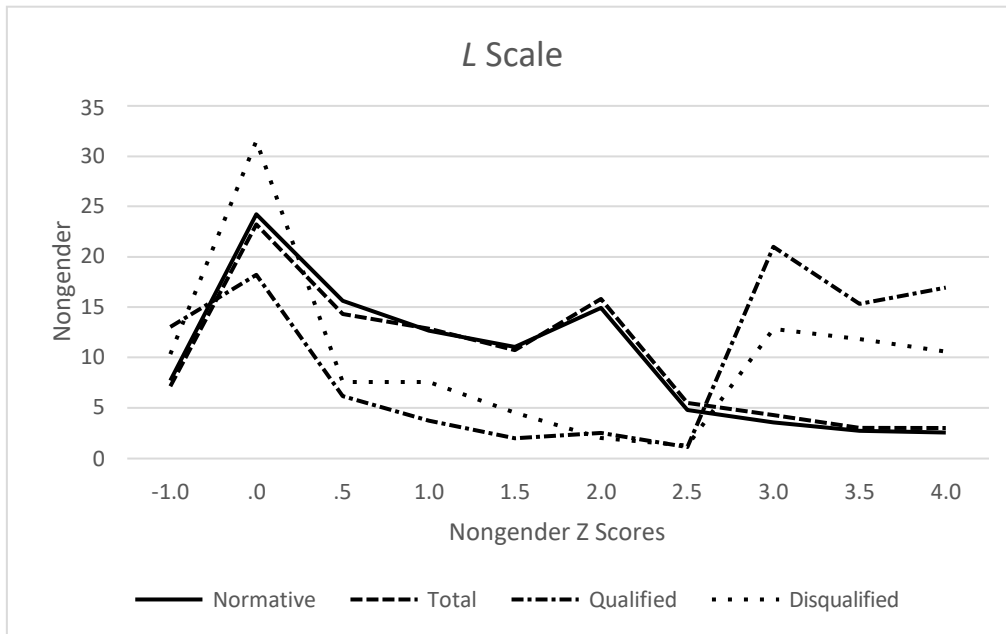
L Scale

Note. FAA = Federal Aviation Administration.

Figure 3. Distribution of Nongender Z-Scores on the L Scale for All FAA Samples

Positive Impression Management: FAA-L Scale

There are several *L* scale items (139, 203, 260) that are endorsed infrequently by the FAA applicants and the MMPI-2 normative group (Table 4). Several other *L* items (16, 41, 123, 232) are endorsed by most of the FAA applicants and approximately 50% of the MMPI-3 normative group (16, 41, 123, 153). The frequency of someone endorsing the individual items on the *L* scale varies widely, yet each item is scored identically (i.e., given equal weight) on the MMPI-2. That is, one point is added to the person's raw score on the *L* scale for each item endorsed as "False." To increase the accuracy of classification of Disqualified versus Qualified on the *L* scale, the frequency of "False" responses to the *L* scale items in the FAA nongender normative sample was used as the item weight to score each item on the *FAA L* scale (Table 5). Higher weights for the items on the *FAA L* scale reflect how unusual it is for those items to be endorsed as "False," or how intent the applicant is on trying to make a positive impression. Lower weights reflect that the item is commonly endorsed by these applicants. It is the former group of applicants (i.e., those who are endorsing the items with the higher weights) that raises the most concern about whether they are trying to conceal significant levels of psychopathology. The descriptive statistics from the FAA Nongender normative sample also are provided in Table 5 for the *FAA-L* scale. The MMPI-2 Restructured Form (MMPI-2-RF) dropped four of the *L* scale items (no. 16, 93, 123, and 153) from their Uncommon Virtues (*L-r*) scale (Nichols & Greene, 1988). Because these four items are among those with higher weights on the *FAA-L* scale, the loss of these items from *L-r* suggests that the MMPI-2-RF would be less useful than the MMPI-2 in the FAA and other similar screening settings.

Table 4. Frequency of Endorsement of L Scale Items by the FAA Nongender Normative Sample and the MMPI-2 Normative Group

Item	FAA Nongender Normative Sample (N = 6,238) %F	MMPI-2 Normals (N = 2,600) %F
16	81.98	57.54
29	57.74	17.27
41	70.79	40.73
51	41.52	4.92
77	17.52	5.42
93	55.29	10.54
102	39.23	2.58
107	23.58	33.31
123	81.98	64.92
139	8.16	9.77
153	57.53	49.92
183	24.21	10.42
203	14.12	13.54
232	75.25	26.88
260	19.51	5.81

Note. FAA = Federal Aviation Administration; MMPI-2 = Revised Minnesota Multiphasic Personality Inventory.

Table 5. Item Weights for the FAA-L Scale

Item	FAA Nongender Normative Sample (N = 6,238) %F	FAA-L Scale Item Weight
16	81.98	.180
29	57.74	.423
41	70.79	.292
51	41.52	.585
77	17.52	.825
93	55.29	.447
102	39.23	.608
107	23.58	.764
123	81.98	.180
139	8.16	.918
153	57.53	.425
183	24.21	.758
203	14.12	.859
232	75.25	.248
260	19.51	.805

FAA-L Scale	
Mean	2.77
SD	1.47

T Score	Z Score	Raw Score
30-40	-2.00~-1.00	1.3-2.76
41-50	-.99~.00	2.77-3.49
51-55	.01~.50	3.50-4.23
56-60	.51~1.00	4.24-4.96
61-65	1.01-1.50	4.97-5.70
66-70	1.51-2.00	5.71-6.43
71+	2.01+	6.44+

Note. FAA = Federal Aviation Administration.

Table 6 illustrates the differences in scoring procedures for the *L* and *FAA-L* scales. Both cases endorsed six of the *L* scale items for a total score of 6. Case 1 endorsed the six items with the highest weights on the *FAA-L* scale for a total score of 4.929, while Case 2 endorsed the six items with the lowest weights for a total score of 1.747. Using the chart in Table 5, Case 1 has a T-score in the 56–60 range on the *FAA-L* scale, while Case 2 is in the 30–40 range.

Table 6. Scoring Examples for the L and FAA-L Scales

Case 1 Item	%F	L Scale Weight	*FAA-L* Scale Item Weight
77	17.52	1	.825
107	23.58	1	.764
139	8.16	1	.918
183	24.21	1	.758
203	14.12	1	.859
260	19.51	1	.805
Total Score	6	4.929	
Case 2 Item	**%F**	**L Scale Weight**	***FAA-L* Scale Item Weight**
16	81.98	1	.180
29	57.74	1	.423
41	70.79	1	.292
123	81.98	1	.180
153	57.53	1	.425
232	75.25	1	.248
Total Score		6	1.747

Note. FAA = Federal Aviation Administration.

Table 7 provides the distribution of Nongender T and Z-scores on the *FAA-L* scale for all the FAA samples of applicants. Figures 4 and 5 illustrate this same information for T-scores and Z-scores, respectively. A T-score of 60 (Z-score of 1.0) on the *FAA-L* scale seems to reflect where the Qualified and Disqualified samples start to differ from the FAA normative Nongender sample that was partially the result of using a cutting score of 10 (raw) on the *L* scale at Tier 1. However, as noted in Table 6 above, it is critical to know which items on the *L* scale are being endorsed. Table 8 shows the overlap between T-scores and Z-scores for the *L* and *FAA-L* scales when the scales are used jointly. Any applicants who have a T-score of 75 (Z-score: 2.50) or lower on the *L* scale, and have a T-score below 60 (Z-score: 1.00) on the *FAA-L* scal*e*, are endorsing the less extreme items on the *L* scale and might *not* need to be assessed at Tier 2. Among the 469 applicants with T-scores from 71 to 75 (Z-score: 2.01–2.50) on the *L* scale, 429 (91.47%) have a T-score from 55 to 60 (Z-score: 0.51–1.00), and 40 (8.53%) applicants have T-Scores from 61-65 (Z-scores: 1.01–1.50) on the *FAA-L* scale. Among the 1,055 applicants with T-scores from 76 to 80 (Z-scores: 2.51–3.00) on the *L* scale, 873 (82.75%) have a T-score from 55 to 60 (Z-score: 0.51–1.00) and 178 (16.87%) applicants have a T-score from 61–65 (Z-score: 1.01–1.50) on the *FAA-L* scale. Substantial cost savings will result if these 873 applicants with T-scores from 76 to 80 (Z-scores: 2.51–3.00) on the *L* scale and a T-score from 55 to 60 (Z-score: 0.51–1.00) on the *FAA-L* scale do not need to be assessed at Tier 2. If it is assumed that each Tier 2 assessment costs $2,000, then eliminating 873 evaluations would save $1,746,000.

When T-scores on the *L* scale are 81 or higher (Z-scores: 3.01+), the relationships between the *L* scale and *FAA-L* scale reverse, and the majority are above a T-score of 60 (Z-score of 1.01) on the *FAA L* scale. These results show the potential ranges in which cutting scores can be selected for the *L* and *FAA-L* scales. Applicants who have a T-score of 80 or higher (Z-score: 3.01) on the *L* scale and a T-score above 60 (Z-score: 1.01) on the *FAA-L* scale would seem to warrant a closer review. Of course, these criteria can be adjusted up or down as needed.

Table 7. Distribution of Nongender T and Z-Scores on the FAA-L Scale for All FAA Applicant Samples

	Nongender T-Score								
Sample	**30-40**	**41-50**	**51-55**	**56-60**	**61-65**	**66-70**	**71-75**	**76-80**	**81-85**
Normative	23.68%	20.65%	23.95%	17.99%	8.13%	3.53%	1.49%	.50%	.10%
Total	23.14	19.77	24.99	18.08	8.27	3.74	1.46	.48	.08
Tier 2	10.68	17.95	20.22	12.58	14.03	13.97	7.48	2.53	0.54
Qualified	9.84	14.21	26.40	16.34	12.86	12.09	5.46	2.21	.59
Disqualified	16.58	29.21	16.05	10.00	10.79	10.26	5.79	1.05	.26
Z-score	**-2.00~-1.00**	**-.99~.00**	**.01~.50**	**.51~1.00**	**1.01-1.50**	**1.51-2.00**	**2.01-2.50**	**2.51-3.00**	**3.01-3.50**

	Nongender Z-score								
	Mean	2.77							
	SD	1.47							
T Score	**Z Score**	**Raw Score**							
30-40	**-2.00~-1.00**	1.30							
41-50	**-.99~.00**	2.77							
51-55	**.01~.50**	3.50							
56-60	**.51~1.00**	4.24							
61-65	**1.01-1.50**	4.97							
66-70	**1.51-2.00**	5.71							
71+	**2.01+**	6.44							
		7.18							

Note. FAA = Federal Aviation Administration; SD = standard deviation.

Table 8. Overlap of T-Scores and Z-Score Ranges for the L Scale and FAA-L Scale

L Scale	FAA L Scale				
T-scores	**56-60**	**61-65**	**66-70**	**70+**	**Total**
71-75	429	40	0	0	469
76-80	873	178	4	0	1055
81-85	125	454	40	14	619
86-90	0	0	219	75	233
91+	0	0	0	12	12
Z-scores	**.51~1.00**	**1.01-1.50**	**1.51-2.00**	**2.01+**	

Note. FAA = Federal Aviation Administration.
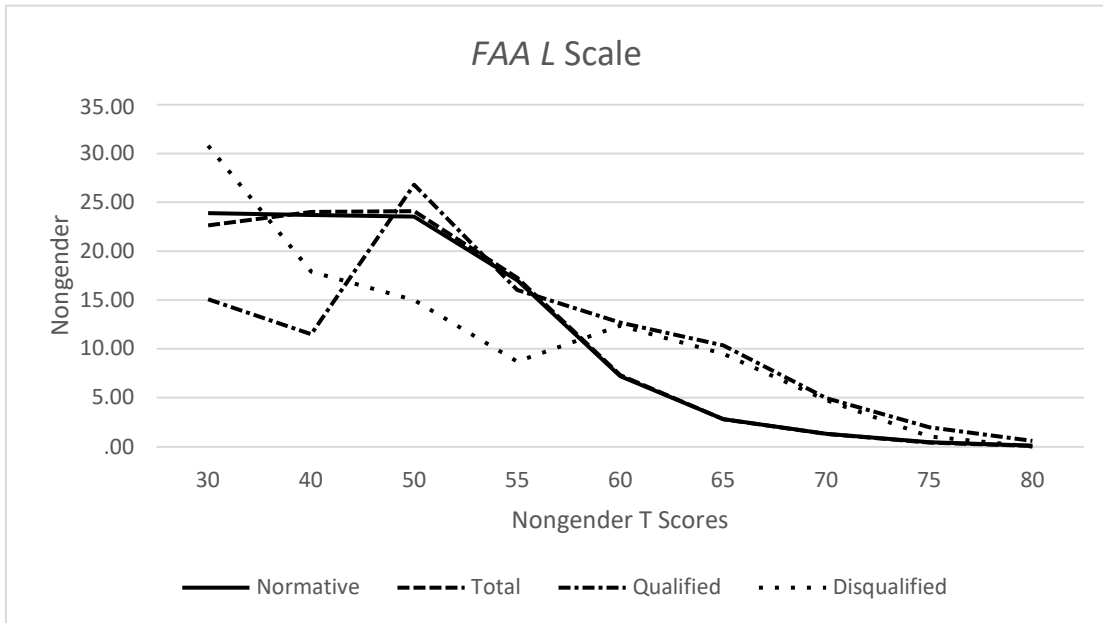
Note. FAA = Federal Aviation Administration.

Figure 4. Distribution of Nongender Z-Scores on the L Scale for All FAA Samples
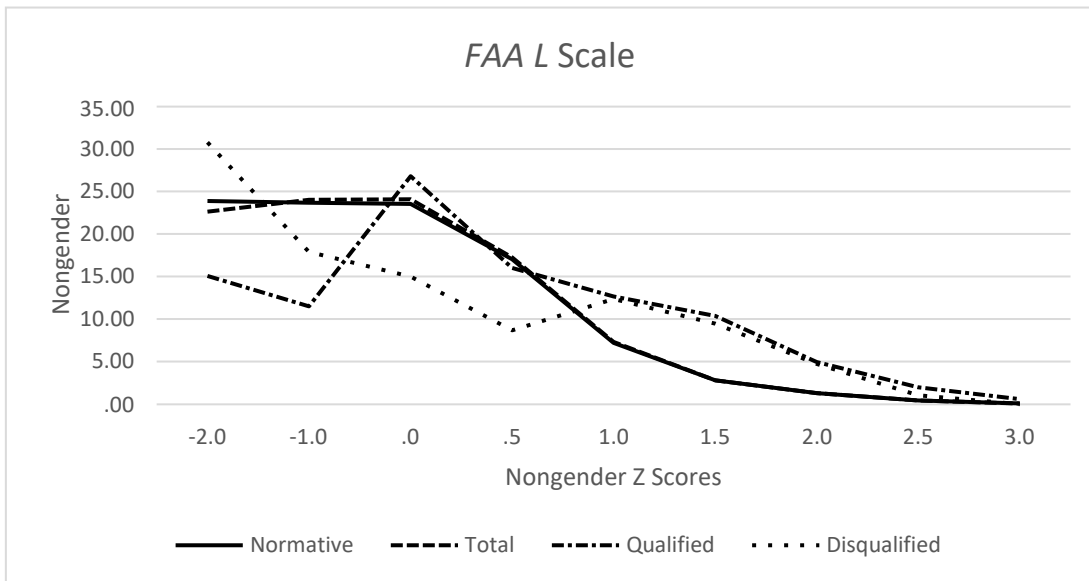
Note. FAA = Federal Aviation Administration.

Figure 5. Distribution of Nongender Z-Scores on the L Scale for All FAA Samples

Consideration of Other Potential MMPI-2 Items to Identify Self-Favorable Distortion

As noted above, the *K* and *S* scales, along with the *L* scale, are used to identify individuals who seek to respond very self-favorably when they take the MMPI-2. The *K* and *S* scales were evaluated to see if any items might be added to the *FAA L* scale to make it more robust and reliable. Although a few potential items were identified, an inspection of item content showed that none of these items were appropriate. Table 9 shows the items on the *K* and *S* scales that differentiated between the FAA Nongender Normative Group and the Disqualified Group. The content of the *K* and *S* items favored by the Disqualified Group proved very different from the content of the *L* scale items. The content areas within the *L* scale include denial of minor moral failings or lapses, personal dishonesties, denial of aggression, bad thoughts, and weakness of character. In contrast, the content areas of these potential items from the *K* and *S* scales included acknowledging inappropriate behaviors, alienation, irritability, cynicism, disinhibition, and self-preoccupation.

The MMPI-2 items that are not found on the *L*, *K*, and *S* scales also were correlated with the *L* scale in a further attempt to identify any potential items that might be added to the *FAA-L* scale. However, the best correlating items, like many of the item correlates of *K* and *S,* described above, emphasized disinhibition, irritability, cynicism, and discontent. None of the items gathered on this basis appeared thematically consistent with the items of the *L* scale, or with positive impression management. All of these items were endorsed "True" by the Disqualified sample except items 127, 158, and 243 on the *K* scale and item 560 on the *S* scale.

These findings support the perspective proposed by Nichols and Greene (1988, 1997) and expanded by Front (2020) that the *L* scale taps into a different aspect of self-favorability response distortion compared to the *K* and *S* scales. Research on response distortion and attempts to measure its expression on self-report measures has led to useful classifications. Expanding on work by Wiggins (1964), Paulhus (1984, 1986, 1998) proposed a 2-factor model to characterize the response distortions that comprise socially desirable responding. Paulhus distinguished between unconscious, personality-based, *self*-deception (consistent with Wiggins' *Alpha*) and conscious, goal-oriented, *other*-deception (consistent with Wiggins' *Gamma*). Paulhus developed and validated a self-report inventory—the Paulhus Deception Scales (PDS; 1998)—that reliably measures the relative contributions of each subtype to the combined, total distortion present in an individual's pattern of responses. Paulhus used the term "Self-Deceptive Enhancement" (SDE) to describe the unconscious distortion that results when a respondent displays a tendency to bias inventory responses based in an actual belief about one's positive qualities such that the responses are believed to be accurate, true, and justified but are, nevertheless, a distorted exaggeration of actual abilities. In contrast, Paulhus (1998) used the term "Impression Management" (IM) to describe the conscious and intentional distortion employed in efforts to promote a favorable impression of the self.

Paulhus (1984) demonstrated that only the IM factor is sensitive to changes in a social setting. Unlike SDE, IM scores are higher when obtained in a social/interpersonal context than when responding anonymously. In contrast, SDE appears to be more stable across contexts. Hence, IM is a form of conscious response distortion that may or may not be employed tactically based upon contextual demands, whereas SDE is a form of response distortion stemming from an enduring, trait-based construct.

Based upon well-established normative personality profiles among aerospace personnel, Front (2020) proposed that moderate elevations on scales such as Paulhus' SDE scale and MMPI-2 scales (*K* and *S*) that tap into the SDE construct are expected and acceptable when screening aerospace personnel. In

contrast, elevations of Paulhus' IM scale and the MMPI-2 *L* scale must be viewed with caution and further scrutiny. The *FAA-L* scale provides a new tool in that regard.

Table 9. K and S Scale Items That Differentiated the FAA Nongender Normative Group and Disqualified Group

Item	*K* Scale Disqualified -Normative Nongender Difference
37	15.76
346	15.09
267	15.02
365	12.97
122	12.92
348	12.76
284	12.37
290	12.10
127	-11.92
110	11.83
58	11.16
158	-11.07
338	10.88
341	10.75
243	-10.71
330	10.67
196	10.23

Item	*S* Scale Disqualified -Normative Nongender Difference
445	18.23
374	18.18
423	17.09
373	16.01
547	15.37
486	15.34
346	15.09
104	14.23
449	13.84
81	13.25
264	12.98
225	12.74
523	12.72
560	-12.54
284	12.37
442	12.15
290	12.1
487	11.83
110	11.83
542	11.45
430	11.42

58	11.16
15	10.77
341	10.75
196	10.23

MMPI-2 SCALES ASSESSING PSYCHOPATHOLOGY

Table 2 summarizes the MMPI-2 scales that were elevated beyond the cutting score at Tier 1 and resulted in the Tier 2 referrals. The validity scales (*L*, *K*, *S*), measures of self-favorability, were reviewed above. The scales that assess psychopathology are reviewed in this section, with a focus on three scales that assess various aspects of behavioral disinhibition: *Scale 9 Ma*, MacAndrews Alcoholism Scale – Revised *MAC-R*, and Disconstraint *DISC*.

Scale 9 *Ma*

McKinley and Hathaway (1944) used an empirical approach to construct the MMPI scales. The items had to be answered differently by the criterion group (e.g., manic patients) compared with normal groups. Their approach was strictly empirical (i.e., no theoretical rationale was posited as the basis for accepting or rejecting items on a specific scale). Consequently, it is not always possible to discern why an item distinguished the criterion group from normal groups. Instead, items were selected solely because the criterion group answered them differently than other groups. The criterion group for Scale 9 *Ma* consisted of 24 manic patients of moderate or mild severity because more severe cases would not cooperate with testing (McKinley & Hathaway, 1944). The item endorsements of this criterion group were contrasted with those of the original Minnesota normative group to develop the 46-item *Ma* scale. This methodology results in empirical scales containing a heterogeneous set of items, many of which bear no obvious relationship with the construct to be assessed. To identify the core of the *Ma* scale in the FAA applicants, item-total correlations were obtained for all the items. A criterion of a correlation greater than ±.30 produced 24 items, slightly over half (52.1%) of the items on the *Ma* scale. Table 10 shows the 10 items that had at least a 10% difference between the Disqualified and FAA Nongender normative samples. This criterion of at least a 10% difference between groups was a standard used by McKinley et al. and has been used routinely in creating MMPI scales. Ten items is *not* a sufficient number to make an *FAA Ma* scale. One item (250) on the *Ma* scale overlaps with the *DISC* scale that will be discussed below.

Table 10. Scale 9 *Ma*

Item	T/F	Disqualified - Normative Difference
169	T	18.08
55	T	17.38
250	F	16.77
182	T	-13.18
122	T	12.92
211	T	12.46
253	F	11.50
212	T	11.04
15	T	10.77
243	T	10.71
168	T	8.17
229	T	-7.79
100	T	7.51
13	T	6.48
50	T	5.10
87	T	4.66
106	T	4.62
248	F	-4.00
190	T	2.90
131	T	-2.22
218	T	1.80
238	T	1.77
85	T	1.70
205	T	1.30

Note. Ma = hypomania.

MacAndrews Alcoholism – Revised MAC-R **Scale**

The MacAndrews Alcoholism – Revised *MAC-R* scale was also created empirically, so all the issues raised above for the *Ma* scale also apply to the *MAC-R* scale, and will not be repeated. To identify the core of the *MAC-R* scale in the FAA applicants, item-total correlations were obtained for all 49 items. A criterion of a correlation greater than ±.30 yielded 35 items, which is 71.1% of the items on the *MAC-R* scale. Table 11 shows the 11 items that had at least a 10% difference between the Disqualified and FAA Nongender normative samples. These 11 items were not sufficient to make an *FAA MAC-R* scale. Four items on the *MAC-R* scale overlapped with the *DISC* scale.

Table 11. *MAC-R* Scale

	Disqualified - Normative Sample	
Item	**T/F**	**Difference**
414	**T**	19.31
445	**T**	18.23
266	**T**	-18.11
84	**T**	18.11
224	**T**	14.11
412	**T**	14.10
103	**T**	12.96
549	**F**	12.01
172	**T**	-10.92
502	**F**	10.50
280	**T**	10.08

Note. MAC-R = MacAndrews alcoholism.

Disconstraint DISC **Scale**

The PSY-5 Disconstraint *DISC* scale (Harkness & McNulty, 1994) was created using factor analysis, so it consists of a single set of homogenous items. Table 12 shows the 14 items that had at least a 10% difference between the Disqualified and FAA nongender normative samples. These 14 items were not sufficient to make an *FAA DISC* scale. As noted above, four items on the *DISC* scale overlapped with items on the *MAC-R* scale and one item matched with an item with the *Ma* scale.

Table 12. *DISC* Scale

	Disqualified - Normative Sample	
Item	**T/F**	**Difference**
477	**T**	21.79
417	**T**	21.48
35	**T**	19.61
266	**F**	-18.11
84	**T**	18.11
418	**T**	16.97
385	**T**	16.86
250	**T**	16.77
123	**T**	15.93
412	**T**	14.10
103	**T**	12.96
284	**T**	12.37
402	**T**	11.70
126	**T**	10.52

Note. DISC = disconstraint.

FAA Reactivity (FAA-RCTY) Scale

All nonredundant items from the *Ma*, *MAC-R*, and *DISC* scales that showed at least a 10% difference separating the Disqualified sample from the FAA Nongender normative sample were combined with any other items from the MMPI-2 item pool that met the same criterion to form the preliminary version of the *FAA Reactivity* (*FAA-RCTY*) scale. *L* scale items were deleted to eliminate any redundancy with the *FAA L* scale. Seventeen items were dropped from this pool of items because they did not separate the Disqualified sample from the Qualified sample (Table 13). These 17 items have content that seems troublesome when read, but the FAA data reveal that they are not relevant for identifying the Disqualified applicants. Six of these deleted items are scored on the *Ma* scale, six items on the *MAC-R* scale, and five items on the *DISC* scale. Eliminating these 17 items will make the *FAA-RCTY* scale more sensitive to the behaviors that distinguish the Disqualified applicants from the other applicants.

The *FAA-RCTY* scale (Table 14) consists of 53 items: 47 keyed "True" and six "False." The keying of the items is not indicated to protect the integrity of the MMPI-2. The FAA-RCTY scale is sensitive to the manifestations of psychopathology that occur with some frequency among FAA ATCS employment candidates. The content and patterns of these items reflect a theme of dyscontrol, elements of impetuous thought, impulsiveness, emotional overarousal, and behavioral reactivity. The *FAA-RCTY* scale shares six items with the *Ma* scale, 10 items with the *MAC-R* scale, and nine items with the *DISC* scale (Table 15). Several items are scored in the opposite direction on the *FAA-RCTY* scale and the *Ma* scale (n = 2), *MAC-R* scale (n = 1), and *DISC* scale (n = 1). The *FAA-RCTY* scale will serve to identify those applicants whose pattern of item endorsement warrants a closer review of historical behavior patterns. Table 16 provides the T-scores and Z-scores for the *FAA-RCTY* scale by sample (represented graphically in Figures 6 and 7, respectively). The Disqualified sample starts having higher scores than the Qualified sample around a T-score of 55 (Z-score: .50), and this would seem like a potential criterion that the applicant warrants a Tier 2 assessment (Table 15). A T-score of 55 (Z-score of .50) is exceeded by 28.11% of the Qualified sample and 50.00% of the Disqualified Sample. Of course, this criterion can be adjusted up or down as needed.

Table 13. Items Deleted from Consideration for the *FAA-RCTY* Scale

Item	T/F	True %	True %
15	**T**	32.18	33.95
37	**T**	19.49	26.58
55	**T**	26.58	34.21
103	**T**	35.11	38.95
211	**T**	35.33	40
212	**T**	23.65	28.95
225	**T**	25.72	26.84
250	**T**	24.86	31.05
253	**T**	22.16	24.47
264	**T**	42.19	37.11
271	**T**	47.43	44.74
373	**T**	33.14	38.68
402	**T**	23.88	30
408	**T**	22.26	28.16
412	**T**	17.29	25
502	**T**	11.65	17.37
549	**T**	12.63	18.28

Note. FAA-RCTY = Federal Aviation Administration-Reactivity.

Table 14. *FAA-RCTY Scale*

	Qualified	**Disqualified**	**Normative**	**Disqualified - Nongender**
Item	**True %**	**True %**	**True %**	**Difference**
21	17.73	29.21	15.93	13.28
33	72.52	60.79	76.79	-16.00
35	15.12	35.00	15.39	19.61
81	27.03	40.26	27.01	13.25
84	13.09	26.84	8.74	18.11
86	47.92	59.74	47.16	12.57
104	29.60	37.63	23.40	14.23
113	29.56	55.79	38.30	17.49
115	39.35	57.89	43.75	14.15
122	36.55	54.21	41.30	12.92
126	49.01	71.05	60.53	10.52
169	29.11	44.21	26.13	18.08
172	44.81	26.32	37.24	-10.92
182	35.60	22.11	35.28	-13.18
224	52.30	73.68	59.57	14.11
242	39.12	52.37	38.17	14.20
243	53.25	32.11	42.82	-10.71
266	54.33	41.32	59.43	-18.11
267	30.05	44.74	29.72	15.02
280	43.14	64.47	54.39	10.08
284	27.21	40.79	28.42	12.37
286	17.92	28.42	14.04	14.38
304	42.78	50.79	38.12	12.67

329	51.49	27.37	40.08	-12.71
334	39.21	21.05	34.80	-13.75
345	62.86	76.32	63.35	12.96
346	26.81	36.32	21.22	15.09
365	41.88	63.16	50.19	12.97
374	25.33	41.32	23.14	18.18
385	52.05	68.95	52.08	16.86
386	16.03	29.74	14.88	14.86
390	17.88	30.53	14.77	15.76
393	13.81	22.63	9.83	12.80
396	15.26	30.53	16.16	14.36
398	30.16	51.32	37.43	13.89
406	16.48	33.16	17.96	15.20
410	21.26	28.42	14.93	13.49
414	30.16	40.53	21.22	19.31
417	42.35	56.84	35.36	21.48
418	35.98	46.84	29.87	16.97
419	22.66	37.11	20.83	16.27
422	56.16	78.16	62.33	15.83
423	23.02	35.79	18.70	17.09
434	26.64	49.47	31.27	18.20
440	36.30	57.63	44.26	13.37
445	26.64	39.74	21.50	18.23
477	30.88	53.16	31.37	21.79
481	38.74	59.74	43.54	16.20
486	22.93	40.79	25.45	15.34
501	48.40	75.00	57.59	17.41
507	11.69	20.53	7.17	13.36
518	18.74	28.68	10.84	17.84
523	16.75	31.84	19.12	12.72

Note. FAA-RCTY = Federal Aviation Administration-Reactivity.

Table 15. Items from the *Ma, MAC-R, DISC* Scales on the *FAA-RCTY* Scale

Ma			**MAC-R**			**DISC**		
Item	**T/F**	**Scored**	**Item**	**T/F**	**Scored**	**Item**	**T/F**	**Scored**
			84	**T**	Same	35	**T**	Same
113	**T**	Same	113	**T**	Same	84	**T**	Same
122	**T**	Same	172	**T**	Opposite	126	**T**	Opposite
169	**T**	Same	224	**T**	Same	284	**T**	Same
182	**T**	Opposite	280	**T**	Same	385	**T**	Same
242	**T**	Same	414	**T**	Same	417	**T**	Same
243	**F**	Opposite	422	**T**	Same	418	**T**	Same
			434	**T**	Same	477	**T**	Same
			445	**T**	Same	266	**F**	Same
			266	**F**	Same			
Overlap	6			10				9

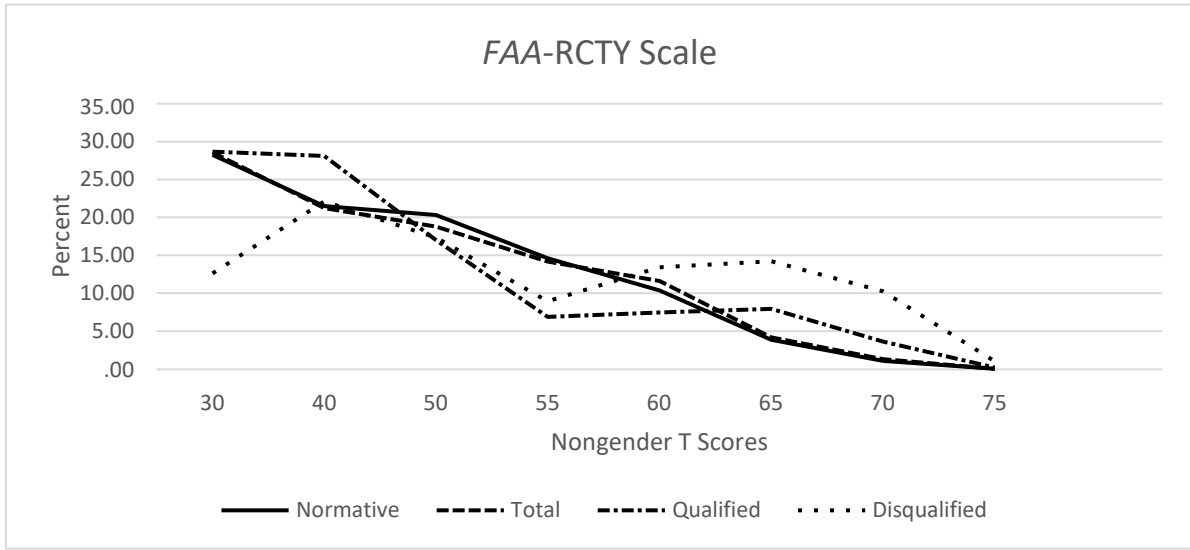*Note.* FAA-RCTY = Federal Aviation Administration-Reactivity; Ma = hypomania; MAC-R = MacAndrews alcoholism.

Table 16. Distribution of Nongender T-Scores and Z-scores on the FAA Reactivity Scale

T-Score Sample	FAA Reactivity Scale Nongender T-Score							
	30-40 %	**41-50** %	**51-55** %	**56-60** %	**61-65** %	**66-70** %	**71-75** %	**76+** %
FAA Normative	22.12	25.28	19.09	15.37	11.86	4.57	1.59	.11
Total	22.52	25.23	17.63	14.78	12.83	5.04	1.85	.13
Tier 2	25.50	19.58	21.24	8.54	9.02	9.05	6.37	.69
Cleared	27.26	26.71	17.92	7.54	7.40	7.58	5.05	.54
Not Cleared	21.05	11.05	17.89	8.16	13.68	12.89	13.16	2.11
Z-score	**-2.00~-1.00**	**-.99~.00**	**.01~.50**	**.51~1.00**	**1.01-1.50**	**1.51-2.00**	**2.01-2.50**	**2.5+**

FAA Reactivity Scale	Raw Score
M	17.57
SD	10.99

T Score	Z Score	Raw Score
30-40	**-2.00~-1.00**	6.58-17.56
41-50	**-.99~.00**	17.57-23.06
51-55	**.01~.50**	23.07-28.55
56-60	**.51~1.00**	28.56-34.05
61-65	**1.01-1.50**	34.06-39.55
66-70	**1.51-2.00**	39.56-45.04
71+	**2.01+**	45.05+

Note. FAA = Federal Aviation Administration.

Note. FAA-RCTY, Federal Aviation Administration-Reactivity.

Figure 6. Distribution of Nongender T-Scores on the *FAA-RCTY* Scale
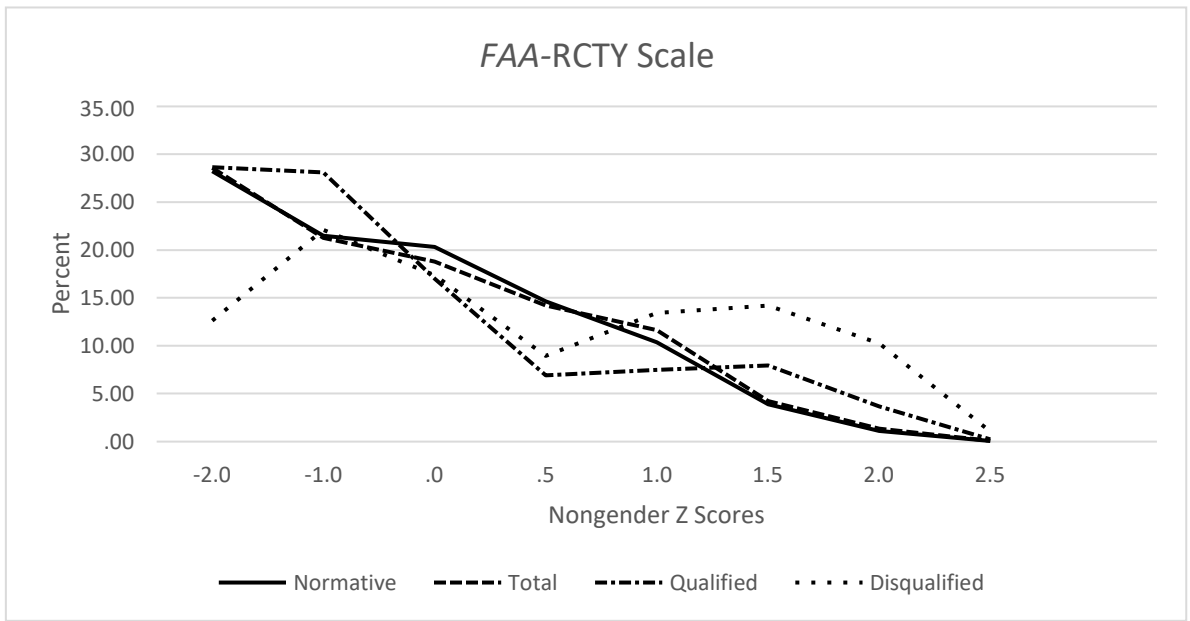
Note. FAA-RCTY, Federal Aviation Administration-Reactivity.

Figure 7. Distribution of Nongender Z-Scores on the *FAA-RCTY* Scale

SUMMARY

The FAA has been administering the MMPI-2 operationally for over 12 years in screening ATCS applicants. King et al. (2008) summarized the initial FAA use of the MMPI-2, which was a research effort that posed no job jeopardy. Retzlaff et al. (2011) reported on the normative scores during operational use of the MMPI-2 with the first 5,500 examinees. This report is a summary of the use of the MMPI-2 for the last 12 years. The MMPI-2 has been administered to a total of 20,385 applicants during this period with virtually no changes in the mean scores.

ATCS applicants whose MMPI-2 performance raised concerns about whether the applicant was aeromedically qualified, as defined in FAA ORDER 3930.3C (2019b), to conduct the duties of an ATCS were referred for a Tier 2 assessment. Nearly one-third of those applicants exceeded the cutting scores that are expected in a personnel screening situation on *L*, *K*, and *S* scales that assess self-favorable descriptions (Table 2). Another set of scales that assess psychopathology (*Ma*, *MAC-R*, and *DISC*) were elevated in 8% to 30% of those applicants. These two sets of scales, positive impression management, and possible psychopathology, were the primary focus of this report.

The *L* scale has been well substantiated as a measure of positive impression management that is used routinely in personnel selection. The scores on the *L* scale increase precipitously at a T-score of 75 (Z-score of 2.50) in the Qualified and Disqualified samples (Table 3). Interestingly, the Disqualified sample shows less of an increase than the Qualified samples. Slightly over half of the Qualified sample (53.25%) have T-scores of 76 or higher (Z-scores of 2.51 or higher) as compared to 35.27% of the Disqualified sample. Although the Qualified sample may appear to be working harder to make a positive impression than the Disqualified sample, a subset of the Disqualified group produced valid profiles (i.e., low or moderate elevations on *L*, *K*, and *S*), combined with elevations of one or more clinical and/or substance use scales at the time of screening. This result suggests that most individuals who are identified at Tier 1 due to an elevated *L* scale will ultimately be found aeromedically qualified. The challenge is identifying those who will be disqualified.

To increase the accuracy of classification between the Qualified and Disqualified samples on the *L* scale, the frequency of "False" responses to the *L* scale items in the FAA Nongender normative sample was used as the item weight to score each item on the *FAA-L* scale (Table 5). Higher weights for the items on the *FAA-L* scale reflect how unusual it is for these items to be endorsed "False," or how much effort the applicant is exerting to make a positive impression, and vice versa. A T-score of 60 (Z-score: 1.00) on the *FAA-L* scale seems to reflect where the Qualified and Disqualified samples start to differ from the FAA normative Nongender sample. The *L* and *FAA-L* scales also can be used jointly. Applicants, who have a T-score of 75 or higher on the *L* scale, yet have a T-score below 60 on the *FAA-L* scale are endorsing the less extreme items on the *L* scale and might *not* need to undergo a Tier 2 assessment. Those applicants who have a T-score of 75 or higher on the *L* scale and a T-score above 60 on the *FAA-L* scale would seem to warrant a Tier 2 assessment. Of course, these criteria can be adjusted up or down as needed.

The three MMPI-2 scales that were flagged in the initial review process (*Ma*, *MAC-R*, and *DISC*) yielded similar findings. There were too few items on each of the *Ma*, *MAC-R*, and *DISC* scales that had at least a 10% difference in item endorsement between the Disqualified and FAA Nongender normative samples to make individual FAA scales. Consequently, all nonredundant items from the *Ma*, *MAC-R*, and *DISC* scales that showed at least a 10% difference separating the Disqualified sample from the FAA Nongender normative sample and any other items from the MMPI-2 item pool that met the same criterion,

excluding *L* scale items, were combined into a preliminary *FAA-RCTY* scale of 70 items. Seventeen items, whose item content seemed troublesome when read, were dropped because they did not separate the Disqualified sample from the Qualified sample (Table 13). That is, these items are not relevant for identifying the Disqualified applicants, which is the sole purpose of this process. The *FAA-RCTY* scale shares six items with the *Ma* scale, 10 items with the *MAC-R* scale, and nine items with the *DISC* scale (Table 15). The *FAA-RCTY* scale will serve to identify those applicants whose pattern of item endorsement warrants a closer review. The Disqualified sample starts having higher scores than the Qualified sample around a T-score of 55, which would seem like a potential criterion that the applicant warrants a Tier 2 assessment (Table 16). Again, this criterion can be adjusted as desired.

REFERENCES

Americans with Disabilities Act of 1990, Pub. L. 101-336, 104 Stat. 328 (1991).

Ben-Porath, Y. S. & Tellegen, A. (2008). *MMPI-2-RF (MMPI-2-Restructured Form) Manual for Administration, Scoring, and Interpretation.* University of Minnesota Press.

Butcher J. N. (1994). Psychological assessment of airline pilot applicants with the MMPI-2. *Journal of Personality Assessment, 62*(1), 31–44. https://doi.org/10.1207/s15327752jpa6201_4

Butcher, J. N., Front, C. M., & Ones, D.S. (2018). Assessing psychopathology in high-risk occupations. In Butcher, J.N. (Ed.), *Handbook of Psychopathology: Vol. 1. Psychopathology: Understanding, Assessing, and Treating Adult Mental Disorders.* American Psychological Association.

Civil Rights Act of 1964, §7, 42 U.S.C. §2000e et seq (1964).

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom L. E. (1972). *An MMPI-2 Handbook: Vol I. Clinical Interpretation (Rev. ed.)*. University of Minnesota Press.

Federal Aviation Administration. (2019a, October). FAA Order 8500.4A – Administering the Minnesota Multiphasic Personality Inventory – 2 (MMPI-2) Test to Air Traffic Control Specialist (ATCS) Candidates. https://www.faa.gov/regulations_policies/orders_notices/index.cfm/go/document.information/doc umentID/1036759

Federal Aviation Administration. (2019b, June). FAA Order 3930.3C - Air Traffic Control Specialist Health Program. https://www.faa.gov/documentLibrary/media/Order/FAA_Order_3930.3C.pdf

Front, C. M. (2020). Selecting personnel for safety sensitive positions: Managing response distortion. In R. Bor, C, Eriksen, T. P. Hubbard, & R. King (Eds). *Pilot Selection: Psychological Principles and Practice*. CRC Press.

Harkness, A. R., & McNulty, J. L. (1994). *The Personality Psychopathology Five (PSY-5): Issues from the pages of a diagnostic manual instead of a dictionary.* In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality* (p. 291–315). Springer Publishing Company.

King, R. E., Schroeder, D. J., Manning, C. A., Retzlaff, P. D., & Williams, C. A. (2008, June). *Screening air traffic control specialists for psychopathology using the Minnesota Multiphasic Personality Inventory-2*. DOT/FAA/AM-03/20. Federal Aviation Administration. https://apps.dtic.mil/sti/citations/ADA482976

McKinley, J. C., & Hathaway, S. R. (1944). The Minnesota multiphasic personality inventory. V. Hysteria, hypomania and psychopathic deviate. *Journal of Applied Psychology, 28*(2), 153–174. https://doi.org/10.1037/h0059245.

Millon, T., Grossman, S., & Millon, C. (2015). *MCMI-IV: Millon Clinical Multiaxial Inventory-IV-manual*. National Computer Systems.

Morey, L. C., Warner, M. B. & Hopwood, C. J. (2007). The Personality Assessment Inventory: Issues in Legal and Forensic Settings. In Goldstein, A. M. (Ed). *Forensic psychology: Emerging topics and expanding roles*. (pp. 97-126). John Wiley & Sons Inc.

Nichols, D. S. & Greene, R. L. (1988). *Adaptive or defensive: An evaluation of Paulhus' two-factor model of social desirability responding in the MMPI with non-college samples*. 23rd Annual Symposium on Recent Developments in the Use of the MMPI, St. Petersburg, FL.

Nichols, D. S., & Greene, R. L. (1997). Dimensions of deception in personality assessment: the example of the MMPI-2. *Journal of Personality Assessment, 68*(2), 251–266. https://doi.org/10.1207/s15327752jpa6802_3

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598. https://doi.org/10.1037/0022-3514.46.3.598

Paulhus, D. L. (1986). Self-deception and impression management in test responses. In *Personality Assessment via Questionnaires* (pp. 143-165). Springer.

Paulhus, D. L. (1998). *Paulhus Deception Scales (PDS): the Balanced Inventory of Desirable Responding-7: User's Manual*. Multi-Health Systems.

Wiggins, J. S. (1964). Convergences among stylistic response measures from objective personality tests. *Educational and Psychological Measurement, 24*(3), 551-562. https://doi.org/10.1177/001316446402400310

Williams, C.A. & King, R. E. (2010, February). The effects of testing circumstance and education level on MMPI-2 correction scale scores. DOT/FAA/AM-10/3. FAA Office of Aerospace Medicine. https://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2010s/media/201003.pdf