7-2006

Reweighting AT-SAT to Mitigate Group Score Differences

Andrew R. Dattel
Civil Aerospace Medical Institute, dattela@erau.edu

Raymond E. King
Civil Aerospace Medical Institute

Federal Aviation
Administration

DOT/FAA/AM-06/16
Office of Aerospace Medicine
Washington, DC 20591

Reweighting AT-SAT to Mitigate Group Score Differences

Andrew R. Dattel
Raymond E. King
Civil Aerospace Medical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

July 2006

Final Report

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-06/16		

4. Title and Subtitle	5. Report Date
Reweighting AT-SAT to Mitigate Group Score Differences	July 2006
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Dattel AR, King RE	

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
FAA Civil Aerospace Medical Institute P.O. Box 25082 Oklahoma City, OK 73125	11. Contract or Grant No.

12. Sponsoring Agency name and Address	13. Type of Report and Period Covered
Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591	14. Sponsoring Agency Code

15. Supplemental Notes

Work was accomplished under approved subtask HRR-523.

16. Abstract

The Air Traffic Selection and Training (AT-SAT) test battery is the selection tool for applicants for Air Traffic Control Specialist (ATCS) positions within the Federal Aviation Administration (FAA) who have not previously been employed as an air traffic controller. AT-SAT is an aptitude test developed to predict the likelihood of successfully learning ATCS skills. Before operational use, however, concerns were raised about the low passing rate of incumbent (who are fully trained and certified) ATCS personnel (who participated in the initial research) and score differences between groups, which could result in adverse impact (possible unfair discrimination). To address these concerns, the subscores of AT-SAT were reweighted, and the additive constant was changed to yield a new total score. The present study compares the original and new scoring methods using data from 724 developmental ATCSs who volunteered to take AT-SAT. An average increase of 4.86 points was found with the new scoring method; the notional passing rate (achieving a score ≥ 70) changed from 58.8% to 80%. American Indian/Alaskan Native, Hispanic, and black participants showed the greatest average increase in overall scores, 6.97, 6.98, and 7.02, respectively. The increase in scores of Hispanic and black participants was significantly higher than the increase in scores for white participants [$F(4, 689) = 6.186$, $p < .001$]. However, a chi square analysis showed no differences between groups for the number of participants whose failing score with the original scoring method changed to a passing score with the new scoring method. Additionally, a Spearman rank correlation coefficient of .85 was found between the two scoring methods, indicating that the ranking of individual participants did not change significantly. Moreover, no differences were found between groups in rank ordering of the two scoring methods. No significant gender differences were found between the scoring methods, with the scores for males increasing an average of 4.58 points and scores for females increasing an average of 5.67 points under the new weighting method. This study found that the new weighting formula has benefited all groups and is likely to reduce the potential of adverse impact.

17. Key Words	18. Distribution Statement
Aptitude Testing, Computerized Testing, Personnel Selection	Document is available to the public through the Defense Technical Information Center, Ft. Belvior, VA 22060; and the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	12	

Reweighting AT-SAT to Mitigate Group Score Differences

The Air Traffic Selection and Training (AT-SAT) battery, a six-and-a-half-hour computerized battery of tests (Heil & Reese, 2002; King & Dattel, 2005; Ramos, Heil, & Manning, 2001), was developed to identify applicants with the necessary aptitude to learn to become air traffic control specialists (ATCSs). AT-SAT is currently composed of eight subtests: *Dials, DI; Applied Math, AM; Scan, SC; Angles, AN; Letter Factory, LF; Air Traffic Scenarios, ATST; Analogies, AY;* and the *Experiences Questionnaire, EQ* (See Table 1 for a short description of the subtests). These eight subtests yield 22 individual "part scores" that, when weighted and combined (with a constant), yield an overall score.

Before operational use of AT-SAT for hiring purposes, concerns were raised about differences in AT-SAT scores among protected groups.[1] Consequently, FAA management met with representatives from these protected groups to solicit their input. The original passing score of 70 had been calibrated so that 62% of fully certified incumbent controllers would achieve an AT-SAT score equal to, or greater than, 70. The intent was to minimize FAA Academy and on-the-job training failures and to compensate for the need for ATCSs to perform potentially more difficult duties in the future (Waugh, 2001).

After these concerns were raised and the representatives' comments were heard, FAA management directed AT-SAT researchers to explore the possibility of reducing potential adverse impact without unduly comprising the validity of the test. Additionally, management decided that most fully qualified incumbent FAA controllers should be able to pass FAA's entry-level aptitude test. AT-SAT researchers were also asked to determine if reasonable changes could be made to AT-SAT to mitigate differences between groups without sacrificing validity as a predictor of ATCS job performance.

Wise, Tsacoumis, Waugh, Putka, and Hom (2001) reported on the consequent reweighting of AT-SAT subtests to reduce variability between groups. (The specific weighting of subtests are not noted here due to concerns over potential coaching efforts that would attempt to target the most heavily weighted subtests to inflate scores for the benefit of applicants.) The content of the subtests was not changed; rather, the subtests were weighted differently. The challenge was to retain adequate validity while reducing differences in scores between groups that could result in adverse impact.[2]

Table 1. Description of the eight AT-SAT subtests.

Subtest	Description
Dials (DI)	Scanning and interpreting readings from a cluster of analog instruments
Applied Math (AM)	Solve basic math problems as applied to distance, rate, and time
Scan (SC)	Scan dynamic digital displays to detect targets that regularly change
Angles (AN)	Determine the angle of intersecting lines
Letter Factory (LF)	Participate in an interactive dynamic exercise that requires categorization skills, decision making, prioritization, working memory (incidental learning), and situation awareness
Air Traffic Scenarios (ATST)	Control traffic in interactive, dynamic low-fidelity simulations of air traffic situations requiring prioritization
Analogies (AY)	Solve verbal and nonverbal analogies that require working memory and the ability to conceptualize relationships
Experience Questionnaire (EQ)	Respond to Likert scale questionnaire about life experiences

[1] Including, but not limited to, female and black group members.

[2] Adverse impact is determined by the "Four-Fifths Rule" as stated in the Uniform Guides (Sec. 1607.4 D): "A selection rate for any race, sex, or ethnic group which is less than four-fifths (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact…"

One method of measuring test validity (job-relatedness) is to correlate test scores with job performance. After reweighting,[3] the AT-SAT validity co-efficient went from .69 to .60, and is, therefore, still considered to have a strong relationship to job performance. The relationship to job performance is especially important in this context as any remaining differences in scores between groups can be justified by "business necessity."[4]

The purpose of this paper is to examine, with empirical data, the impact of the reweighting effort and its effectiveness in reducing differences in scores between groups. Wise et al. (2001) computed the reweighting formula using data from the original concurrent validation study. Those participants were incumbent controllers. The current study uses participants more similar to future applicants as they, too, are applicants, albeit successful ones (they represent those who were hired). Notional pass rates (the voluntary participants in the present study were not required to achieve a passing score) will be considered in terms of race/ethnicity and gender. The potential change in overall pass rates will also be empirically examined.

METHOD

Participants

Data were collected from 724 students ("developmentals") who were enrolled in the Air Traffic Training program at the FAA Academy. These developmentals had been selected into the air traffic training program by methods other than passing AT-SAT, such as by passing the Office of Personnel Management written test (mostly College Training Initiative, CTI, applicants) or based on previous employment as an air traffic controller (such as in a branch of the military), and they voluntarily agreed to take AT-SAT for research purposes upon entering training.

Students who volunteered to take the AT-SAT were enrolled in either initial en route or terminal training. Of the 724 participants, 292 took Version 1.0 of the AT-SAT (158 were enrolled in en route, 132 in terminal). The remaining 432 participants took Version 2.0, the reweighted version (165 were enrolled in en route, 269 were enrolled in terminal). The content of these two versions were identical; only the weighting of the subtests varied, and these differences were transparent to the participants.

[3] Throughout this paper, "reweighting" refers to the change in weights of subtests as well as the changed constant.

[4] Business necessity ensures that the selection procedure is closely coupled to the requirements of the job, usually as demonstrated by job analysis.

Procedure

Participants were recruited during the first few days of their two- to three-month (depending on option - terminal or en route - respectively) initial training curriculum at the Academy. They were offered the opportunity to volunteer as research participants in a continuing effort to validate AT-SAT as a selection measure. Each student was assured his or her score on the AT-SAT was not part of the training evaluation and that none of the instructors would have access to the results. It takes between 6.5 to 8 hours to complete the AT-SAT; the entire test is presented via computer and responses are recorded via numeric keypad and mouse. As previously described, the content of the subtests themselves were not changed from the original-weighting version (which is termed version 1.0) to the reweighted version (version 2.0), and participants were totally unaware of the change in weighting.

Recalculation of Scores

To calculate the new (reweighted) score from the AT-SAT version 1.0 results, scores from the AT-SAT subtests were converted to raw scores and recalculated with the new weighting formula. The basis for recalculating the scores was drawn from the example found in Wise et al. (2001). Conversely, this formula also specified a method for taking subtest scores from the reweighted version of AT-SAT (version 2.0), weighting them with the original method, and applying the formerly used constant to arrive at the overall score that would have been achieved under the original weighting scheme. The subtest and overall scores of the 292 developmentals who took AT-SAT under the original weighting scheme (version 1.0) were converted, as described above. Likewise, scores from 432 developmentals who took the reweighted version of AT-SAT (version 2.0) were converted to the scores that would have been achieved under the original weighting scheme, as described above. The two groups differed only according to the weighting scheme in place when they took AT-SAT, which, as previously noted, was totally transparent to each of the participants. The presentation of the subtests was identical. Thus, each of the total 724 cases could be scored under both weighting schemes for the purposes of this paper.

RESULTS

Gender and race/ethnicity information, as self-reported by participants on OPM Form 1468, were collected from the 724 participants. Nine participants indicated they were American Indian/Alaskan Native, 21 indicated they were Asian/Pacific Islander, 54 indicated they were black (not of Hispanic origin), 71 indicated they were

Hispanic, and 539 indicated they were white (not of Hispanic origin). Thirty participants chose not to answer the race/ethnicity question. Five hundred and fifty nine were male, 145 were female, and 20 participants elected to not specify their gender.

The average total increase in overall score between the original version (version 1.0) and the reweighted version (version 2.0) was 4.86 *(SD*=6.65). Although most overall scores increased, slightly over 20% of the overall scores decreased. Of the overall scores that showed a decrease, the average decrease was 4.18 (*SD*=3.18). Of the overall scores that increased, the average increase was 7.59 (*SD*=4.75).

Table 2. Change in notional pass/fail status between original scoring method and reweighted scoring method.

		Reweighted Scores		
		Pass	Fail	Total
Original Scores	Pass	423	3	426
	Fail	153	145	298
	Total	576	148	724

Using the original weights, 426 of the 724 research participants (58.8%) would have achieved a passing score (70 or above). The reweighted scores changed 153 individuals' failing scores to passing scores but also changed three individuals' passing scores to failing scores. The reweighting formula resulted in a net gain of 150 individuals, for a total of 576 (80%) individuals who would have achieved a passing score. A chi-square analysis showed this change to be significant $X^2(1)=244.28$, $p<.001$. Table 2 shows the number of participants whose scores changed from pass to fail, fail to pass, and no change in pass or fail when rescoring the original scores to the reweigted scores.

Table 3 depicts the pass rate by race/ethnic group and gender with AT-SAT scored under both weighting schemes. Such a display demonstrates the potential for score differences that could result in adverse impact, under both weighting schemes (recall that adverse impact is determined by the "Four-Fifths Rule"). In this example, a passing rate of less than 80% (for protected race/ethnic groups and women) would suggest a group score difference that could result in adverse impact (because one group has a passing rate of 100%).

The next area of concern is the impact on *individuals* as well as groups under both weighting schemes. Consequently, analyses of rank order for the two scoring methods were conducted. A Spearman rank correlation coefficient found a strong correlation between the two scoring methods r_s (724) = .85, $p<.001$, with a R^2 of .72.

Table 3. AT-SAT notional passing rate (≥70) by race/ethnic group and gender for both weighting methods.

Group	**Original scoring method**	**Revised scoring method**	**Net increase**
	Number (%) of passing scores	*Number (%) of passing scores*	*Number (%) of passing scores*
American Indian or Alaskan Native	7 (77.8%)	9 (100%)	2 (22.2%)
Asian or Pacific Islander	13 (61.9%)	15 (71.4%)	2 (9.5%)
Black, not of Hispanic Origin	20 (37%)	39 (72.2%)	19 (35.2%)
Hispanic	28 (51.9%)	47 (87%)	19 (35.1%)
White, not of Hispanic Origin	343 (63.6%)	443 (82.2%)	100 (18.6%)
Unknown race/ethnicity group	15 (50%)	23 (76.7%)	23 (76.7%)
Male	343 (61.4%)	458 (81.9%)	115 (20.5%)
Female	73 (50.3%)	102 (70.3%)	29 (20%)
Unknown gender	10 (50%)	16 (80%)	6 (30%)

Additionally, the change in rank between the two scoring methods by race/ethnicity and gender was calculated (Table 4). A chi square analysis was also conducted for the change in rank position from the original scoring formula and the reweighting scoring formula. Reweighting the scores showed no differences in increase/decrease of rank by race/ethnicity group $X^2(4)=2.767$, $p=.598$, or gender $X^2(1)=.805$, $p=.370$.

The next set of analyses contrasts scores attained using the original scoring method with those attained using the reweighted scoring method. ANOVAs comparing different race/ethnic groups and across genders were computed for each scoring method.

Original scoring method

The mean scores (with standard deviation in parentheses) on the AT-SAT by gender and race/ethnic group when scored by the original weighting scheme are shown in Table 5.

Because of the large variation in the number (n) of participants by race/ethnic group and gender, one-way ANOVAs and t-tests were conducted separately for race/ethnic group and gender. An ANOVA, using AT-SAT scores as the dependent variable and race/ethnic group as the independent variable, revealed a main effect for Race/Ethnic Group, $F(4,689) = 8.612$, $MS_E = 170.405$, $p<.001$. Tukey post hoc analyses showed significantly

Table 4. Change in rank between two scoring methods by race/ethnicity group and gend

Group	Total members	Participants ↑ overall rank		Participants ↓ overall rank		No rank change	
		n	%	n	%	n	%
American Indian or Alaskan Native	9	4	44.44	5	55.56	0	
Asian or Pacific Islander	21	13	61.91	7	33.33	1	.05
Black, not of Hispanic Origin	54	26	48.15	28	51.85	0	
Hispanic	71	33	46.48	38	53.52	0	
White, not of Hispanic Origin	539	281	52.13	255	47.31	3	.01
Male	559	283	50.63	272	48.66	4	.01
Female	145	80	55.17	65	44.83	0	

Table 5. Mean scores of AT-SAT by Gender and Race/Ethnic Group when scored by original weighting application (standard deviations in parentheses).

Race/ethnic group	Male	Female	Combined
American Indian or Alaskan Native	76.81 (10.40) n=7	67.84 (4.37) n=2	74.81 (9.96) n=9
Asian or Pacific Islander	72.59 (15.01) n=17	79.34 (13.80) n=4	73.87 (14.70) n=21
Black, not of Hispanic Origin	67.96 (11.84) n=43	63.02 (13.52) n=11	66.96 (12.23) n=54
Hispanic	68.56 (13.37) n=52	63.01 (9.57) n=19	67.08 (12.64) n=71
White, not of Hispanic Origin	75.52 (12.86) n=431	70.96 (13.75) n=108	74.61 (13.16) n=539
All groups	74.07 (13.21) n=559	69.55 (13.50) n=145	73.01 (13.31) n=724*

* Includes participants that did not indicate their gender

higher scores for white participants when compared with both black and Hispanic participants. A t-test, using AT-SAT results as the dependent variable and gender as the independent variable, showed higher scores for males than for females, t (702) = 3.652, p<.001. See Figure 1 for a graphical representation of the mean AT-SAT scores and standard deviations by Race/ethnic group and Figure 2 for a graphical representation of the mean AT-SAT scores and standard deviation by gender, scored by the original scoring method (version 1.0).

Reweighted Scoring Method

The mean overall AT-SAT scores and standard deviations for the reweighted scoring method are shown in Table 6.

An ANOVA using the reweighted AT-SAT scores as a dependent variable found a significant main effect for Race/Ethnic Group F(4,689) = 6.186, MS_E = 105.746, p<.001. Tukey post hoc analyses also showed significantly higher reweighted AT-SAT scores for white participants when compared with both black and Hispanic participants

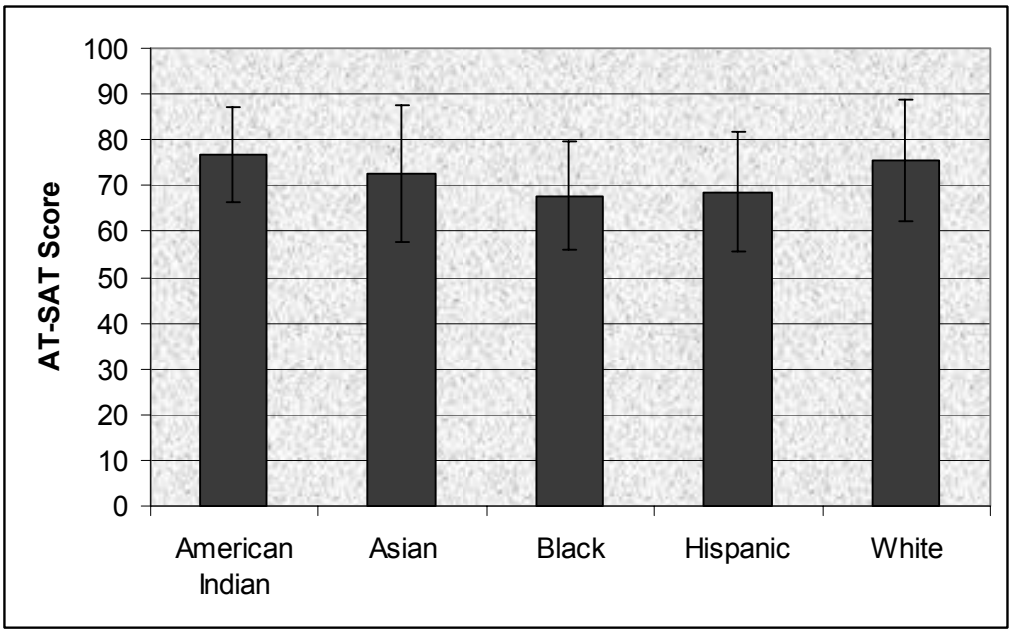

Figure 1. Version 1.0 (original weighting) AT-SAT scores and standard deviations by race/ethnic group.

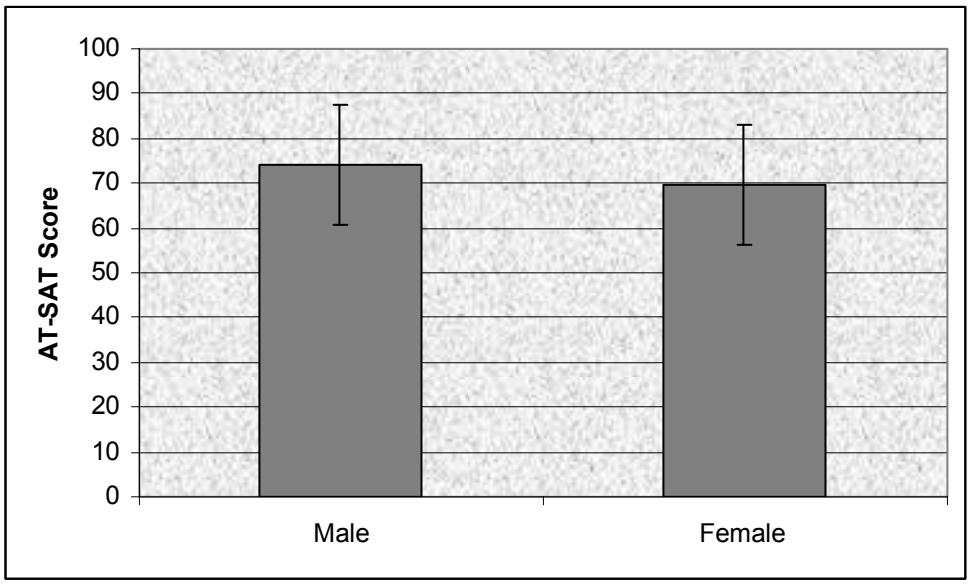

Figure 2. Version 1.0 (original weighting) AT-SAT scores and standard deviations by gender.

Table 6. Mean scores of AT-SAT by Gender and Race/Ethnic Group when scored by revised weighting (standard deviations in parentheses).

Race/group	**Gender**		
	Male	*Female*	**Combined**
American Indian/ Alaskan Native	82.56 (5.13) n=7	79.06 (6.30) n=2	81.78(5.20) n=9
Asian or Pacific Islander	76.27 (11.49) n=17	78.82 (7.47) n=4	76.76 (10.72) n=21
Black, not of Hispanic Origin	74.98 (9.27) n=431	70.07 (10.61) n=11	73.98 (9.66) n=54
Hispanic	74.96 (10.65) n=52	71.58 (7.66) n=19	74.05 (10.00) n=71
White, not of Hispanic Origin	79.66 (10.19) n=431	76.08 (10.85) n=108	78.94 (10.42) n=539
All groups	78.65 (10.43) n=559	75.23 (10.49) n=145	77.86 (10.51) n=724*

* Includes those participants that did not indicate their gender

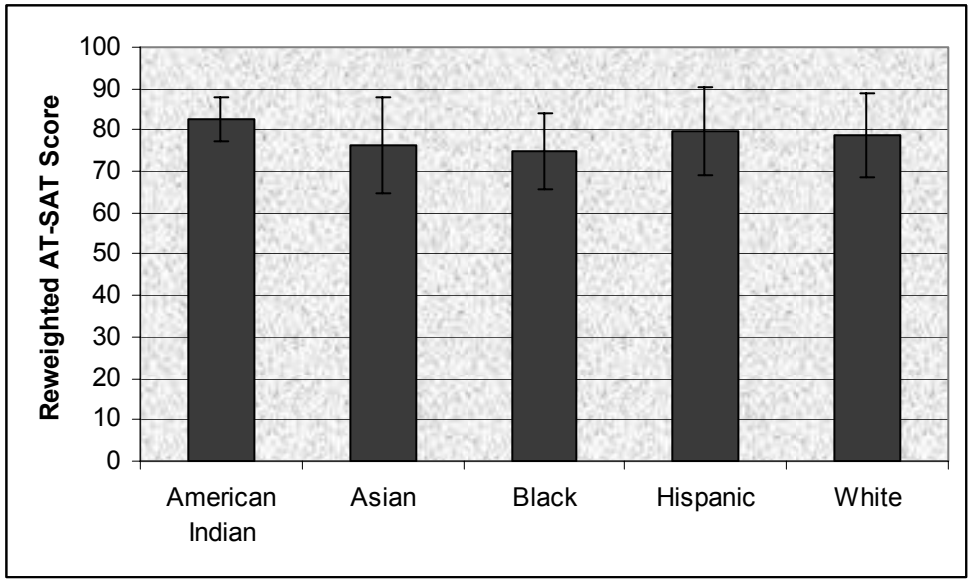

Figure 3. Reweighted AT-SAT scores by Race/ethnic group.

when the reweighted scoring method was applied. A t-test, using reweighted AT-SAT scores as the dependent variable and gender as the independent variable, showed higher scores for males than for females, t (702) = 3.513, p<.001. See Figure 3 for a graphical representation of the mean AT-SAT scores and standard deviations by Race/ethnic group and Figure 4 for a graphical representation of the mean AT-SAT scores and standard deviation by gender, scored by the reweighted scoring method (Version 2.0).

Differences Between Original and Reweighted Scores

Analyses were conducted on the difference in scores calculated by the original, versus the reweighting, scheme (Table 7). A oneway ANOVA of change in AT-SAT scores by race/ethnic group found a main effect $F(4,689) = 4.718$, MS_E=22.28, p=.001. Tukey post hoc analyses showed the increase in scores for blacks and Hispanics were significantly greater than the increase in scores for whites. A t-test of change in AT-SAT scores by gender found only a marginally larger increase in scores for females when compared to males t (702) = 1.756, p = .080.

Subtests

At the more elemental level, analyses were conducted on the subtests to examine the differences in their scores as a function of race/ethnic group and gender. An ANOVA with AT-SAT subtests scored using the original weighting method as dependent variables showed a significant main effect for race/ethnic group for the following subtests: *Dials, Applied Math, Angles, Letter Factory, Air Traffic Scenarios* (*ATST*), and the *Experience Questionnaire* (See Table 8). Tukey post hoc analyses showed whites and Asians scored higher than blacks for the *Dials* subtest, whites scored higher than both blacks and Hispanics on the *Applied Math* and *Angles* subtests, whites scored higher than both blacks and Asians on the *ATST*, and whites, Hispanics, and American Indians scored higher than Asians on the *Experience Questionnaire*. The less conservative LSD post hoc analyses found whites and Asians scored higher than Hispanics on the *Letter Factory* subtest. When the reweighting method was applied, the *Letter Factory* and the *Experience Questionnaire* subtest

Table 7. Improvement in mean AT-SAT scores due to reweighting of scores for gender and race/ethnic group (standard deviations in parentheses).

Race Ethnic/Group	**Mean change in score**
American Indian/Alaskan Native	6.97 (5.35) n=9
Asian/Pacific Islander	2.88 (5.43) n=21
Black, not of Hispanic Origin	7.02 (2.64) n=54
Hispanic	6.98 (7.16) n=71
White, not of Hispanic Origin	4.33 (6.71) n=539
Gender	
Male	4.58 (6.75) n=559
Female	5.67 (6.58) n=145

Table 8. ANOVA and follow up tests of subtests by race/ethnic group when scored by original weighing method.

Post Hoc(Tukey)	***Dials***	***Applied Math***	***Scan***	***Angles***	***Letter Factory***	***ATST***	***Analogies***	***Experience Quest***
Omnibus F(4,689)	6.84**	9.54**	1.43^{ns}	8.74**	2.43*	6.85**	.91ns	4.12**
White>Blacks	**	**		**		**		
White>Hispanics		**		**	*^{LSD}			
Asians>Blacks	*							
Asians>Hispanics					*^{LSD}			
White>Asians						**		**
Hispanics>Asians								*
American Indian > Asians								*

**<.01

*<.05

scores were no longer significantly different for any race/ethnic group, but a significant change was found for *Analogies* in that Tukey post hoc analyses showed whites scored higher than both blacks and Hispanics. A Tukey post hoc analysis no longer showed that whites scored higher than blacks on the *ATST*, but the LSD analysis did show whites scored higher than blacks on the *ATST*. Additionally, the Tukey post hoc analyses no longer showed whites scoring higher than Asians on the *ATST* subtest, but LSD post hoc analyses showed whites scored higher than Hispanics on the reweighted scoring of the *ATST* subtest (See Table 9).

Men scored significantly higher than women on the *Dials, Applied Math, Angles*, and *Air Traffic Scenarios* subtests when they were scored both with the original weighting scheme and the reweighted scheme. T-test analyses showed women scored higher than men when using the original weighting scheme for the *Experience Questionnaire*, but no differences between men and women were found when the *Experience Questionnaire* was scored by the reweighting scheme (See Table 10).

Table 9. ANOVA and follow up tests of subtests by race/ethnic group when scored by the reweighted method.

Post Hoc(Tukey)	***Dials***	***Applied Math***	***Scan***	***Angles***	***Letter Factory***	***ATST***	***Analogies***	***Experience Quest***
Omnibus F(4,689)	6.84**	9.54**	1.43^{ns}	8.75**	2.38^{ns}	3.15*	5.10^{**}	$.68^{ns}$
White>Blacks	**	**		**		$*^{LSD}$	*	
White>Hispanics		**		**		$**^{LSD}$	**	
Asians>Blacks	*							

**<.01

*<.05

Table 10. ANOVA of subtests by gender when scored by both earlier and revised weighting method.

	Original Scoring Method			**Revised Scoring Method**	
	Mean (SD)			Mean (SD)	
Subtests	Male	Female	**Subtests**	Male	Female
Dials $t(702)=2.047^{*}$	10.54(1.32)	10.29(1.33)	*Dials* $t(702)=2.047$*	1.80(.23)	1.76(.23)
Applied Math $t(702)=5.923$**	15.67(4.39)	13.20(4.79)	*Applied Math* $t(702)=5.923$**	20.03(5.61)	16.88(6.12)
Scan $t(702)=1.336^{ns}$	11.92(3.30)	11.51(3.26)	*Scan* $t(702)=1.336^{ns}$	8.0(2.22)	7.72(2.18)
Angles $t(702)=4.003$**	13.11(2.01	12.34(2.36)	*Angles* $t(702)=4.003$**	1.55(.24)	1.46(.28)
Letter Factory $t(702)=1.063^{ns}$	12.64(6.47)	12.06(6.49)	*Letter Factory* $t(702)=.966^{ns}$	4.31(2.20)	4.11(2.23)
ATST $t(702)=2.432$*	4.98(1.52)	4.63(1.65)	*ATST* $t(702)=4.962$**	1.99(.59)	1.71(.59)
Analogies $t(702)=1.078^{ns}$	6.86(2.17)	6.65(2.08)	*Analogies* $t(702)=.127^{ns}$	5.23(1.24)	5.25(1.35)
Experience Questionnaire $t(702)=2.437$*	8.95(2.63)	9.53(2.40)	*Experience Questionnaire* $t(702)=.577^{ns}$	25.25(7.39)	24.85(7.83)

**<.01

*<.05

CONCLUSIONS AND DISCUSSION

Reweighting has indeed reduced group differences and hence the potential for adverse impact. Improvements in scores were found for each race/ethnic group and both genders. Using reweighted subtest scores reduced some of the group differences across individual subtest scores. The reweighting effort did not substantially inflate subtest scores and consequently, the overall scores for any particular group.

Reweighting was based on data collected from incumbent ATCSs who took AT-SAT on a research basis; some of these employees achieved overall scores less than 70 (that was one of the reasons for the reweighting effort – a belief that incumbent employees should be able to pass the entry-level selection test). When AT-SAT is used for hiring purposes, overall pass rates are likely to increase; this issue requires continual monitoring and assessment via longitudinal validation.

The present study used empirical data from participants hired (on the basis of successfully negotiating one of several selection systems other than passing AT-SAT) to train in the ATCS career field. Thus, there was not only a restriction in range, as participants consisted only of those individuals who had been selected, but also the present sample contains only individuals who had successfully negotiated a selection system. Another important limitation in the study was the low stakes these individuals had in the results of their AT-SAT efforts, as they were explicitly told that their results would have no impact on their careers. While the reweighting scheme seems to be working on the subtest level to reduce some group differences and, thus, potential adverse impact, score differences between groups will be continually monitored. Such monitoring will continually assess the potential for group differences that could result in adverse impact as AT-SAT results are acquired from actual applicants (including those who pass and those who fail), assessed with AT-SAT for selection purposes.

DISCLAIMER

This is a statistical snapshot of the workforce demographics. The use of this data in any employment decision is PROHIBITED without the express written authorization of the Deputy Chief Counsel for Operations, AGC-3.

REFERENCES

Heil, M.C. & Reese, N.S. (2002). FAA air traffic controller selection in the United States. In Hinnerk Eißfeldt, Mike C. Heil, and Dana Broach (Eds). Staffing the ATM system: The selection of air traffic controllers, (pp. 109-20). Burlington, VT: Ashgate.

King, R.E. & Dattel, A.R. (2005). The air traffic selection and training battery: What it is and isn't (and how it has changed and hasn't). Proceedings of the 13th International Symposium on Aviation Psychology, Oklahoma City, OK.

Ramos, R.A., Heil, M.C., & Manning, C.A. (2001). Documentation of the validity for the AT-SAT computerized test battery, Volume I (DOT/FAA/AM-01/05). Washington, DC: FAA Office of Aviation Medicine.

Waugh, G. (2001). Analyses of group differences and fairness. In R.A. Ramos, M.C. Heil, & C.A. Manning (Eds.) Documentation of validity for the AT-SAT computerized test battery,Volume II (DOT/FAA/AM-01/6) (pp.43-7). Washington, DC: FAA Office of Aviation Medicine.

Wise, L.L., Tsacoumis, S., Waugh, G.W., Putka, D.J., & Hom, I. (2001). Revision of the AT-SAT. (Unpublished Manuscript).