3-2020

Undergraduate Pilot Training Attrition: An Analysis of Individual and Class Composition Component Factors

Christina M. Akers

UDERGRADUATE PILOT TRAINING ATTRITION: AN ANALYSIS OF INDIVIDUAL AND CLASS COMPOSITION COMPONENT FACTORS

THESIS

Christina M. Akers, Captain, USAF

AFIT-ENS-MS-20-M-126

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-20-M-126

UDERGRADUATE PILOT TRAINING ATTRITION: AN ANALYSIS OF INDIVIDUAL AND CLASS COMPOSITION COMPONENT FACTORS

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Logistics and Supply Chain Management

Christina M. Akers, BS

Captain, USAF

March 2020

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENS-MS-20-M-126

UDERGRADUATE PILOT TRAINING ATTRITION: AN ANALYSIS OF INDIVIDUAL AND CLASS COMPOSITION COMPONENT FACTORS

Christina M. Akers, BS

Captain, USAF

Committee Membership:

Dr. Adam D. Reiman
Chair

Lt Col Jason R. Anderson
Member

AFIT-ENS-MS-20-M-126

Abstract

The Air Force is working hard to reduce the shortage of nearly 2,000 pilots that threatens the Air Force's core mission. Officials have focused on increasing retention and training throughput. Despite this, the first Pilot Training Next class graduated in August of 2018 with 13 of the initial 20 students (65.5% graduation rate). The purpose of this research is to explore attrition reduction by understanding how class composition of individual abilities and personalities affects the class graduation rate. Using AFOQT scores, SDI+ scores, PCSM scores, flight hours, and college GPAs, correlations were studied and a simple linear regression was run with the variables to determine relationships. This study resulted in the creation of models to help decision makers plan classes to optimize success rates. Additionally, correlations between group scores and graduation rates were compared to correlations between individual scores and individual performance. Decision makers can employ these findings in the creation of future classes to increase performance and decrease attritions.

Acknowledgments

I would like to express my sincere appreciation to my faculty advisor, Dr. Adam D. Reiman (Retired Colonel), for his guidance and support throughout the course of this thesis exploration. I'd also like to thank my classmates, Capt Adam Brubakken, Capt Daniel Cherobini, Capt Kyle Gladney, Capt Mike Ingram, and Capt Tom O'Neal for their insight and support during the entire program. I would, also, like to thank my spouse and children for supporting me during my countless hours of research.

Christina M. Akers

Table of Contents

Page

List of Figures

Page

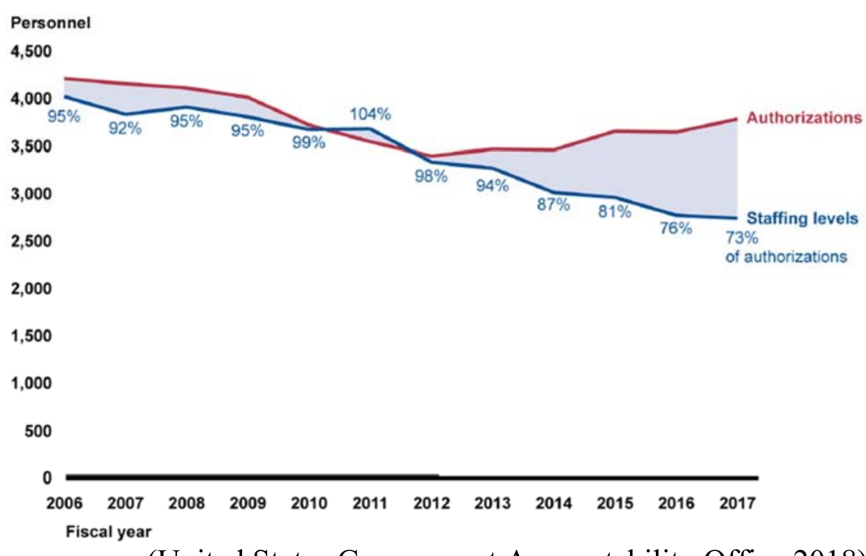
List of Tables

UDERGRADUATE PILOT TRAINING ATTRITION: AN ANALYSIS OF INDIVIDUAL AND CLASS COMPOSITION COMPONENT FACTORS

I. Introduction

General Issue

The success of the United States Air Force (USAF) mission relies on the availability of resources, both aircraft and people. The Air Force needs approximately 20,000 pilots to support its 5,500 aircraft (Axe 2018). Unfortunately, the Air Force fell short of this requirement. In September 2016, the Air Force announced there was a shortage of 1,555 pilots and by December, 2017, the shortage had grown to nearly 2,000 (Secretary of the Air Force Public Affairs, 2017; Panzino, 2018). The decrease in pilots can be attributed to experienced pilots leaving the Air Force at a faster rate than they can train new pilots. The Air Force chose to focus their efforts on both ends of the problem, retention and training throughput (Losey 2018).



(United States Government Accountability Office 2018)

Figure 1: Air Force Active Component: Fighter Pilot Actual Staffing Levels Compared with Authorizations, Fiscal Years 2006-2017

Over recent years the Air Force began the process of implementing initiatives to retain experienced pilots. These initiatives included an increase in aviation incentive pay, expanding the Aviation Bonus Program, and implementing the Voluntary Rated Return to Active Duty program which will enable retired pilots to return to service (Secretary of the Air Force Public Affairs 2017). The Air Force will also be testing an Aviator Technical Track in which certain mobility pilots can remove themselves from non-flying duties to get more time in the air and even remain at the same base for up to five years as desired (Panzino 2018).

To increase pilot production, programs are in the pipeline to address training capacity limitations. The Air Force updated the Undergraduate Pilot Training (UPT) syllabi and reduced training time from 54 to 49 weeks (Pawlyk 2018). They also introduced Pilot Training Next (PTN), with virtual reality training devices that can be used outside of the classroom. This changed the typical read and visualize study session into realistic, read, visualize and practice session. Trainees that excel are moved up faster while trainees who need extra time move up at a slower pace, allowing each member to get an individualized level of instruction (Pawlyk 2018). These changes turned the normal 12 month timeframe into a flexible six to eight months (Pawlyk 2019). The goal is to increase training capacity to 1,500 trainees each year, by 2022 (Losey 2018).

Pilot candidates are administered cognitive and personality tests and competitively vetted through selection boards by senior leaders. The process is designed to weed out questionable candidates. Despite this, trainees are still removed from training, primarily because of the following reasons: personal (Drop on Request, DOR), medical, academic performance, or flying performance (Schulker et al., 2018, p.34).

Schulker's study included pilot training data from 2009-2014 in which students were eliminated at the following rates: 9% during Phase 1: Academic/Ground Training, 8% in Phase 2: Primary Flying Training (T-6, PFT), and 2% during Phase 3: Advanced Flying Training (T-1, T-38, TH-1H. or RPA) (2018). Of those rates, flying performance, DOR, and academic eliminations accounted for 8%, 7%, and 1%, respective to each level of training (2018). Applying these historical rates to the increased capacity of 1500 would produce 269 attritions. In fact, the first updated PTN class graduated in August of 2018 with 13 of the initial 20 students (65.5% graduation rate) (Pawlyk 2019). This suggested that increasing capacity should be coupled with a focus on decreasing attritions to increase training throughput and alleviate the manning gap.

Problem Statement

The Air Force has a pilot manning shortage and attrition in UPT is a contributing factor.

Purpose Statement

The purpose of this study is to identify and explore linear relationships between class mean cognitive and personality scores and the class graduation percentage.

Research Focus

The research focus will be on analyzing pilot training records to identify components that have the greatest impact on the probability of graduation. These components could be from to Air Force Officer Qualifying Test (AFOQT), Self-Description Inventory + (SDI+), or Pilot Candidate Selection Method (PCSM) scores, prior pilot experience (Flight Hours), or college performance (GPA). The goal is to determine if any of these components could be exploited to group classes in such a way to reduce attrition.

Limitations

This data contains only USAF trainees. Data for students from other services and countries are not available and therefore created holes in the classes. Classes with significant missing student data could skew the results so only classes with at least 75% of the students were included in the analysis. This percentage was chosen in order to optimize the number of classes while still providing an adequate student sample.

Students who do not complete training do not have training performance data. Therefore, a direct comparison of scores to performance is not capable for attritions. Additionally, while the type of attrition is provided, there is no way to determine the context of the attrition so each attrition is treated the same, even though they are not.

The data primarily contained students who had taken the Form S version of the AFOQT. There were only three classes with 75% of the data on the updated Form T facets. Therefore, these facets were not analyzed in this research.

II. Literature Review

Pre-Accession Testing

Pilot selection and training is a highly competitive process involving multiple tests and phases. Candidates must complete the AFOQT, a measure of cognitive ability, and the included SDI+ subtest, a measure of personality. Eligible members are boarded by members of their commissioning source and only the top candidates are accepted into the program.

AFOQT.

The AFOQT is a multiple-choice test consisting of cognitive subtests and the SDI+ section (Barron, Carretta, and Rose 2016). It has been modified over the years in an attempt to find the best means of testing for military career field aptitude. From 2005 to 2014, the AFOQT Form S was administered. It had eleven subtests included Verbal Analogies, Arithmetic Reasoning, Word Knowledge, Math Knowledge, General Science, Table Reading, Hidden Blocks, Rotated Blocks, Instrument Comprehension, Block Counting, and Aviation Information. These subtest scores were distributed into five composite scores: Verbal (V), Quantitative (Q), Academic Aptitude (AA), Pilot (P), and Navigator/Technical [also known as Combat Systems Officer (CSO)]. Table 1 shows the distribution of the test scores into each of the composites.

Table 1: AFOQT Form S Composite Composition

	V	Q	AA	P	CSO
Verbal Analogies (VA)	X		X		X
Arithmetic Reasoning (AR)		X	X	X	X
Word Knowledge (WK)	X		X		
Math Knowledge (MK)		X	X	X	X
Instrument Comp. (IC)				X	
Block Counting (BC)					X
Table Reading (TR)				X	X
Aviation Information (AI)				X	
Rotated Blocks (RB)					
General Science (GS)					X
Hidden Figures (HF)					

Note: RB and HF do not contribute to any composites score.

In August of 2014, the AFOQT Form T was introduced. This version removed Hidden Figures and Rotated Blocks and replaced them with Reading Comprehension and Situational Judgement (Situational Judgement is still considered experimental and not included in this assessment). Additionally, Physical Science replaced the General Science test. The Form T composite scores are: V, Q, AA, P, CSO, and Air Battle Management (ABM) (Carretta, King, Ree, Teachout, & Barto, 2016). Table 2 is a distribution of the Form T's ten tests into the re-designated six composites.

Table 2: AFOQT Form T Composite Composition

	V	Q	AA	P	CSO	ABM
Verbal Analogies (VA)	X		X			X
Arithmetic Reasoning (AR)		X	X			
Word Knowledge (WK)	X		X		X	
Math Knowledge (MK)		X	X	X	X	X
Instrument Comp. (IC)				X		X
Block Counting (BC)					X	X
Table Reading (TR)				X	X	X
Aviation Information (AI)				X		X
Reading Comprehension (RC)	X		X			
Physical Science (PS)						

Note: PS does not contribute to any composites score.

SDI+.

The SDI+ is a 220-question subtest of the AFOQT, which assesses the Big Five personality domains: Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness; and an Air Force specific measure: Machiavellianism. The Form S version was broken down into the six domains with twenty facets. The Form T maintained the six domains but was modified with the deletion of six facets and addition/update of sixteen facets for a total of thirty (Manley and Weissmuller 2017). Table 3 depicts the six domains and facets under both versions.

Table 3: AFOQT Form S and Form T SDI+ Domains and Facets

Domain	Form S Facets	Form T Facets
A – Agreeableness	Team Player	Team Player
	Pleasant	Pleasant
	Helpful-Altruistic	Helpful-Altruistic
	Considerate	Optimist
	Hyper Competitive	Well-Adjusted
N – Neuroticism	Stress-Under-Pressure	Stress-Under-Pressure
	Temperamental	Temperamental
	Worry	Worry
		Angry-Hostility
E - Extraversion	Unassertive	Reserved
	Sociable	Dominance-Leader
	Dominance	Excitement-Seeking
		High-Intensity Pleasure
		Activity
		Spontaneous-Variety
C - Conscientiousness	Achievement-Striving	Achievement-Striving
	Order	Order
		Self-Discipline
		Deliberation
		Unconventional
O - Openness	Creative	Creative
	Reflective	Reflective
	Scientific Interest	Scientific Interest
	Cultured	Cultured
		Imagination

M - Machiavellianism	Envious	Envious
	Individualistic	Cynical View
	Self-Serving	Interpersonal Tactics
		Influence Tactics
		Independent

TBAS/PCSM.

The Test of Basic Aviation Skills (TBAS) is a computer-administered cognitive and perceptual-motor test battery designed to measure pilot aptitude. The TBAS battery consists of eight subtests that assess psychomotor skills, psychomotor multitasking, and spatial orientation. (Rose et al. 2014). Weighted scores from the AFOQT Pilot Composite and TBAS composite are combined with a prior flying experience to form the PCSM score, a percentile ranking between 1-99 (Carretta, 2013).

Relevant Research

Several studies have validated the ability of the AFOQT to predict aircrew training success (Arth et al. 1990; Carretta 2008; 2013; Carretta and Ree 1995; 2000; Finegold and Rogers 1984; Olea and Ree 1993). Carretta and Ree noted that of the AFOQT subtests, the best predictors of success during Phase 1 was Arithmetic Reasoning; Phase 2 was Aviation Information and Instrument Comprehension; and Phase 3 was Scale Reading (1995). In addition to the AFOQT, Carretta determined the PSCM score correlation with T-6 completion was .53 (2011).

A study by Manley showed the internal consistency of the SDI+ domains as Agreeableness: .97, Neuroticism: .95, Extraversion: .96, Conscientiousness: .95, Openness: .89, and Machiavellianism: .75. The study further explains that Machiavellianism is lower due to it having less questions that feed into it. With only 9 questions that make up the scale, it has more than 2 to 5 times less than the other domains. The study concludes the SDI+ has many possible uses including the classification of members into Air Force Specialties for increased training success (2011).

Other studies have validated the use of cognitive and personality testing to individual pilot trainee performance (Carretta, 2011; Carretta et al., 2014; King et al., 2013; Teachout et al., 2013). Statistically significant relationships were found between the AFOQT, and SDI+, showing the USAF is using measures that correlated to some degree the performance of an individual in training.

Other studies link the AFOQT to success after pilot training. One such study by Rose et all, studied the success of pilot training by comparing AFOQT scores of graduated students to whether they were stratified on their first officer performance report (OPR). The study found that the Pilot composite and all the subtests which form it (AR, MK, IC, TR and AI) were predictors for stratification on the first OPR (2014).

The Gap

There has been a lot of research into the correlation of individual success through these measures. What has not been studied is the correlation of these tests to group success. If there is a linear relationship between cognitive and personality results amongst members in a class and the overall success (graduation rate) of that class, could class members be selected in a manner which forms a better team, and thereby decreases attrition rates?

III. Methodology

UPT Data

Air Force Personnel Center (AFPC) provided UPT student data from 2010 to 2018 which was used as a representative sample of pilot training classes. The data was for 12,001 students across all levels of training for a total of 27,897 lines of data. The data was scrubbed down to focus on T-6 training at Columbus Air Force Base (AFB), Laughlin AFB, and Vance AFB. While Sheppard AFB also has a T-6 training program, it was removed due to the uniqueness of its Euro-NATO Joint Jet Pilot Training Program (ENJJPT) program. After scrubbing the data, there were a total of 5,565 students, 406 of these were attritions.

The data includes personnel information, AFOQT scores, SDI+ scores, PCSM scores, and student performance scores and rankings. Some student information was incomplete, lending to gaps in the information, as illustrated by the accession source and demographics tables below, showing less than 5,565 personnel.

Exploring the Data

Using the available data, attritions were examined by accession source, demographics, base, and type.

Table 4: Attrition by Accession Source

	USAFA	ROTC	OTS	Other	Total
Source Total	1162	2485	1384	463	5031
Attritions	84	199	90	28	373
Attrition %	7.23%	8.01%	6.50%	6.05%	
Average Age	20.65	20.00	24.31	24.51	

Table 4 details attritions by accession source. Of the three main accession sources, the majority of candidates came from the Reserve Officer Training Core (ROTC). This is understandable because ROTC programs are in hundreds of civilian universities across the nation and supply a large body of candidates. ROTC candidates also suffer from the highest attrition rate which may be harder to explain. Some would argue that United States Air Force Academy (USAFA) students have a strict college regiment so transitioning to pilot training may be less of a culture shock than to those transitioning from a civilian institution. Officer Training School (OTS) candidates have the lowest graduation rate of the three main sources. This could be for a multitude of reasons. One reason could be a greater maturity level as the average age of OTS accessions is roughly four years older than ROTC and USAFA candidates. Or maybe the OTS rate is driven down by those candidates with prior enlisted, and possibly aircrew, experience who are already familiar with the structure and expectations of the military and military training.

Table 5: Attrition by Demographics

	Gender Totals	**Hispanic**	**Indian/ Alaska Native**	**Asian**	**Black or African American**	**Hawaiian / Other Pacific Islander**	**White**
Male	5068	273	26	182	156	43	4399
Attritions	351	44	2	18	24	5	265
Male %	6.93%	16.12%	7.69%	9.89%	15.38%	11.63%	6.02%
Female	451	20	7	18	9	6	403
Attritions	49	5	1	1	4	0	39
Female %	10.86%	25.00%	14.29%	5.56%	44.44%	0.00%	9.68%

Table 5 shows a breakdown of the demographics for T-6 training. There is a large disparity between the number of white males and non-white males. Additionally, white males have the lowest attrition rate with the exception of Female Asians and

Hawaiian/Other Pacific Islander. The highest attrition rate is Female Black or African American. This inequality deserves further exploration but is outside the scope of this research.

Table 6: Attritions by Base

	Columbus AFB	Laughlin AFB	Vance AFB	Total
Students	1839	1904	1822	5565
Student %	33.05%	34.21%	32.74%	
Attritions	138	138	130	406
Attrition %	7.50%	7.25%	7.14%	7.30%

Table 6 lists the three bases used for this research and their number of students and attritions. The students were fairly evenly spread across the bases. Of the 406 attritions, Columbus AFB and Laughlin AFB were tied with 138 each but the highest percentage of attritions, based on students in attendance, was Columbus AFB.

Table 7: Attrition Reason by Base

	Columbus AFB	Laughlin AFB	Vance AFB	Total
Flying	16.01%	16.26%	19.70%	51.97%
DOR	7.39%	4.93%	4.68%	17.00%
Medical	2.96%	3.94%	2.96%	9.85%
Academic	2.96%	1.97%	1.48%	6.40%
Other	2.71%	4.68%	2.22%	9.61%
Fear / MOA	1.97%	2.22%	0.74%	4.93%
Military	0.00%	0.00%	0.25%	0.25%
Total	33.99%	33.99%	32.02%	

Attritions by base and reason are listed in Table 7. Of the seven types of attrition reasons, flying was the greatest at nearly 52%. Columbus produced a higher than normal number of DOR attritions. Laughlin had higher than normal Medical and Other attritions, and Vance was above average for Flying attritions. The average graduation percentage for classes at each base is as follows: Columbus AFB: 92.8%, Laughlin AFB: 93.9%, and Vance AFB: 93.2%

After examination of the T-6 student data, the students were separated by their classes. This produced 302 classes with a range of 3 to 33 USAF students. Figure 2 shows the distribution of graduation percentage by class size. The figure does not depict the number of classes within each data point.

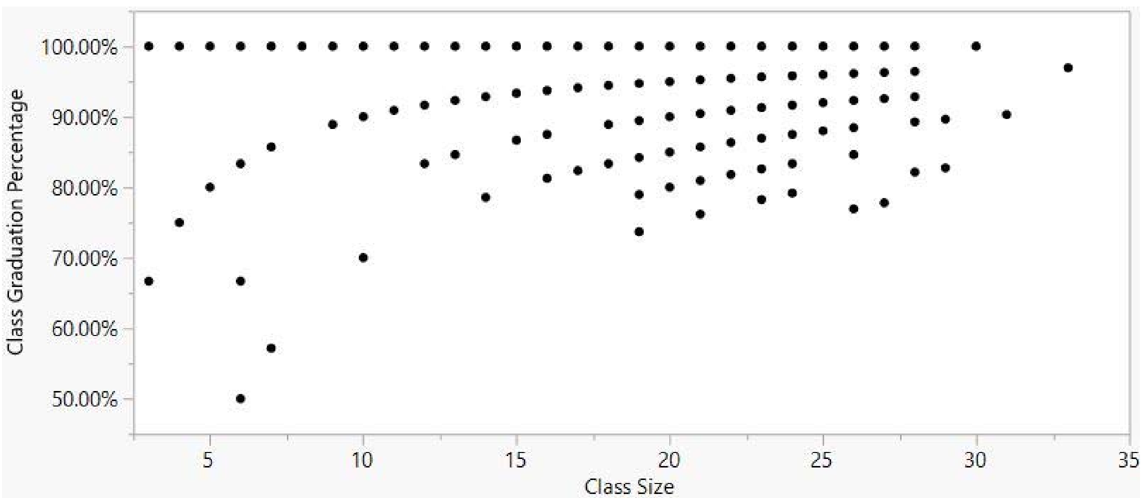


Figure 2: Class Size Class by Graduation Percentage

Some classes were missing data from multiple students. For example, the class size was 24 but data was available for only 4 of those students. When comparing class dynamics, it was deemed important to have sufficient data from each class. So only classes with data for at least 75% of the class was included in the analysis. Once the classes with less than 75% of the student data were removed, the data set contained 188 viable classes for analysis.

Class Graduation Percentage

A metric was required to compare performance against. Since attritions are not assigned an overall performance score, a new metric was created. The total number of

USAF graduating students was divided by the total number of USAF students to attain a graduation percentage of USAF students (Equation 1). The new metric became the dependent variable for all further analysis.

$$y = \frac{Total\ USAF\ Graduating\ Students}{Total\ USAF\ Students} = USAF\ Graduation\ Percentage \qquad (1)$$

The independent (x) variables were the scores from the AFOQT and SDI+ listed in Table 8. In addition to these, the Class size, PCSM Score, Flight Hours, and GPA (at the time of application) were also considered. After the students were grouped by class, the individual student scores were averaged to compare against that class's graduation percentage.

Table 8: Available AFOQT and SDI+ Scores

AFOQT Composite	AFOQT Raw Scores	SDI+ Domain	SDI+ Facet Scores
Pilot	Verbal Analogies	Agreeableness	A - Team Player
CSO	Arithmetic Reasoning	Neuroticism	A - Pleasant
Academic	Word Knowledge	Extraversion	A - Considerate
Verbal	Math Knowledge	Conscientiousness	A - Helpful Altruistic
Quantitative	Instrument Comprehension	Openness	A - Hyper-Competitive
	Block Counting	Machiavellianism	N - Stress Under Pressure
	Table Reading		N - Temperamental
	Aviation Information		N - Worry
	Rotated Blocks		E - Reserved
	Hidden Figures		E - Sociable
	Data Interpretation		E - Dominance
	Electrical Knowledge		C - Achievement Striving
	Scale Read		C - Order
			O - Creative
			O - Reflective
			O - Scientific Interest
			O - Cultured
			M - Envious
			M - Individualistic
			M - Self Serving

IV. Analysis and Results

Overview

The 188 classes were analyzed using JMP Pro 13. An initial correlation and regression were accomplished to find significant variables. Then a Stepwise function was utilized to determine if there was a good model which correlated score averages and standard deviations to the success of the class. That model was tweaked for currency and then simplicity. Correlations and significant variables were analyzed and discussed.

Initial Regression Exploration of Mean Scores

The AFOQT raw scores and SDI+ facet scores were normalized between the Form S and Form T by their absolute lows and highs for a value between 0 and 100. After the data was normalized, a correlation matrix was run to compare the components to the graduation percentage. The components with the highest correlation were Instrument Comprehension, Pilot Composite, Scale Read, Flight Hours, and E - Sociable. The entire list of correlation results is in Table 9.

Table 9: Correlation Matrix: Mean Scores to Graduation Percentage

Category	Correlation	Category	Correlation
Instrument Comprehension	0.20830	Openness	-0.06183
Pilot Composite	0.17966	Agreeableness	-0.06064
Scale Read	0.16240	M - Individualistic	-0.05955
Flight Hours	-0.15522	Hidden Figures	-0.05738
E - Sociable	-0.15174	C - Achievement Striving	-0.05674
Rotated Blocks	0.14814	O - Reflective	-0.05518
Block Counting	0.12774	A - Pleasant	-0.05372
Extraversion	0.12156	Conscientiousness	-0.05232
Quantitative Composite	0.12017	O - Creative	-0.04718
A - Hyper-Competitive	0.11800	E - Reserved	0.04066
Electrical Knowledge	-0.11673	Table Reading	0.03748

N - Worry	-0.10908	M - Envious	0.03583
A - Helpful Altruistic	-0.10896	Verbal Analogies	-0.03399
Math Knowledge	0.10830	Verbal Composite	0.02814
Arithmetic Reasoning	0.10387	A - Team Player	-0.02712
PCSM Score	0.09122	Data Interpretation	0.02246
E - Dominance	-0.08513	N - Stress Under Pressure	0.02245
Academic Composite	0.08016	Neuroticism	0.02019
O - Cultured	-0.07970	M - Self Serving	0.01801
Aviation Information	0.07867	Word Knowledge	0.01297
CSO Composite	0.07676	C - Order	0.01082
Class Size	-0.07253	O - Scientific Interest	0.00495
N - Temperamental	0.06828	Machiavellianism	0.00261
A - Considerate	-0.06818	GPA	-0.00115

Using JMP, an ordinary least squares regression was run on the top ten correlated components to determine if they would produce a good model. When that failed, a regression was completed using all the components. Extensive multi-collinearity was found because some scores go into others. When the composite and raw/domain and facet scores were analyzed separately it reduced some of the multi-collinearity and highlighted some significant variables. Significant variables are those with a *p*-value less than 0.05.

AFOQT Composite Score Regression.

Figure 3 show the regression of the AFOQT composite scores. Academic Aptitude has a lot of multi-collinearity, as shown by its high variance inflation factor (VIF) score. This is explained by the composite being made up of elements from the Verbal and Quantitative Composites. If the Academic Aptitude Composite is removed, the other VIF scores drop below seven, and the pilot composite appears as a significant variable with a *p*-value of 0.0255 (Figure 4).

Parameter Estimates					
Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	0.8129393	0.147776	5.50	<.0001*	.
Pilot Composite	0.0026223	0.001676	1.56	0.1206	2.0364073
CSO Composite	-0.003092	0.002595	-1.19	0.2360	6.6105876
Academic Aptitude Composite	0.007648	0.02166	0.35	0.7247	370.92379
Verbal Composite	-0.003177	0.012052	-0.26	0.7926	123.81062
Quantitative Composite	-0.002936	0.012041	-0.24	0.8078	109.77999

Figure 3: AFOQT Composite Regression Analysis

Parameter Estimates					
Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	0.7191198	0.079104	9.09	<.0001*	.
Pilot Composite	0.0027943	0.00124	2.25	0.0255*	2.0798692
CSO Composite	-0.001812	0.001907	-0.95	0.3435	6.3431361
Verbal Composite	-7.661e-5	0.001508	-0.05	0.9595	3.6301545
Quantitative Composite	0.0016655	0.001195	1.39	0.1652	1.97632

Figure 4: AFOQT Composite Regression Analysis Without Academic Aptitude

AFOQT Raw Score Regression.

When analyzing the AFOQT raw scores (Figure 5), they all appear to be insignificant until Scale Read is removed because of the high VIF score (Figure 6). Once removed, the Instrument Comprehension *p*-value drops to 0.0301 and other high VIF scores drop below three.

Parameter Estimates					
Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	156.23102	68.2535	2.29	0.0233*	.
Verbal Analogies	-0.473104	0.256422	-1.85	0.0667	2.8682546
Arithmetic Reasoning	0.0908617	0.166793	0.54	0.5866	2.051223
Word Knowledge	0.0519272	0.148165	0.35	0.7264	2.0640489
Math Knowledge	-0.456427	0.353266	-1.29	0.1981	10.178146
Instrument Comprehension	0.0749243	0.257696	0.29	0.7716	5.0341934
Block Counting	-0.116009	0.269707	-0.43	0.6676	3.7098892
Table Reading	-0.618952	0.367984	-1.68	0.0944	10.614574
Aviation Information	-0.252764	0.199517	-1.27	0.2069	5.00102
Rotated Blocks	0.1149031	0.128193	0.90	0.3713	1.8151714
Hidden Figures	-0.169541	0.118604	-1.43	0.1547	1.4806208
Data Interpretation	-0.022576	0.701796	-0.03	0.9744	1.7679281
Electrical Knowledge	-0.098244	0.075652	-1.30	0.1958	1.7092908
Scale Read	0.9374307	0.576777	1.63	0.1059	52.541648

Figure 5: AFOQT Raw Scores Regression Analysis

Parameter Estimates					
Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	89.153491	54.61648	1.63	0.1044	.
Verbal Analogies	-0.236517	0.212087	-1.12	0.2663	1.9439278
Arithmetic Reasoning	0.0582649	0.166357	0.35	0.7266	2.0215646
Word Knowledge	0.0531774	0.148856	0.36	0.7213	2.0639933
Math Knowledge	0.0555719	0.160622	0.35	0.7298	2.0845968
Instrument Comprehension	0.3832412	0.175233	2.19	0.0301*	2.3061766
Block Counting	0.1281453	0.225046	0.57	0.5698	2.5589742
Table Reading	-0.073661	0.151867	-0.49	0.6283	1.7910939
Aviation Information	0.0185121	0.109821	0.17	0.8663	1.501113
Rotated Blocks	0.1526997	0.126656	1.21	0.2296	1.7554358
Hidden Figures	-0.193538	0.118232	-1.64	0.1034	1.4576743
Data Interpretation	-0.146036	0.700938	-0.21	0.8352	1.7472156
Electrical Knowledge	-0.116714	0.075144	-1.55	0.1222	1.67072

Figure 6: AFOQT Raw Scores Regression Analysis without Scale Read

SDI+ Domain and Facet Score Regression.

The SDI+ domain scores all appear to be insignificant (Figure 7), but Extraversion is the most significant with a *p*-value of 0.0899 and Machiavellianism was the least significant with a *p*-value of 0.8770. Figure 8 shows the regression of the facet scores. N - Worry, E - Sociable, and M - Individualistic, have a *p*-value of 0.0399, 0.0103, and 0.0349 respectively (Figure 8).

Parameter Estimates					
Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	105.9005	22.99337	4.61	<.0001*	.
Agreeableness	-0.106045	0.244558	-0.43	0.6651	4.1434921
Neuroticism	-0.225923	0.233216	-0.97	0.3340	3.3529251
Extraversion	0.238166	0.139701	1.70	0.0899	1.3221306
Conscientiousness	-0.042273	0.182568	-0.23	0.8172	2.724359
Openness	-0.120248	0.17698	-0.68	0.4977	1.3246309
Machiavellianism	-0.030008	0.193535	-0.16	0.8770	1.6307694

Figure 7: SDI+ Domain Regression Analysis

Parameter Estimates					
Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	134.92422	36.55726	3.69	0.0003*	.
A - Team Player	-0.013182	0.36087	-0.04	0.9709	10.903694
A - Pleasant	0.1575617	0.26628	0.59	0.5548	5.4298674
A - Considerate	0.2341342	0.198135	1.18	0.2390	2.6601622
A - Helpful Altruistic	-0.258287	0.237859	-1.09	0.2791	3.6024865
A - Hyper-Competitive	0.3427635	0.19938	1.72	0.0874	2.6022905
N - Stress Under Pressure	-0.217234	0.324327	-0.67	0.5039	6.7782199
N - Temperamental	0.5307697	0.293447	1.81	0.0723	6.6527644
N - Worry	-0.416018	0.200919	-2.07	0.0399*	1.8743458
E - Reserved	-0.381093	0.288046	-1.32	0.1876	5.3046379
E - Sociable	-0.640812	0.247017	-2.59	0.0103*	3.5950338
E - Dominance	-0.251708	0.217169	-1.16	0.2481	2.1294533
C - Achievement Striving	-0.27041	0.330819	-0.82	0.4149	10.465604
C - Order	0.3102329	0.217474	1.43	0.1556	3.3250119
O - Creative	0.2053717	0.242467	0.85	0.3982	4.2055636
O - Reflective	0.105737	0.207119	0.51	0.6104	1.8655178
O - Scientific Interest	0.0797359	0.194726	0.41	0.6827	1.884103
O - Cultured	-0.112523	0.190509	-0.59	0.5556	1.3398797
M - Envious	0.1179841	0.17964	0.66	0.5122	1.8029557
M - Individualistic	-0.377999	0.177692	-2.13	0.0349*	1.6111751
M - Self Serving	-0.029358	0.186508	-0.16	0.8751	1.7002645

Figure 8: SDI+ Facet Regression Analysis

Miscellaneous Factors Regression.

An analysis of Class Size, Flight Hours, PCSM Score and GPA showed Flight Hours and PCSM score were significant with a *p*-value of 0.0019 and 0.0224 respectively (Figure 9).

Parameter Estimates					
Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	0.9747191	0.079527	12.26	<.0001*	.
Class Size	-0.001684	0.001218	-1.38	0.1685	1.0928336
Flight Hours	-0.000184	5.811e-5	-3.16	0.0019*	1.219636
PCSM Score	0.0015579	0.000677	2.30	0.0224*	1.4707943
GPA	-0.026276	0.027899	-0.94	0.3475	1.3176708

Figure 9: Class Size, Flight Hours, PCSM Score and GPA Regression Analysis

Correlations.

Figures 10 and 11 are charts of the scores with the lowest *p*-values from the regression analysis. These charts are a visual representation of the correlations found in

JMP. PCSM, Pilot Composite, and Instrument Comprehension show a positive correlation and Sociable and Flight Hours show a negative correlation.

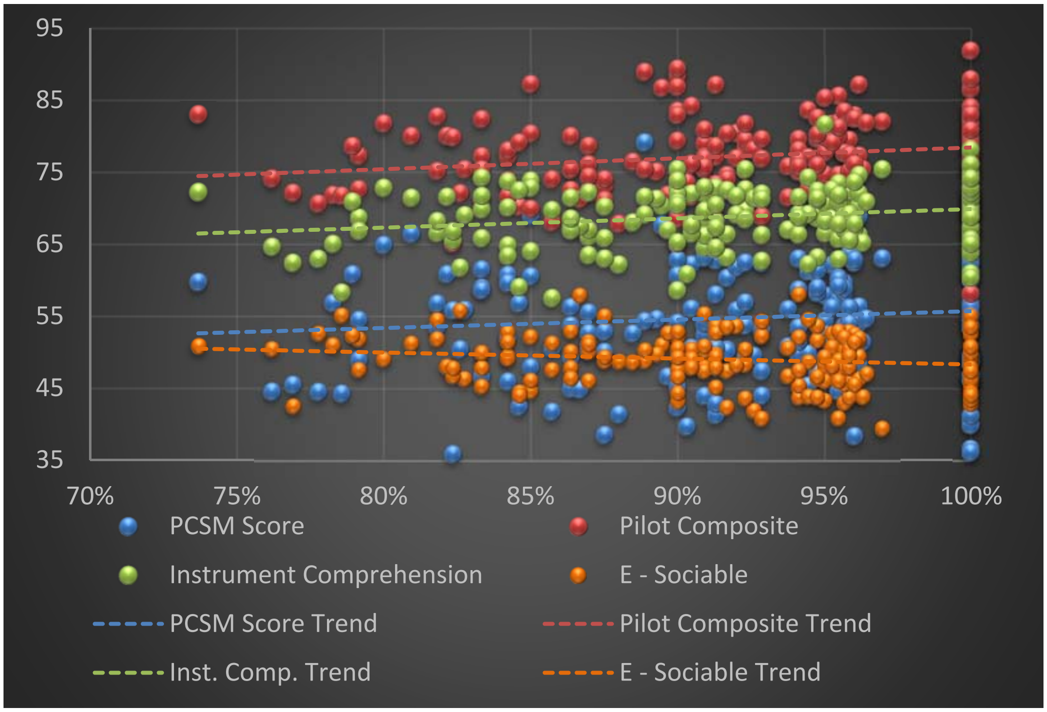


Figure 10: Significant Components by Graduation Percentage

The Pilot Composite and PCSM score's positive correlation seems self-explanatory. Since these scores are designed to measure pilot aptitude, then a higher average score should be indicative of a higher graduation rate. Instrument Comprehension measures the ability to recognize an aircraft's attitude through provided instrument pictures (Weissmuller and Schwartz 2007). As it is vital for pilots to be able to read their instruments, it makes sense for this test score to be positively correlated to the class success.

Sociable is negatively correlated, which according to its definition in Attachment 2, seems slightly counter-intuitive. The pilot stereotype is not that of a loner, so an average below 55 begs the questions: are pilots more introverted than they appear or do those not focused on being social, have better study habits?

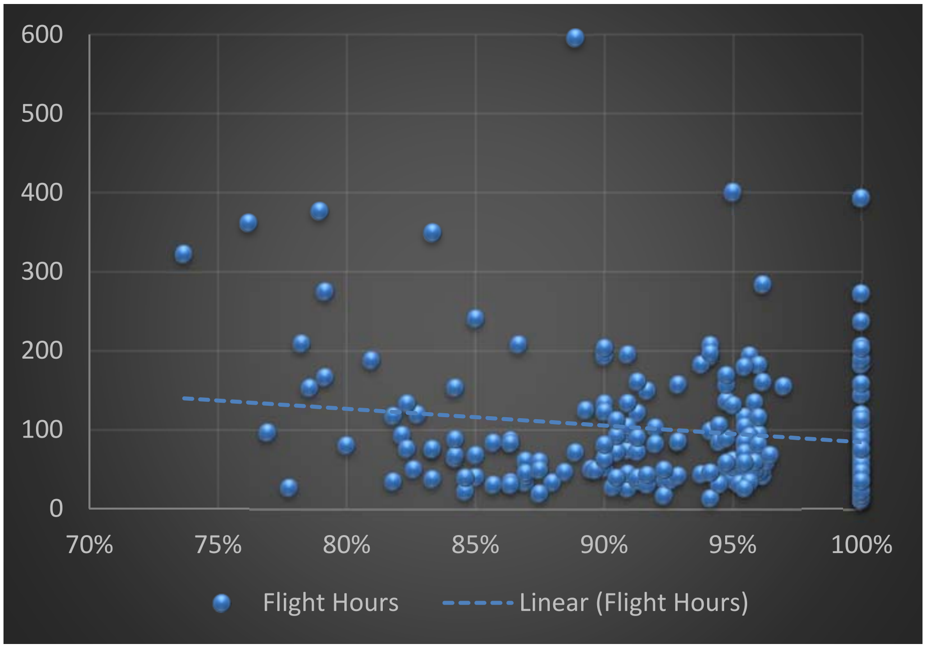


Figure 11: Flight Hours by Graduation Percentage

The negative correlation of average Flight Hours seems intuitive to a small degree. From assessing individual performance factors, increased flight hours are positively correlated to enhanced performance in pilot training. It is possible that by evaluating the performance of the high performing individuals in a group that the low performing individuals would appear even worse through comparison leading to further attrition.

Stepwise Regression Analysis

Using the most significant variables from the initial exploration failed to produce a good model, so the JMP Stepwise function was used. The data was analyzed with a validation breakdown of 60% Training (*n*=113), 20% Validation (*n*=37), and 20% Test (*n*=38). The function was set to produce the best Validation R2. From this, two models were created, one for Form S analysis, and one for Form T analysis.

Results of Regression Analysis, Form S Model.

This model was labeled as the Form S Model because it contains a test element from the AFOQT Form S which was removed from the Form T version. It was the first created by JMP and contained seven variables: Flight Hours; three raw AFOQT scores: Verbal Analogies, Instrument Comprehension, and Hidden Figures; and three SDI+ facets: N-Temperamental, N-Worry, and E-Dominance. Figure 12 shows the model analysis. The Root Average Squared Error (RASE) delta was 0.0090, or 0.9% average variation. This indicates the model created with the training set was a good fit for the test set as well.

Summary of Fit

RSquare	0.138769
RSquare Adj	0.081354
Root Mean Square Error	0.053418
Mean of Response	0.927169
Observations (or Sum Wgts)	113

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	0.04827648	0.006897	2.4169
Error	105	0.29961362	0.002853	Prob > F
C. Total	112	0.34789010		0.0246*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.1614861	0.238643	4.87	<.0001*
Flight Hours	-8.019e-5	0.000061	-1.31	0.1916
Verbal	-0.002292	0.001896	-1.21	0.2295
Instrument Comprehension	0.0052739	0.0016	3.30	0.0013*
Hidden Figures	-0.001617	0.001382	-1.17	0.2446
N - Temperamental	0.24631	0.178125	1.38	0.1697
N - Worry	-0.408454	0.213944	-1.91	0.0590
E - Dominance Leader	-0.423749	0.200061	-2.12	0.0365*

Effect Tests

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.1388	0.05149	113
Validation Set	0.1178	0.06816	37
Test Set	0.1584	0.06051	38

Figure 12: Form S Model JMP Regression Analysis with Validation

When this model was applied to the entire data set, the results showed four significant variables: Flight Hours, Instrument Comprehension, N-Temperamental, and N-Worry. Figure 13 shows the model's Summary of Fit, Analysis of Variance, Parameter Estimates, and Profiler outputs. The F(Model)=5.2911 > $F_{[.05]}$=2.0608 (where k=7 numerator degrees of freedom and n=188, so n-(k+1)=180 denominator degrees of freedom), and the model's p-value is less than 0.05, both suggesting, that at a 0.05 level of significance, this model of variables is better than a model with only the intercept.

Whole Model

Summary of Fit

RSquare Adj	0.138398
Root Mean Square Error	0.057448
Mean of Response	0.921929
Observations (or Sum Wgts)	188

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	0.12223611	0.017462	5.2911
Error	180	0.59405876	0.003300	Prob > F
C. Total	187	0.71629486		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	0.9946262	0.201254	4.94	<.0001*	.
Flight Hours	-0.000146	5.331e-5	-2.74	0.0068*	1.1386293
Verbal Analogies	-0.002285	0.001542	-1.48	0.1401	1.1342554
Instrument Comprehension	0.006294	0.001312	4.80	<.0001*	1.4272391
Hidden Figures	-0.001529	0.001068	-1.43	0.1540	1.3123595
N - Temperamental	0.003202	0.00142	2.26	0.0253*	1.7326052
N - Worry	-0.004276	0.001634	-2.62	0.0096*	1.379455
E - Dominance	-0.003008	0.001578	-1.91	0.0582	1.2513277

Prediction Profiler

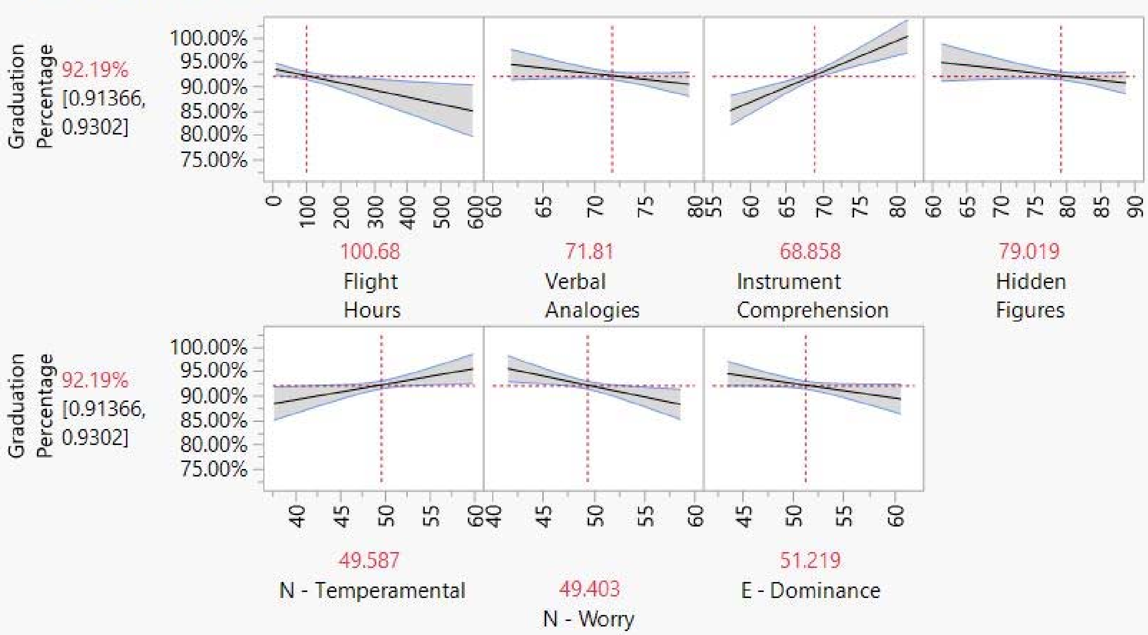


Figure 13: Form S Model JMP Regression Analysis without Validation

Figure 14 shows the actual versus predicted graduation rates when using the Form S Model. Figure 15 is a summary of the residual's distribution.

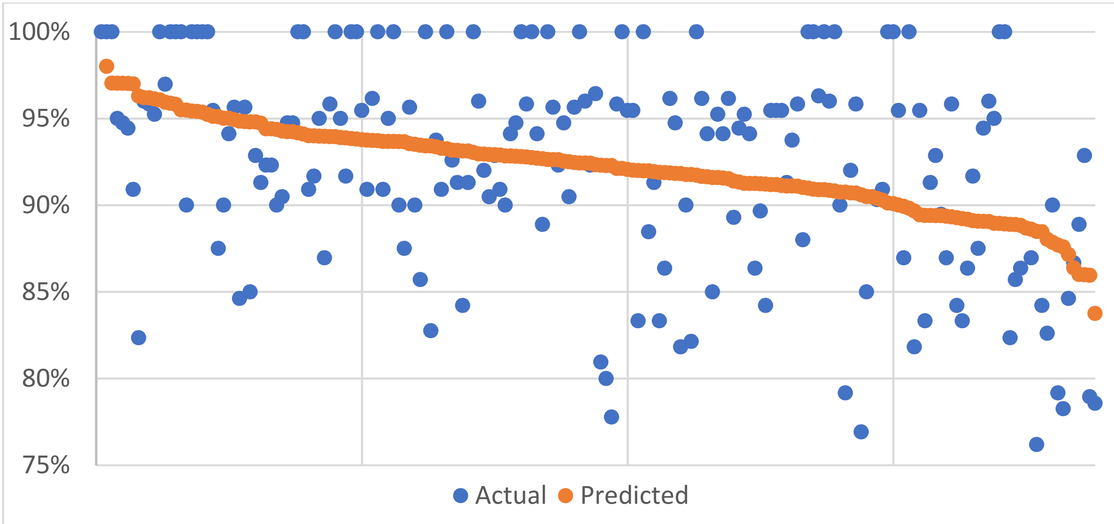
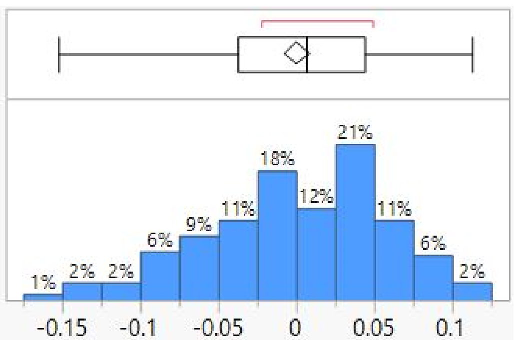


Figure 14: Form S Model Actual vs Predicted Graduation Percentage



Quantiles

100.0%	maximum	0.1118557711
99.5%		0.1118557711
97.5%		0.0983188256
90.0%		0.0673019616
75.0%	quartile	0.0432048266
50.0%	median	0.0060631783
25.0%	quartile	-0.037043082
10.0%		-0.079208535
2.5%		-0.126917417
0.5%		-0.151956816
0.0%	minimum	-0.151956816

Summary Statistics

Mean	-3.19e-17
Std Dev	0.0566005
Std Err Mean	0.004128
Upper 95% Mean	0.0081435
Lower 95% Mean	-0.008143
N	188

Figure 15: Form S Model JMP Residuals Distribution

Results of Regression Analysis, Form T Model.

The first model is good for assessing student composition under the Form S version but not the Form T. In order to assess Form T components only, slight tweaks were made to the Stepwise function. By removing those raw scores not on the Form T version, another model appeared in which Hidden Figures was replaced with Word Knowledge, but all other components remained, as shown in Figure 16. The RASE delta was 0.0097, or 0.97% average variation.

Summary of Fit

RSquare	0.136119
RSquare Adj	0.078526
Root Mean Square Error	0.053503
Mean of Response	0.927171
Observations (or Sum Wgts)	113

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	0.04736014	0.006766	2.3635
Error	105	0.30057281	0.002863	**Prob > F**
C. Total	112	0.34793295		0.0278*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	1.0761486	0.227404	4.73	<.0001*	.
Flight Hours	-0.000108	5.931e-5	-1.83	0.0706	1.171188
Verbal Analogies	-0.004225	0.002547	-1.66	0.1001	2.0681066
Word Knowledge	0.0016894	0.001653	1.02	0.3090	1.9264527
Instrument Comprehension	0.0049395	0.001572	3.14	0.0022*	1.4727301
N - Temperamental	0.0021204	0.001819	1.17	0.2464	2.0815817
N - Worry	-0.003719	0.002135	-1.74	0.0844	1.5875393
E - Dominance	-0.004186	0.002016	-2.08	0.0403*	1.251227

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.1361	0.05157	113
Validation Set	0.1135	0.06833	37
Test Set	0.1371	0.06127	38

Figure 16: Form T Model JMP Regression Analysis with Validation

When the Form T Model was applied to the entire data set, the results were similar to the Form S Model, with the same significant variables. Figure 17 shows the model's Summary of Fit, Analysis of Variance, Parameter Estimates, and Profiler. The

F(Model)=5.0314 > $F_{[.05]}$=2.0608 (maintaining the same *k* and *n* from the Form T Model) and the model's *p*-value is less than 0.0001, as with the Form S Model.

Summary of Fit

RSquare	0.163645
RSquare Adj	0.131121
Root Mean Square Error	0.057691
Mean of Response	0.921929
Observations (or Sum Wgts)	188

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	0.11721837	0.016745	5.0314
Error	180	0.59907650	0.003328	Prob > F
C. Total	187	0.71629486		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	0.9158469	0.196068	4.67	<.0001*	.
Flight Hours	-0.000168	5.287e-5	-3.18	0.0018*	1.1105373
Verbal Analogies	-0.003538	0.001994	-1.77	0.0777	1.8809338
Word Knowledge	0.0009626	0.00133	0.72	0.4700	1.8022195
Instrument Comprehension	0.0058937	0.001279	4.61	<.0001*	1.3449293
N - Temperamental	0.0030155	0.001455	2.07	0.0396*	1.8046161
N - Worry	-0.003839	0.00163	-2.35	0.0196*	1.3618217
E - Dominance	-0.003047	0.001592	-1.91	0.0571	1.2623295

Prediction Profiler

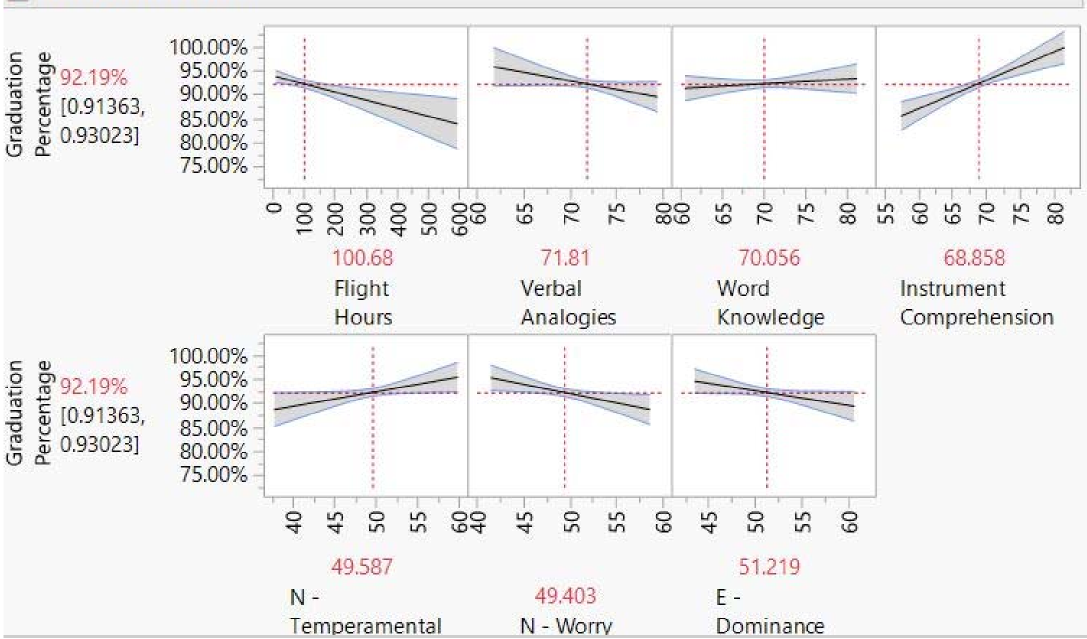


Figure 17: Form T Model JMP Regression Analysis without Validation

Figure 18 shows the Actual versus Predicted Graduation Percentage using the Form T Model. Figure 19 shows the residual's distribution.

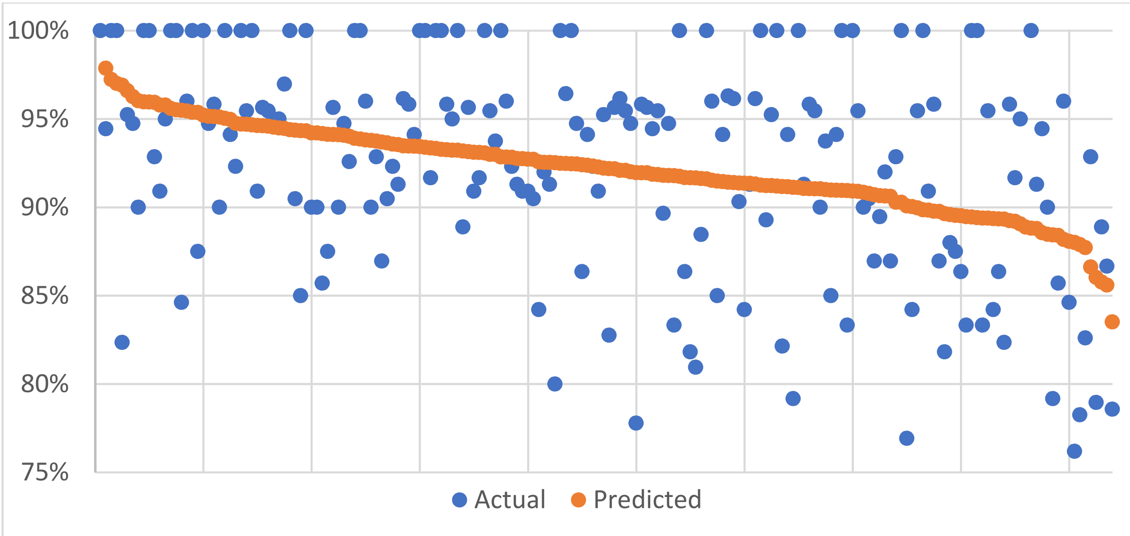
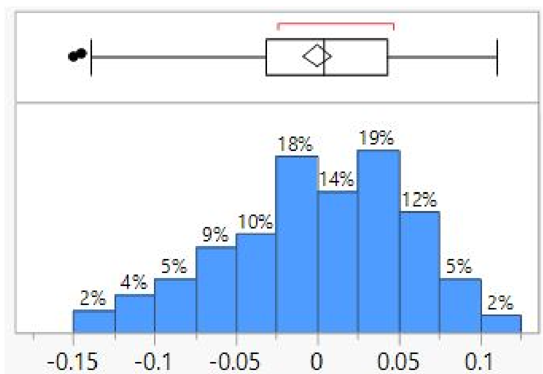


Figure 18: Form T Model Actual vs Predicted Graduation Percentage



Quantiles

100.0%	maximum	0.11086324
99.5%		0.11086324
97.5%		0.0989098545
90.0%		0.0688459894
75.0%	quartile	0.0426967071
50.0%	median	0.0036588451
25.0%	quartile	-0.031466958
10.0%		-0.079144483
2.5%		-0.126714067
0.5%		-0.149905271
0.0%	minimum	-0.149905271

Summary Statistics

Mean	-2.54e-17
Std Dev	0.056363
Std Err Mean	0.0041107
Upper 95% Mean	0.0081093
Lower 95% Mean	-0.008109
N	188

Figure 19: Form T Model JMP Residuals Distribution

Results of Regression Analysis, FINN Model.

The Form S and Form T Models were created using the Stepwise function in JMP, providing an ideal model with an optimized R2. The results of both models indicated the same four significant variables. Running those variables as a model by themselves also created a good model (Figure 20). The RASE delta between the training and test set for this model is 0.00732 or 0.732%, indicating this model's predicted variables were slightly better than the first two models. This simplistic model was dubbed the FINN Model.

Summary of Fit

RSquare	0.07426
RSquare Adj	0.039973
Root Mean Square Error	0.054611
Mean of Response	0.927171
Observations (or Sum Wgts)	113

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	0.02583742	0.006459	2.1658
Error	108	0.32209553	0.002982	**Prob > F**
C. Total	112	0.34793295		0.0777

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	0.6898578	0.161637	4.27	<.0001*	.
Flight Hours	-6.086e-5	5.776e-5	-1.05	0.2944	1.0658581
Instrument Comprehension	0.0036725	0.001515	2.42	0.0170*	1.3132602
N - Temperamental	0.0033088	0.001784	1.85	0.0663	1.9212749
N - Worry	-0.003498	0.002162	-1.62	0.1085	1.5625787

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.0743	0.05339	113
Validation Set	0.0931	0.06911	37
Test Set	0.1528	0.06071	38

Figure 20: FINN Model JMP Regression Analysis with Validation

When the FINN Model was applied to the entire data set, the results were similar to the other models. Figure 21 shows the model's Summary of Fit, Analysis of Variance, and Parameter Estimates. The profile of the four variables is the same as in Figures 14 and 18. The F(Model)=7.0234 > $F_{[.05]}$= 2.4205 (where k=4 numerator degrees of freedom and n=188, so n-(k+1)=185 denominator degrees of freedom) and the model's p-value is less than 0.0001, as with the other models.

Summary of Fit

RSquare	0.133086
RSquare Adj	0.114138
Root Mean Square Error	0.058252
Mean of Response	0.921929
Observations (or Sum Wgts)	188

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	0.09532915	0.023832	7.0234
Error	183	0.62096571	0.003393	**Prob > F**
C. Total	187	0.71629486		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	0.5780614	0.134468	4.30	<.0001*	.
Flight Hours	-0.000138	5.182e-5	-2.66	0.0085*	1.0464141
Instrument Comprehension	0.0050044	0.001243	4.03	<.0001*	1.2457567
N - Temperamental	0.0042524	0.001368	3.11	0.0022*	1.5638633
N - Worry	-0.004002	0.001635	-2.45	0.0153*	1.3432706

Figure 21: FINN Model JMP Regression Analysis without Validation

Figure 22 shows the actual versus predicted graduation percentage. Figure 23 shows the distribution of the residuals. This model's residuals are disbursed slightly wider than the other two models, signifying that a small amount of accuracy was sacrificed for simplicity.

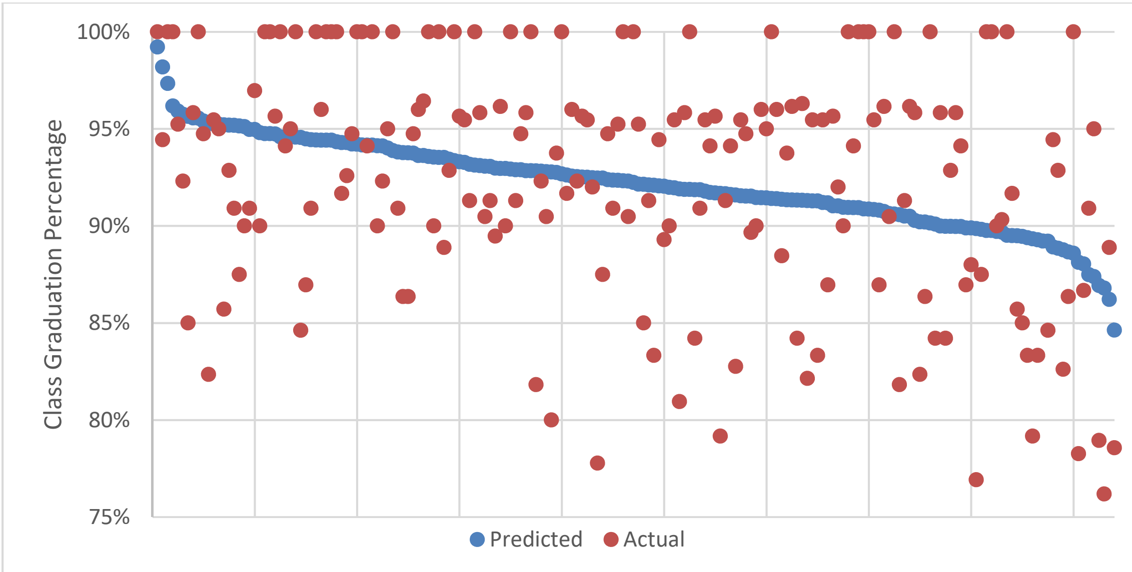
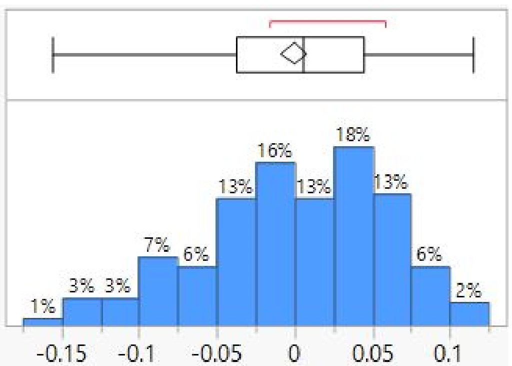


Figure 22: FINN Model Actual vs Predicted Graduation Percentage



Quantiles

100.0%	maximum	0.1140979168
99.5%		0.1140979168
97.5%		0.0995852209
90.0%		0.0688886297
75.0%	quartile	0.0439668364
50.0%	median	0.0046512191
25.0%	quartile	-0.037679794
10.0%		-0.080579704
2.5%		-0.128196666
0.5%		-0.155346339
0.0%	minimum	-0.155346339

Summary Statistics

Mean	-2.13e-17
Std Dev	0.0576253
Std Err Mean	0.0042028
Upper 95% Mean	0.0082909
Lower 95% Mean	-0.008291
N	188

Figure 23: FINN Model Residuals Distribution

Model Comparison.

The test data root average square error (RASE) for the Form S, Form T and FINN models was 0.06051, 0.06127 and 0.06071 respectively. The Form S had the lowest error of any model on the test data, the FINN model came in second and Form T model performed the worst. The three models represent different variations of relationships between mean cognitive and personality test scores of the class to the class graduation percentage. When comparing the Form S to the Form T, they are very similar in outcomes. When comparing the validation against the models, the RASE delta was 0.0090 for the Form S and 0.0097 for the Form T showing there was little difference between the predicted graduation percentages. Of the two, the Form T Model is probably more useful as the Form S version is becoming obsolete.

When comparing the Form T and FINN models the FINN model appears to be better. The RASE delta of the FINN model was better at only 0.0073, showing a better average predicted graduation percentage. This model is also a simplistic model and therefore easier to implement.

Exploring the Significant Variables

The correlations of Flight Hours and Instrument Comprehension remained the same although their coefficients were reduced. N - Temperamental was positively correlated and N - Worry was negatively correlated. According to the definitions of the facets in Appendix A, N - Worry seems intuitive, but N - Temperamental does not. Too much worry could be detrimental to the group, but an average amount (as indicated by the profiler) could lead to healthy amount of concern and induce studying. One would think a person who is more emotionally stable and less erratic would be desirable in a high stress situation like pilot training.

The four significant variables were further analyzed to look for patterns between attritions and graduates. Table 10 shows the average number of Flying Hours and average scores for Instrument Comprehension, N-Temperament, and N-Worry. The attrition average flying hours is less than the graduates suggesting a more flying experience prior to UPT is an advantage to the individual. Instrument Comprehension has a similar finding, that a greater understanding of instruments is beneficial to the individual. On the other hand, the average scores for N - Temperamental and N - Worry are only slightly higher than those of the graduates.

Table 10: Comparison of Averages Scores of Attritions and Graduates

Component	All Attritions	Graduates
Flying Hours	38.77	114.63
Instrument Comprehension	63.84	70.09
N - Temperamental	48.48	47.47
N - Worry	49.84	48.13

Table 11 details the significant variables and their correlation to the performance of graduates using the total merit assignment selection system score (TOTMASS) and performance of the class using the graduation percentage. The TOTMASS is a sum of four weighted scores: Category Check T-score, Daily Maneuver T-score, Academic T-score, FLT/CC Ranking T-score, and is 91.7% correlated to class ranking for this data set. The correlation to graduation percentage data listed in Table 11 is the same as that listed in Table 9.

Table 11: Significant Variable Correlation to TOTMASS of Graduates

Component	Correlation to TOTMASS	Correlation to Graduation %
Flight Hours	0.21377	-0.15522
Instrument Comprehension	0.21635	0.20830
N - Temperamental	-0.04750	0.06828
N – Worry	-0.02645	-0.10908

This comparison suggests that more Flight Hours are beneficial to individual success while a larger mean is detrimental to group success. Instrument Comprehension is positively correlated to the success of the individual and group. N-Temperamental is negatively correlated to individual success but positively correlated to group success. N – Worry is negatively correlated for individual and group success. The difference in correlation strength between the individual and group suggests that a higher group average has a greater impact on the group than it does the individual.

An entire list of the category correlations to group and individual success is in Appendix B. The delta between the correlations is also listed. The largest deltas were for Flight Hours: 0.36899, PCSM Score: 0.17167, A - Hyper-Competitive: 0.16323, and Electrical Knowledge:0.14955. The fifth largest was Hidden Figures which was removed with the implementation of the Form T.

Initial Regression Exploration of the Standard Deviation of the Mean Scores

After the mean scores were analyzed, the standard deviation of the class means was analyzed. A correlation matrix was run to compare the components to the graduation rate. The correlation matrix results are in Table 12. The five components with the greatest correlation were Pilot Composite: -0.24093, Scale Read: -0.23845, Flight Hours: -0.18963, Instrument Comprehension: -0.18043, and Aviation Information: -0.16884. The negative correlation suggests the scores of the students within the class should be grouped closer as opposed to wider spread. An interesting note is the top four components from this list also appeared as the top four correlations for the averages and graduation rate.

Table 12: Correlation Matrix: Standard Deviation of Mean Scores to Graduation Percentage

Category	Correlation	Category	Correlation
Pilot Composite	-0.24093	Math Knowledge	-0.06778
Scale Read	-0.23845	Block Counting	-0.06409
Flight Hours	-0.18963	E - Sociable	0.06325
Instrument Comprehension	-0.18043	Neuroticism	0.06129
Aviation Information	-0.16884	Academic Composite	-0.06042
CSO Composite	-0.16337	N - Stress Under Pressure	0.06025
PCSM Score	-0.14750	M - Self Serving	0.05973
Machiavellianism	0.12485	Data Interpretation	0.05424
C - Achievement Striving	0.12073	Word Knowledge	-0.04977
N - Temperamental	0.12001	Arithmetic Reasoning	-0.04870
Table Reading	-0.11236	Quantitative Composite	-0.04724
Conscientiousness	0.09570	A - Helpful Altruistic	0.04213
E - Reserved	0.09352	O - Scientific Interest	0.03818
A - Team Player	0.09213	GPA	-0.02580
Hidden Figures	-0.09093	Verbal Composite	-0.02521
A - Hyper-Competitive	0.08939	A - Considerate	0.02391
Agreeableness	0.08706	C - Order	0.01515
Extraversion	0.08459	O - Reflective	-0.01114
M - Envious	0.07939	E - Dominance	0.01005
Openness	0.07850	N - Worry	-0.00947
O - Cultured	0.07839	Verbal Analogies	0.00569
Class Size	-0.07253	Rotated Blocks	-0.00560
A - Pleasant	0.07169	M - Individualistic	0.00414
O - Creative	0.06914	Electrical Knowledge	0.00235

A linear regression was run on the top ten correlated components to determine if they would produce a good model. A model appeared after removing the high VIF scores and highest *p*-values. The model contained four variables: Flight Hours, CSO Composite, Instrument Comprehension, and Openness. Figure 24 shows the model analysis with validation. The RASE delta was 0.00345, or 0.345% average variation.

Summary of Fit

RSquare	0.050641
RSquare Adj	0.015479
Root Mean Square Error	0.058932
Mean of Response	0.923096
Observations (or Sum Wgts)	113

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	0.02000789	0.005002	1.4402
Error	108	0.37508533	0.003473	Prob > F
C. Total	112	0.39509322		0.2257

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.9882743	0.057626	17.15	<.0001*
Flight Hours	-2.573e-5	0.000022	-1.16	0.2466
CSO Composite	-0.002089	0.001898	-1.10	0.2734
Instrument Comprehension	-0.002932	0.001793	-1.64	0.1049
Openness	0.0016414	0.002619	0.63	0.5322

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.0506	0.05761	113
Validation Set	0.1146	0.06045	37
Test Set	0.1560	0.06106	38

Figure 24: Standard Deviation Model JMP Analysis with Validation

When the model was applied to the entire data set, Flight Hours, CSO Composite, and Instrument Comprehension were all significant below a *p*-value of 0.05 and Openness was slightly less significant at 0.0949. Figure 25 shows the model's Summary of Fit, Analysis of Variance, Parameter Estimates and Profiler. The F(Model)=5.3947 > $F_{[.05]}$=2.4205 (using the same *k* and *n* from the FINN Model) and the model's *p*-value is 0.0004.

Summary of Fit

RSquare	0.10548
RSquare Adj	0.085928
Root Mean Square Error	0.059171
Mean of Response	0.921927
Observations (or Sum Wgts)	188

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	0.07555313	0.018888	5.3947
Error	183	0.64072594	0.003501	Prob > F
C. Total	187	0.71627907		0.0004*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	1.0012066	0.043708	22.91	<.0001*	.
Flight Hours	-4.981e-5	0.000018	-2.77	0.0062*	1.0015094
CSO Composite	-0.003003	0.001469	-2.04	0.0423*	1.007492
Instrument Comprehension	-0.0038	0.001369	-2.78	0.0061*	1.0520612
Openness	0.0033142	0.001974	1.68	0.0949	1.0439279

Prediction Profiler

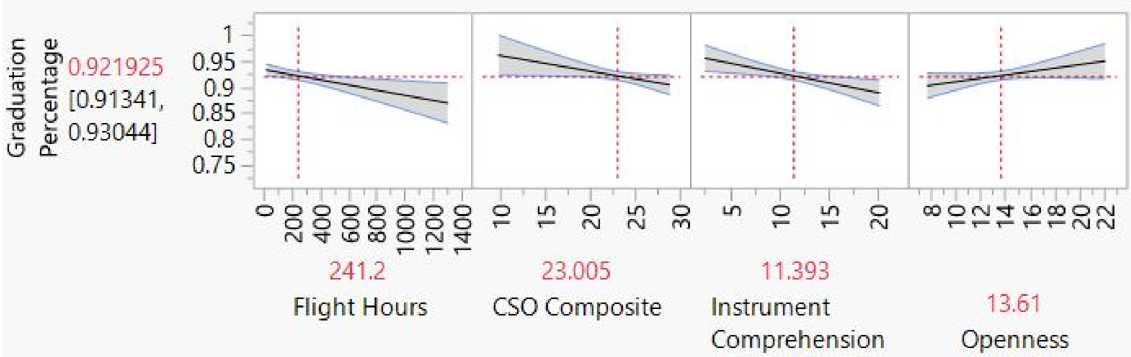


Figure 25: Standard Deviation Model JMP Analysis without Validation

Figure 26 shows the actual versus predicted graduation percentage. Figure 27 shows the distribution of the residuals.

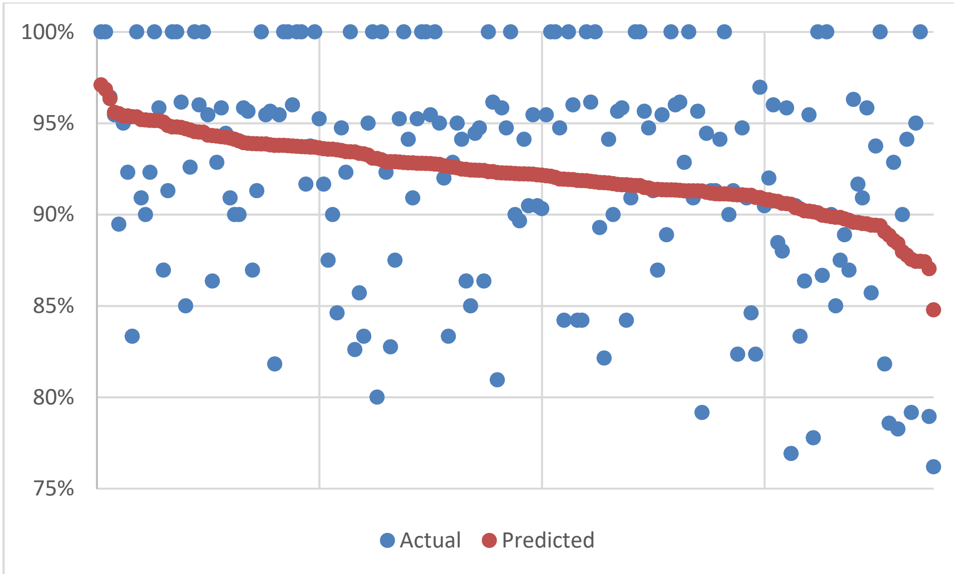
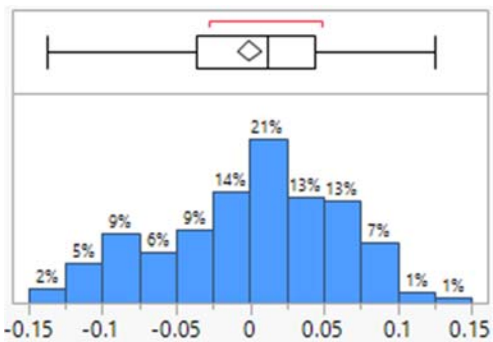


Figure 26: Standard Deviation Model Actual vs Predicted Graduation Percentage



Quantiles

100.0%	maximum	0.125654467
99.5%		0.125654467
97.5%		0.0915593669
90.0%		0.0721198901
75.0%	quartile	0.0434039158
50.0%	median	0.0115238179
25.0%	quartile	-0.036723706
10.0%		-0.086055675
2.5%		-0.121824
0.5%		-0.137342307
0.0%	minimum	-0.137342307

Summary Statistics

Mean	-1.06e-16
Std Dev	0.058535
Std Err Mean	0.0042691
Upper 95% Mean	0.0084218
Lower 95% Mean	-0.008422
N	188

Figure 27: Standard Deviation Model Residuals Distribution

V. Conclusions and Recommendations

Conclusions of Research

This research studied USAF pilot class average test scores and how they relate to pilot training group performance. The research showed there is a modest linear relationship between the group average scores and graduation rate.

While the highest correlations were found between the graduation percentage and Instrument Comprehension, Pilot Composite, Scale Read, Flight Hours, and E - Sociable, only two of those components were represented in any of the models, Flight Hours and Instrument Comprehension. Those two, along with N - Temperamental, and N - Worry held the most significance and were analyzed further.

Recommendations for Action

These findings are able to be implemented when placing students in classes. According to the FINN Model, small changes in the average class scores could mean notable changes in the graduation rate. For example, for a one-point increase in the Instrument Comprehension average, the graduation percentage would increase 0.5%. With a one-point increase in N-Temperamental and N-Worry, the graduation rate would increase 0.4% and decrease 0.4%, respectively. More generally speaking, classes should be stacked to create higher averages for Instrument Comprehension and N - Temperamental, and lower averages for Flight Hours and N - Worry.

The correlations with Flight Hours presents an interesting question. Why is more Flight Hours beneficial to an individual but detrimental to the group? Flight hours outside of UPT develops experience and is positively correlated to better performance. The comparison of exceedingly high performing individuals may make the performance of low performing individuals seem weaker. Two options are available to managers to address this problem. First, the classes could be grouped by flight hours. This would reduce the disparities in performance caused by these differences in experience. Couple this with the results of the standard deviation regression, and the case for tightening class grouping around flight hours is bolstered. Second, individuals with high flight hours could be accelerated through the program by being given the option to take check rides early. This would also have the added benefit of increasing throughput and thereby reducing the pilot shortage.

Recommendations for Future Research

The created models are able to account for a small portion of the model variability so more research is warranted. A better understanding of the relationship amongst variables would be helpful. A machine learning type neural network could be used in understanding the outcomes of how components that are high or low interact with each other to create a more accurate assumption of the data.

This research viewed all attritions as the same, a reduction in the graduation percentage. But these students were not all the same and their attrition may not have been a sign of their ability. Therefore, it would also be beneficial to understand the nature of each students' attrition. There are sometimes cases where individuals DOR because of reasons outside of their control, or are medically disqualified although, by all the indicators, they are a perfect candidate.

There was insufficient data to include the newer Form T raw scores and facets in this analysis. The modification of the AFOQT was driven by a desire to better predict officer as well as pilot aptitude, so understanding how these new sections affect the graduation rate could be an indicator of their success or failure. Future research could include a newer data set with the new Form T data points.

Summary

This research showed a linear relationship between group averages and standard deviations and the graduation percentage. Classes should be filled with students whose scores yield a higher average Instrument Comprehension and N-Temperamental score and a lower average N-Worry. Based on average score findings, Flight Hour averages should be lower, but the standard deviation analysis suggest grouping students with similar individual flight hour amounts would be beneficial. More research is required to better understand the differences in attritions, the Form T facets, and how interaction between components affects the graduation percentage.

Appendix A. Facet Definitions

Facet	Definition
A - Team Player	The tendency to work well with others to reach a common goal.
A – Pleasant	Have an agreeable manner and appearance to others, pleasing to be around.
A – Considerate	The tendency to treat others with kindness and consideration.
A – Helpful-Altruistic	The level of active concern for the welfare of others.
A – Hyper-Competitive	Being very competitive in nature without concern for others.
N- Stress Under Pressure	Level of susceptibility to stress, especially in pressure situations.
N – Temperamental	The level to which one is easily upset emotionally and erratic in behavior.
N- Worry	The level of anxious concern for things, especially those that have not yet happened.
E – Reserved	Lacking Social ascendancy and forcefulness of expression
E – Sociable	Enjoying or requiring the company of others, fondness of companionship.
E – Dominance	Having social ascendancy and forcefulness of expression.
C – Achievement-Striving	The extent to which one has need for personal achievement and sense of direction for goal attainment.
C – Order	The level of preference for order, arraignment, and tidiness in life.
O – Creative	Extent of intellectual curiosity and innovative thinking.
O – Reflective	The level of receptivity to one's own inner feelings, emotions, and thoughts.
O – Scientific Interest	The extent to which one is interested in science and theory.
O – Cultured	Level of appreciation for art and beauty.
M – Envious	Resentment towards others due to their success or achievements.
M – Individualistic	Level of preference for working alone and doing things one's own way.
M -Self-Serving	Tendency to serve one's own selfish interests, especially at the expense of others.

(Manley, 2011)

Appendix B. Group and Individual Success Correlation and Correlation Deltas

Category	Group Correlation	TOTMASS Correlation	Delta
Flight Hours	-0.15522	0.21377	0.36899
PCSM Score	0.09122	0.26289	0.17167
A - Hyper-Competitive	0.11800	-0.04523	0.16323
Electrical Knowledge	-0.11673	0.03282	0.14955
Hidden Figures	-0.05738	0.08227	0.13965
E - Sociable	-0.15174	-0.02508	0.12666
Aviation Information	0.07867	0.20325	0.12458
Table Reading	0.03748	0.15840	0.12092
Extraversion	0.12156	0.00299	0.11857
N - Temperamental	0.06828	-0.04750	0.11577
E - Dominance	-0.08513	0.02767	0.11279
A - Helpful Altruistic	-0.10896	-0.00269	0.10627
Pilot Composite	0.17966	0.28370	0.10404
Verbal Analogies	-0.03399	0.06701	0.10100
Scale Read	0.16240	0.26249	0.10009
C - Achievement Striving	-0.05674	0.03539	0.09213
CSO Composite	0.07676	0.16591	0.08915
N - Stress Under Pressure	0.02245	-0.06032	0.08277
N - Worry	-0.10908	-0.02645	0.08263
Data Interpretation	0.02246	0.10430	0.08185
Conscientiousness	-0.05232	0.02949	0.08181
Neuroticism	0.02019	-0.05404	0.07423
Agreeableness	-0.06064	0.01239	0.07303
Class Size	-0.07253	-0.00002	0.07252
GPA	-0.00115	0.07135	0.07251
E - Reserved	0.04066	-0.03086	0.07152
Verbal Composite	0.02814	0.09877	0.07063
A - Pleasant	-0.05372	0.01604	0.06976
M - Individualistic	-0.05955	0.00904	0.06859
Academic Composite	0.08016	0.14399	0.06383
A - Team Player	-0.02712	0.03439	0.06151
O - Creative	-0.04718	0.01210	0.05928
A - Considerate	-0.06818	-0.01560	0.05258
Arithmetic Reasoning	0.10387	0.15552	0.05165
M - Self Serving	0.01801	-0.02545	0.04346
Word Knowledge	0.01297	0.05640	0.04343
Quantitative Composite	0.12017	0.15645	0.03629

Block Counting	0.12774	0.09159	0.03615
Openness	-0.06183	-0.03255	0.02928
O - Reflective	-0.05518	-0.02979	0.02539
Rotated Blocks	0.14814	0.12489	0.02325
M - Envious	0.03583	0.01320	0.02264
Math Knowledge	0.10830	0.13039	0.02210
O - Cultured	-0.07970	-0.09660	0.01690
Instrument Comprehension	0.20830	0.21635	0.00805
C - Order	0.01082	0.01420	0.00339
Machiavellianism	0.00261	0.00569	0.00308
O - Scientific Interest	0.00495	0.00346	0.00149

Bibliography

Arth, Thomas O., Kurt W. Steuck, Christopher T. Sorrentino, and Eugene F. Burke. 1990. *Air Force Officer Qualifying Test (AFOQT): Predictors of Undergraduate Pilot Training and Undergraduate Navigator Training Success: Interim Report, April 1987-August 1989.* Contract AFHRL-TP-89-52. Brooks Air Force Base TX: Air Force Human Resources Laboratory, May 1990.

Axe, David. 2018. "What's Driving the U.S. Air Force Pilot Shortage?" Foreign Policy. 2018. https://foreignpolicy.com/2018/05/04/whats-driving-the-u-s-air-force-pilot-shortage/.

Barron, Laura G., Thomas R. Carretta, and Mark R. Rose. 2016. "Aptitude and Trait Predictors of Manned and Unmanned Aircraft Pilot Job Performance." *Military Psychology* 28 (2): 65–77.

Carretta, Thomas R. 2008. *Predictive Validity of the Air Force Officer Qualifying Test for USAF Air Battle Manager Training Performance: Interim Report, June 2008-August 2008.* Contract AFRL-RH-WP-TR-2009–0007. Wright-Patterson AFB OH: Air Force Research Laboratory, September 2008.

———. 2011. "Pilot Candidate Selection Method: Still an Effective Predictor of US Air Force Pilot Training Performance." *Aviation Psychology and Applied Human Factors* 1 (1): 3–8.

———. 2013. "Predictive Validity of Pilot Selection Instruments for Remotely Piloted Aircraft Training Outcome." *Aviation Space and Environmental Medicine* 84 (1): 47–53.

Carretta, Thomas R., Raymond E. King, Malcolm J. Ree, Mark S. Teachout, and Erica L. Barto. 2016. "Compilation of Cognitive and Personality Norms for Military Aviators." *Aerospace Medicine and Human Performance* 87 (9): 764–71.

Carretta, Thomas R., and Malcolm J. Ree. 1995. "Air Force Officer Qualifying Test Validity for Predicting Pilot Training Performance." *Journal of Business and Psychology* 9 (4): 379–88.

———. *Pilot Selection Methods*: *Interim Report, January 1999-July 2000*. Contract AFRL-HE-WP-TR-2000-0116. Wright-Patterson AFB OH: Air Force Research Laboratory, September 2000.

Carretta, Thomas R., Mark S. Teachout, Malcolm J. Ree, Erica L. Barto, Raymond E. King, and Charles F. Michaels. 2014. "Consistency of the Relations of Cognitive Ability and Personality Traits to Pilot Training Performance." *International Journal of Aviation Psychology* 24 (4): 247–64.

Finegold, Lawrence S., and Deborah Rogers. 1984. *Relationship Between Air Force Officer Qualifying Test Scores and Success in Air Weapons Controller Training: Interim Report, November 1982-Feburuary 1984*. Contract AFHRL-TR-85-13. Brooks Air Force Base TX: Adkins Research Center, June 1985 (AD-A158162).

King, Raymond E., Thomas R. Carretta, Paul Retzlaff, Erica L. Barto, Malcolm J. Ree, and Mark S. Teachout. 2013. "Standard Cognitive Psychological Tests Predict Military Pilot Training Outcomes." *Aviation Psychology and Applied Human Factors* 3 (1): 28–38.

Losey, Stephen. 2018. "Air Force Hopes to Train 1,500 New Pilots Each Year by 2022 to Help Solve Shortage." Air Force Times. 2018. https://www.airforcetimes.com/news/your-air-force/2018/10/10/air-force-hopes-to-train-1500-new-pilots-each-year-by-2022-to-help-solve-shortage/.

Manley, Gregory G. *Development of Domain and Facet Level Scales for the Self-Description Inventory (Redacted Version): Final Report, July 2005-November 2011.*Contract AFCAPS-TR-2011-0007-R. Randolph Air Force Base TX: Air Force Personnel Center, November 2011.

Manley, Gregory G., and Johnny J. Weissmuller. 2017. "Development of the USAF Self-Description Inventory." In *32nd Annual Meetings of the Society for Industrial and Organizational Psychology*, 20.

Olea, Michele M., and Malcolm J. Ree. *Predicting Aircrew Training Performance with Psychometric g: Interim Report, January 1992-June 1992*. Contract AL-TP-1993-0011. Brooks Air Force Base TX: Armstrong Laboratory, April 1993.

Panzino, Charlsy. 2018. "So You Wanna Fly? Air Force Pilots Can Apply to New Aviation-Only Program." Air Force Times. 2018. https://www.airforcetimes.com/news/your-air-force/2018/07/20/so-you-wanna-fly-air-force-pilots-can-apply-to-new-aviation-only-program/.

Pawlyk, Oriana. 2018. "Here's How the Air Force Hopes to Train 1,500 New Pilots a Year." Military.Com. 2018. https://www.military.com/daily-news/2018/10/20/heres-how-air-force-hopes-train-1500-new-pilots-year.html.

———. 2019. "Air Force's Pilot Training Experiment Still Evolving as New Class Begins." Military.Com. 2019. https://www.military.com/daily-news/2019/12/26/air-forces-pilot-training-experiment-still-evolving-new-class-begins.html.

Rose, Mark R., Laura G. Barron, Thomas R. Carretta, Richard D. Arnold, and William R. Howse. 2014. "Early Identification of Unmanned Aircraft Pilots Using Measures of Personality and Aptitude." *International Journal of Aviation Psychology* 24 (1): 36–52.

Schulker, David, Douglas Yeung, Kirsten Keller, Leslie Payne, Lisa Saum-Manning, Kimberly Curry Hall, and Stefan Zavislan. 2018. *Understanding Demographic Differences in Undergraduate Pilot Training Attrition*. *Understanding Demographic Differences in Undergraduate Pilot Training Attrition*. RAND Corporation.

Secretary of the Air Force Public Affairs. 2017. “Air Force Announces Initiatives to Lessen Pilot Shortage.” 2017. https://www.af.mil/News/Article-Display/Article/1290397/air-force-announces-initiatives-to-lessen-pilot-shortage/.

Teachout, Mark S., Malcolm J. Ree, Erica L. Barto, Thomas R. Carretta, Raymond E. King, and Charles F. Michaels. 2013. “Consistency of Pilot Trainee Cognitive Ability , Personality, and Training Performance in Undergraduate Pilot Training.” *United States Air Force Research Laboratory*, 56.

United States Government Accountability Office. 2018. “GAO-18-113: MILITARY PERSONNEL: DOD Needs to Reevaluate Fighter Pilot Workforce Requirements,” no. April: 69.

Weissmuller, Johnny J., and Kenneth L. Schwartz. *Self-Description Inventory: Assault on Occam’s Razor: Final Report.* Contract AFCAPS-TR-2012-0003. Randolph Air Force Base TX: Air Force Personnel Center, October 2007.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
03/26/2020

2. REPORT TYPE
Master's Thesis

3. DATES COVERED *(From - To)*
September 2018 - March 2020

4. TITLE AND SUBTITLE
Undergraduate Pilot Training Attrition: An Analysis of Individual and Class Composition Component Factors

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Akers, Christina M., Capt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
Wright-Patterson AFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENS-MS-20-M-126

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Intentionally Left Blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES
This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT
The Air Force is working hard to reduce the shortage of nearly 2,000 pilots that threatens the Air Force's core mission. Officials have focused on increasing retention and training throughput. The purpose of this research is to explore attrition reduction by understanding how class composition of individual abilities and personalities affects the class graduation rate. Using AFOQT scores, SDI+ scores, PCSM scores, flight hours, and college GPAs, correlations were studied and a simple linear regression was run with the variables to determine relationships. This study resulted in the creation of models to help decision makers plan classes to optimize success rates.

15. SUBJECT TERMS
Undergraduate Pilot Training (UPT), Attrition, T-6, AFOQT, SDI+

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	59	Dr. Adam D. Reiman (937) 255-3636 x4319 Adam.Reiman@afit.edu

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18