FAA ACADEMY

International Journal of Applied Aviation Studies

A publication of
FAA Academy
Oklahoma City, Oklahoma

REVIEW PROCESS

The Federal Aviation Administration Academy provides traceability and oversight for each step of the *International Journal of Applied Aviation Studies (IJAAS)*. *IJAAS* is a peer-reviewed publication, enlisting the support of an international panel of consulting editors. Each consulting editor was chosen for his or her expertise in one or more areas of interest in aviation. Using the blind-review process, three or more consulting editors are selected to appraise each article, judging whether or not it meets the requirements of this publication. In addition to an overall appraisal, a Likert scale is used to measure attitudes regarding individual segments of each article. Articles that are accepted are those that were approved by a majority of judges. Articles that do not meet *IJAAS* requirements for publication are released back to their author or authors.

IJAAS is printed by the Government Printing Office. Individuals wishing to obtain a copy of the journal may contact Beverly Laughead by email at bsl@ou.edu, by telephone at (405) 954-1191, or by writing to the following address:

International Journal of Applied Aviation Studies
c/o Beverly Laughead
AMA-502-OU
PO Box 25082
Oklahoma City, OK 73125

International Journal of Applied Aviation Studies
Fall 2003, Volume 3, Number 2

1st Printing December 2003

POLICY AND DISCLAIMERS

Policy Statement: The Federal Aviation Administration (FAA) Academy strongly supports academic freedom and a researcher's right to publish; therefore, the Federal Aviation Administration Academy as an institution does not endorse the viewpoint or guarantee the technical correctness of any of the articles in this journal.

Disclaimer of Liability: With respect to articles available in this journal, neither the United States Government nor the Federal Aviation Administration Academy nor any of their employees, makes any warranty, express or implied, including the warranties of merchantability and fitness for a particular purpose, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

Disclaimer of Endorsement: Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the United States Government or the Federal Aviation Administration Academy.The views and opinions of authors expressed herein do not state or reflect those of the United States Government or the Federal Aviation Administration, and shall not be used for advertising or product endorsement purposes.

PHILOSOPHY STATEMENT

Cornelius Lanczos, a mathematician working in the field of applied analysis, expressed the history of mathematics in three phases:

1) A given physical situation is translated into the realm of numbers,
2) By purely formal operations with these numbers certain mathematical results are obtained, [and]
3) These results are translated back into the world of physical reality (1988, p. 1). [1]

Formal papers, in subjects related to aviation, roughly follow the same course. However, there appears to be a weakness in aviation research, that being the omission of the third phase.

It is not good enough that conclusions are drawn, if those conclusions fail to improve the system observed. Clearly, the observed have a say in implementing the conclusions of research, but their failure to implement the conclusions drawn by the researcher may be more indicative of a lack of understanding than a lack of desire. Researchers tend to peer into complex systems as through a soda straw, forming formal opinions on the finite without understanding the complete system. Industry, ever mindful of the complete system, may find research irrelevant, because it makes much to do about nothing.

The editorial staff, to include those listed as consulting editors, is committed to the improvement of all individuals within the aviation community. We seek to enhance existing systems bearing in mind that small improvements must not upset the delicate balance between too little and too much help. We also seek to promote safety, not by lip service, but by demonstration in how we execute our studies and how we report our findings.

We feel that the best way to translate results back to the physical world is to incorporate the viewpoints of people around the globe. Without the influence of a worldwide community, we deny the significance of diversity, and ignore the perspectives of gifted scientists from different countries. It is our hope that each reader will feel the same.

B.S.L.

[1] Lanczos, C. (1988). Applied Analysis. Mineola, NY: Dover Publications, Inc.

EDITOR'S NOTES

Formal Papers

Because of the increase in human consumption of caffeine-containing beverages, such as coffee, tea, cola drinks, energy drinks, and chocolate, to cope with performance problems due to fatigue Deixelberger-Fritz, Tischler, and Kallus conducted a study to evaluate the effects of a single standard dose of an Energy Drink on performance of pilots in a fatigue-inducing paradigm. The results cannot be attributed to a caffeine withdrawal effect because the subjects were only nonsmokers and low to moderate caffeine users and the subjects were not deprived from caffeine. Readers who are looking for positive options to cope with fatigue will find the results of this study to be quite interesting.

Many schools and colleges are incorporating electronic methods into their instructional delivery. Although there have not been many studies to determine whether instruction delivered in this way is equal to or better than the more traditional "classroom" method, those interested in distance learning will find this study by Howell, Denning, and Fitzpatrick showed that the academic performance of today's students is not impacted adversely by electronic instruction.

With the ever increasing expense associated with latest generation training aircraft, flight training devices, and materials, Young and Fanjoy recommend a tailored approach to advanced collegiate flight training that will take advantage of cost effective program components and generate the potential for interaction with aviation industry partners. They address issues associated with advanced flight training resources as well as suggested levels of implementation that address the particular needs and financial constraints of each collegiate flight training unit. Although they state that the levels of advanced flight training resources they present are not intended to be prescriptive, Young and Fanjoy propose that they do offer a starting point for consideration by members of a training unit's flight faculty.

The thought that the recent identification of a new spatial ability (dynamic spatial ability) is particularly important in occupations such as air traffic control and piloting, D'Oliveira conducted a study to analyze the potential predictive value of static and dynamic spatial ability and the ability to coordinate information when predicting training results of air traffic controllers and pilots. Her study dealt with whether both dynamic spatial ability and the ability to coordinate information are significantly better predictors of training results than static spatial ability, therefore contributing to potential improvements in current selection procedures adopted for these occupations.

Using one questionnaire—called the Organizational Safety Culture Questionnaire (OSCQ)—Patankar conducted an exploratory study to assess the safety attitudes and opinions among flight operations personnel, maintenance personnel, and other employees at one partner organization to determine the factors that may contribute to accident-free safety record of a specific organization. His study was based on a previous survey of flight operations personnel, maintenance personnel, and other employees to measure individual attitudes regarding safety and their opinions regarding the factors that may contribute toward their accident-free safety record. In the case of this particular partner company, the success factors were identified as follows: emphasis on compliance with the Standard Operating Procedures, collective commitment to safety, individual sense of responsibility toward safety, and the high level of employee-management trust.

Those interested in reducing errors made in aircraft cockpits and those interested in flight crew training/education might find the results of the Bliss and Fallon study, which broadens the applicability of Normative Decision Theory by demonstrating that it can be successfully applied to a specialized teamwork environment, to be of interest. Bliss and Fallon examined the problem of faulty information transfer among flight crew members by manipulating leadership style and studying its effects on performance and satisfaction under conditions of high and low workload. In addition to their findings, Bliss and Fallon discuss the limitations of their study and recommend future research to continue to clarify the conditions where the Normative Decision Theory applies and does not apply.

The FAA pilot knowledge test is a multiple-choice assessment tool designed to measure the extent to which applicants for FAA pilot certificates and ratings have mastered a corpus of required aeronautical knowledge. All questions that appear on the test are drawn from a database of questions that is made available to the public. Casnar, Jones, Puentes, and Irani conducted a study to investigate the FAA's concerns that releasing test questions in advance may: (1) negatively affect the way students learn and ultimately understand required aeronautical knowledge; and (2) reduce the validity of the knowledge test as an assessment tool. In addition to the results of their study, Casnar, Jones, Puentes, and Irani discuss the limitations of their study, as well as the changes made by the FAA to the Knowledge Test in July 2003.

In addition to those who treat individuals with a fear of flying, those with that fear may find the article by Van Gerwen, Van de Wal, Spinhoven, Diekstra, and Van Dyck to be of interest. Based on Bandura's self-efficacy theory that judgments and expectations concerning performance capabilities are relevant for the initiation, persistence, and modification of anxiety problems and specific fears, the authors conducted a study on the differential effects on self-efficacy expectancies of various treatment components in a fear of flying protocol. They provide detailed statistical analyses of the study, discuss the study's shortcomings, and recommend future studies.

Training Development Reports, Studies, and Papers

Readers who are involved in aviation education/training may find the Elliott article on biometric technologies valuable. As a result of the work to improve and standardize airline security, biometrics is being considered for a number of transportation-related applications. Elliott provides an introduction to biometric technologies, outlining the general concepts and definitions that students in aviation technology will come across, the specific classifications, and the specific issues within an airport environment that may affect the performance of individual biometric technologies. He discusses biometric technologies that have been successfully deployed within an aviation setting and suggests that those who develop aviation curricula may want to include biometrics as part of a course in aviation security or airport management.

Utilizing the findings of a 1998 study of the Characteristics of Successful Aviation Leaders of Oklahoma, Kutz presents a review of literature to explore the power of passion in leading aviation professionals—from introducing a description of the term "passion," as it relates to aviation, to discussing the problems with passion in aviation leadership—and suggests that this subject deserves more extensive research.

Book Reviews

We had an exciting experience for this issue of IJAAS – we were asked to provide reviews of several publications. Because of the tremendous success we have had with the increase in manuscript submissions, we were forced to limit the reviews to three.

Todd Hubbard provides a positive review of Psychological Perspectives on Fear of Flying, edited by Robert Bor and Lucas van Gerwen, which, by the way, ties in well with the article about the study conducted by Van Gerwen, Van de Wal, Spinhoven, Diekstra, and Van Dyck, also in this issue. The text provides clinical studies, therapy critiques, technological aides, suggestions for air carrier personnel, and therapy success stories. Hubbard highly recommends the book not only for those suffering from fear of flying, but also for the classroom. He also suggests that readers investigate other works by the authors.

Our next book review (Passenger Behaviour, edited by Robert Bor) is provided by Raymond King. Adding personal quips, King begins by suggesting that most readers will not read this book from cover to cover the way in which he did to prepare for his review review, but will use it as a reference tool. Furnishing a picture of each chapter, he recommends the book, especially to those who fly on a regular basis.

It seems only appropriate in this 100th year of aviation, that Mark Sherman and Deak Arch reviewed a book about aviation history, The U.S. Air Service in the Great War, 1917-1919, by James J. Cooke. Describing the book as one that explores World War One Aviation, not from a pilot or unit perspective, but from a military command and control perspective, Sherman and Arch suggest that those with an interest in American Air Power organization during World War I, aeronautical logistics, or military command and control functions should read this book.

B.S.L.

International Journal of Aviation Research and Development

Volume 3, Number 2 2003

Book Reviews

International Journal of Applied Aviation Studies, 3, 2

Formal Papers

Changes in Performance, Mood State and Workload Due to Energy Drinks in Pilots

D. Deixelberger-Fritz,
M.A. Tischler, and K. Wolfgang Kallus
Department of Psychology
Karl-Franzens-University Graz
Universitaetsplatz 2/III
8010 Graz, Austria
Email: wolfgang.kallus@uni-graz.at

Abstract

Twenty-four pilots and eight non-pilots participated in a randomized double-blind crossover study on performance changes in a fatigue-inducing 6-hour laboratory paradigm in the early evening hours on two different days without caffeine deprivation. After two hours of work on mental performance tests, either 250ml of a placebo (same Energy Drink, without Caffeine, Taurine and Glucuronolactone) or 250ml of an Energy Drink (containing Caffeine 80mg, Taurine 1000mg, and Glucuronolactone 600mg) were administered. Changes in performance were assessed during the following 4 hours. The kind of drink was switched on the second experimental day. Statistical hypotheses based on a preliminary study were tested via analysis of variance. Clear-cut positive effects of the energy drink could be demonstrated on choice reaction time and on performance in a concentration test. Effects persisted for more than two hours after administration. This long-lasting positive effect on performance can hardly be explained by the small dose of caffeine. In addition, pilots performed markedly better on the second day, despite the fact that tests were practiced extensively during baseline conditions. Effects on mood state and heart rate variability were less pronounced.

Performance measures were controlled for errors to make sure that the effects were not due to a mere change in the speed/accu-

Requests for reprints should be sent to Beverly Laughead, FAA Academy, AMA-502-OU, P.O. Box 25082, Oklahoma City, OK 73125.

racy trade-off towards more speed.

The study showed that breaks and nutrition in breaks can easily change the performance potential even in highly selected people like pilots. This offers positive options to cope with fatigue.

Introduction

Caffeine is the most widely used psychoactive substance in the world (Gilbert, 1976). Most of the caffeine consumed comes from dietary sources such as coffee, tea, cola drinks, energy drinks, and chocolate. The early studies used large doses of caffeine and it has been suggested that doses in excess of 500mg are not beneficial (Hasenfratz & Baettig, 1994). Consequently, smaller doses that are closer to those in commonly used caffeinated beverages have been used in more recent studies (Warburton, 1995; Durlach, 1998; Hindmarch, Quinlane, Moore, and Porkin, 1998; Reyner & Horne, 2000; Warburton, 2001; Alford, Cox and Westcott, 2001). Caffeine improves mental performance (Hindmarch et al., 1998) e.g.: concentration (Alford et al., 2001), attention (Warburton, 2001), and memory (Alford et al., 2001) as well as reaction time (Lieberman, Wurtman, Emde, Roberts, and Coviella, 1987; Kerr, Sherwood, and Hindmarch, 1991; Durlach, 1998; Kenemans & Verbaten, 1998; Alford et al., 2001).

A rising number of humans use caffeine-containing beverages to cope with performance problems due to fatigue. Experimental laboratory studies in fatigue paradigms are necessary, if one wants to prove the benefits of energy drinks to compensate fatigue-induced performance impairments. Rules of good clinical practice (GCP, 1997) from pharmacopsychological drug studies (randomized double-blind placebo controlled) should be adopted as far as possible. In addition, a multivariate approach should be used to obtain a picture of the drug effects in different functional areas (attention, concentration, and mood).

Positive results of a small amount of coffee (e.g. two cups, about 150mg caffeine) on alertness in fatigue-inducing laboratory studies have been reported (Lorist, Snel, Kok & Mulder, 1994; Horne & Reyner, 1996; Reyner & Horne, 1997).

The present study was conducted to evaluate the effects of a single standard dose of an Energy Drink on performance of pilots in a fatigue-inducing paradigm. The study followed the rules of GCP.The problem of caffeine deprivation (James, 1997) is not relevant in this study as subjects with low habitual caffeine consumption were studied in the evening (telling subjects to abstain from caffeine for 24 hours, James, 1997).

Method

Participants

Twenty-four pilots and eight non-pilots participated in a randomized double-blind crossover study on performance changes in two tiring six-hour laboratory sessions.

All participants were nonsmokers and low habitual caffeine users (3-6 cups coffee per day). Subjects with a higher average caffeine intake, smokers, and those who took drugs regularly were excluded. On investigation days subjects were not permitted to drink alcohol. Informed written consent was obtained from all subjects.

All pilots had a valid flight license (13 commercial pilots, 4 military jet pilots, 7 VFR-Pilots =visual flight licenses for private pilots). Ranging in age from 23 to 40, the mean age was 29.75 years (std=5.08 years).

Independent variables

In this study, 250ml of an energy drink (Red Bull® Energy Drink, containing carbonated water, 80mg caffeine, 1000mg taurine, 600mg glucuronolacton, sucrose, glucose, citric acid, carbon acid, inositol, vitamins (niacin, panthenol, B6, B12), flavours, colour) was tested. The placebo contained glucose and vitamins. The equivalence of taste for verum and placebo (at 2°C) was ensured before the experiment. In a preliminary study, a tipple comparison taste-tasting, using two temperature levels (2°C N=15, 5-10°C N=8) to ensure equivalence at taste, was conducted with 23 subjects.

Design

A repeated measure, double-blind crossover design was used with each subject receiving both treatments in a randomized order.The two treatment conditions were 250ml Red Bull® Energy Drink and the same amount of a glucose-containing placebo.

Procedure

Both test sessions started at 1700 h and ended at 2300 h. The participants completed a similar six-hour work block on both test days. The investigation took place in a shielded psycho physiological, air-conditioned laboratory.A wash out period of at least 24 hours was kept to avoid potential residual effects (cf. Deixelberger, 2000).

Psychological Testing

The following tests were administered by means of a computer or in paper and pencil form.

Performance Measures for Hypothesis Testing. Three performance measures were selected for hypothesis testing based on a preliminary study (Deixelberger, 2000). The following measures were used:

1. sustained match to sample test "Cognitrone" (Schuhfried, 1994): Score: MC = Mean reaction time of correct responses

The computer-based sustained match to sample task assessed sustained attention (Schuhfried, 1994). A series of four pictures was presented in the first line on the screen and one picture below.The subjects determine whether the picture in the second line is exactly the same as one of those in the first line. For analysis, mean reaction time of correct responses is taken. The split-half-reliability is .98.

2. continuous performance task (arithmetic) CPT (Dueker & Lienert, 1959): Score: C-E (Correct – Error)

The CPT (Dueker and Lienert, 1959) obtains information about quantity and quality of arithmetic performance and short term memory.The subjects add two sets of figures and decide on a further subtraction or addition based on the relation of the two sums. The highly reliable test (r_t=.92) has been used in series of psycho-pharmacological studies in the past 40 years.

3. d2 letter cancellation test (continuous attention performance, Brickenkamp,1994):Score = T-E (Total – Error)

The letter cancellation test d2 (Brickenkamp, 1994) is a performance test, measuring speed and accuracy by doing simple tasks like a vigilant task. The subjects mark every d with two lines (one above and one below, both lines above or both lines below) mixed with other letters. The results from this test can be interpreted in three ways 1. quantitative – how many letters are marked, 2. how many letters are marked within a certain time, 3. qualitative – total of performance minus errors. To prove the hypothesis, total performance minus errors in this study was selected for the data analysis. The re-test-reliability is .92.

Supplementary performance tests. A set of supplementary performance tests was used to simulate tiring, monotonous working conditions. The following tests were used:

- Stroop Color Word Test (German version by Baeumler, 1985).
- Choice reaction test (Vienna Testing System, Schuhfried, 1996).
- Pilot Spatial Test (Vienna Testing System, Schuhfried, 1996).
- W AIS – Wechsler digits backwards (Tewes, 1991).

Assessment of psycho physiological state

- Adjective Check List EWL (Janke & Debus, 1978) and a short version for repeated measurements (BSKE)
- Multidimensional Scale on Physical Symptoms (Erdmann & Janke, 1984)
- ECG during resting conditions and during task execution.
- EEG during resting conditions and during task execution (results not reported here).

Intervening variables

Additional information was obtained for:

- Initial state (RESTQ, Kallus, 1995; FAL Janke et al., 1976)
  Initial state with respect to stress and recovery in the past three days was assessed by psychometric questionnaire (RESTQ). Additional questions on activities and nutrition on the test day was assessed by the questionnaire (FAL).
- Questionnaire on drug consumption (Janke et al., 1988).
  The questionnaire assesses habitual drug consumption and the use of functional foods.
- Personality traits (Eysenck Pers. Inventory; Eggert, 1983);
  The Eysenck Personality Inventory assesses two basic personality traits (extraversion and emotional stability).
- Biographical data
  Biological data were used for description of the sample.

Testing procedures

The study was conducted in a shielded psycho-physiological laboratory in the Department of Psychology, Karl-Franzens-University Graz, Austria. Volunteers were paid for participation.

The study started at 5 p.m. and ended at 11 p.m. Fatigue and monotony are very common at this time and most of the people do not drink much caffeine. Thus, side effects due to abstaining from caffeine were very unlikely.

The next two test phases were 120 minutes each. Before starting with test-phase 1 on day 1 and day b, the Red Bull® Energy Drink or the placebo was administered.

Table 1

Testing procedure

baseline day a	17.00-19.30	baseline day b
application	19.30-19.40	application
test-phase 1	19.40-21.30	test-phase 1
tests-phase 2	21.35-23.00	test-phase 2

Data analysis

Analyses of variance were computed for each of the three basic performance measures using the average of each test block as a dependent variable. The interaction term treatment*time was used to test the cross-over effect following Lehmacher´s decomposition of possible effects in cross-over designs (Lehmacher,1986). Type I error was 0.05, which was adjusted to three multiple tests by the Bonferroni-Holm Procedure (Holm, 1979). SPSS-Manova was used to compute effects. The overall individual level (mean across all measurements on all days of each dependent variable) was subtracted to eliminate initial value effects (Kallus, 1991).

Supplementary descriptive data analysis is based on descriptive p-values from analysis of variance. The results were analyzed "descriptively" to characterize the drug effects more precisely. The model of descriptive data analysis (Abt, 1987) provides a framework for the combination of confirmatory and supplementary data analysis.

Results

Results for hypothesis testing are depicted in table 2. The energy drink showed clear-cut effects on performance at the .05 level. Non-significant interaction terms (treatment*time) were expected on the null hypotheses.

Table 2

*Results of Analysis of Variance (time*Treatment-interaction)*

Variable (ranking according to p-values)	p-values	adjusted type-I-errors
CPT	.000	.0167 *
MC (Cognitrone)	.020	.025 *
d2	.322	.050 n.s.

Continuous Performance Task CPT. A significant time*treatment effect (p=<.001) can be reported for the score total performance minus errors. Subjects performed markedly better in the Red Bull® Energy Drink condition compared to the placebo day.

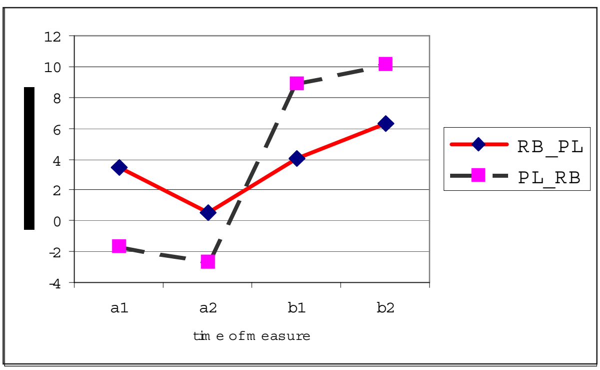

Figure 1. Results in CPT (total minus errors) for crossed treatment conditions (RB_PL means first day Red Bull®, second day placebo; PL_RB means first day placebo, second day Red Bull®) on two test days for both measurement occasions on each day.

Sustained match to sample test MC. A significant time*treatment effect (p=.020) for mean reaction time of correct responses in the sustained match to sample test was observed. The subjects had the shorter reaction times on the day with Energy Drink compared with the performance on the placebo day.

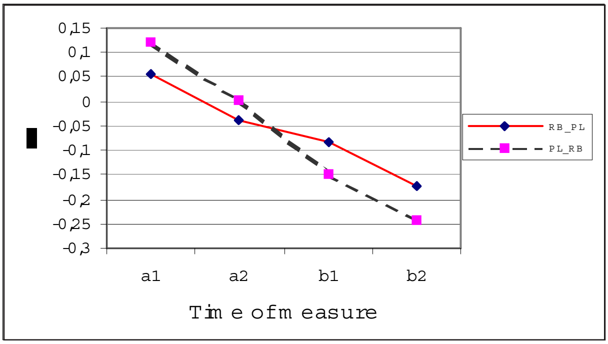

Figure 2. Results in MC for both treatment groups for both phases on both days

D2 letter cancellation test. No significant time*treatment effects (p=.322) appeared for the score total of performance minus errors.

Further results

Positive effects for the energy drink were obtained for Stroop color word performance (p=.035) and motor reaction time (p=.066) in the reaction time test. Figure 5 depicts the results for errors in the Stroop test.

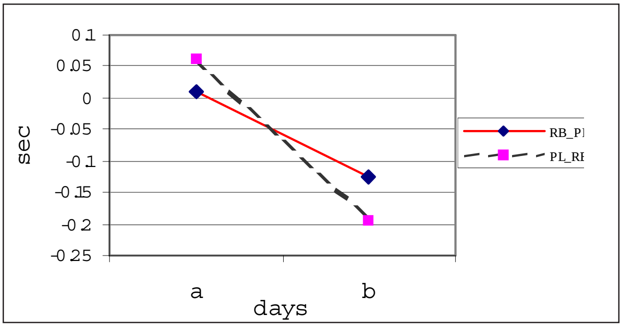

Figure 3. Results for the reaction time task.

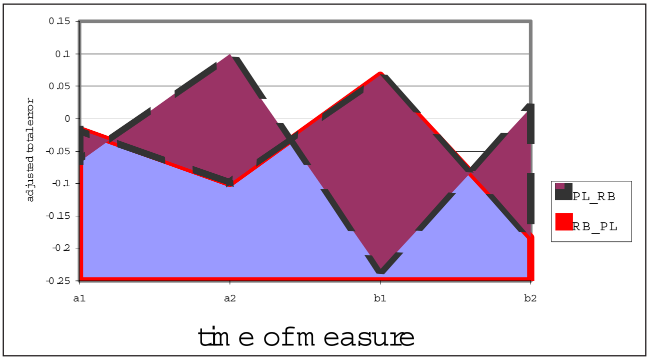

Figure 4. Results for the total of errors in the Stroop test

Mood state. The positive effect of the energy drink was well reflected in the subjective state of the subjects. Especially subjective deactivation was reduced (p=.052).

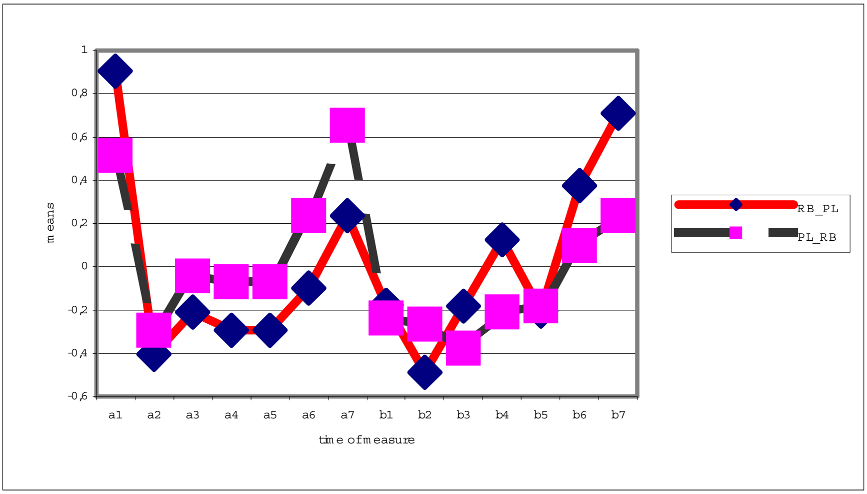

Figure 5. Means of deactivation for both treatment groups and all times of measure

Heart rate. Heart rate reflected the different task conditions and showed no systematic changes due to the energy drink.

Discussion

Clear cut positive effects on attention could be demonstrated for 250ml Red Bull® Energy Drink with controlled type one error (α =.05). The results showed interesting correspondences with the positive results reported by Horne and Reyner (2002).

In our study the subjects were only nonsmokers and low to moderate caffeine users and the subjects were not deprived from caffeine. Thus, the results cannot be attributed to a caffeine withdrawal effect (James, 1997). The present results are in accordance with the results reported in Warburton (1995), who showed, that a small dose of caffeine could improve cognitive performance significantly. In another study with a caffeine and taurine containing Energy Drink by Warburton (2001), the results concerning mental performance were confirmed. In Warburton´s study, the subjects were not deprived from caffeine.

In the present study, effects of a break with an energy drink on performance were considerably large. This implied some practical significance in addition to the statistical significance. Considering the difference in decision time of correct decisions in the continuous match to sample test, even highly selected subjects like pilots show an improvement of about 0.1 second with Red Bull® Energy Drink. Decision time is increased by an amount, which is close to the total amount of a simple reaction time of highly practiced subjects.

Performance increments could be demonstrated in the late evening under tiring and monotonous working conditions, which increases impact of the results for night and shift work, which is common in all areas of aviation.

The observed high stability of enhanced performance in time should receive further attention, especially because Horne and Reyner (2002) obtained similar effects with the same energy drink. The effect can hardly be explained by the low dose of caffeine. Other ingredients like taurine might contribute to a persisting increase in performance. Animal studies with these substances suggest that this might be an interesting and important area of future research (Vohra & Hui, 2000; Schulz, 1988).

The study showed that breaks and nutrition in breaks can easily change the performance potential. This offers positive options to cope with fatigue as well as possible risks. Finally, it should be noted that the positive results could even be obtained in pilots, who are highly trained to cope with fatigue (Caldwell, in print).

References

Abt, K. (1987). Descriptive data analysis: a concept between confirmatory and exploratory data analysis. *Methods of Information in Medicine, 26*, 77-88.

Alford, C., Cox, H. & Wescott, R. (2001). The effects of red bull energy drink on human performance and mood. *Amino Acids, 21* (3), 139-150.

Baeumler, G. (1985). *FWIT: Farbe-Wort-Interferenztest (nach Stroop, 1935) [The Stroop-Test].* Goettingen: Hogrefe.

Brickenkamp, R. (1994). *Test d2: Aufmerksamkeits-Belastungs-Test [The Continuous Attention Performance test d2]*. Goettingen: Hogrefe.

Caldwell, J. A. & Caldwell, J. L. (2003). *Fatigue in aviation. A guide to staying awake at the sick.* Burlington: Ashgate.

Deixelberger, D. (2000). *Koffein, Pausen und Leistung [Caffeine, breaks and performance].* Graz: Not published. Institute for Psychology.

Dueker, H. & Lienert, G. A. (1959). *Konzentrations-Leistungs-Test (KLT). [The Continuous Performance Task].* Goettingen: Hogrefe.

Durlach, P. J. (1998). The effects of a low dose of caffeine on cognitive performance. *Psychopharmacology, 140* (1), 116-119.

Eggert, D. (1983). *Eysenck-Persoenlichkeits-Inventar (EPI), [The Eysenck-Personality-Inventory].* Goettingen: Hogrefe.

Erdmann, G. & Janke, W. (1984). *Mehrdimensionale koerperliche Symptomliste (MKSL) [Multidimensional Scale on Physical Symptoms].* Berlin/Wuerzburg: University, Institute for Psychology.

Gilbert, R. M. (1976): Caffeine as a drug of abuse. In R. J. Gibbins, Y. Israel, H.

Kalant, R. E. Popham, W. Schmidt, R. G. Smart (Eds.), *Research advances in alcohol and drug problems, 3,* (pp. 49-77). New York: Wiley.

Hasenfratz, M. & Baettig, K. (1994). Acute dose-effect relationships of caffeine and mental performance, EEG, cardiovascular and subjective parameters. *Psychopharmacology, 114*(2), 281-287.

Hindmarch, I., Quinlane, P.T., Moore, K. L., Porkin, C. (1998). The effects of black tee and other beverages on aspects of cognition and psychomotor performance. *Psychopharmacology, 139*: 230-238.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65-70.

Horne J. A. & Reyner L. A. (2002). Efficacy of a 'functional energy drink' in counteracting driver sleepiness. *Physiology & Behaviour, 75*, 331-335.

Horne, J. A. & Reyner, L. A. (1996). Counteracting driver sleepiness: Effects of napping, caffeine, and placebo. *Psychophysiology, 33*, 306-309.

James, J. E. (1997). Does caffeine enhance or merely restore degraded psychomotor performance? *Neuropsychobiology, 30*, 124-125.

Janke, W. & Debus, G. (1978). *Die Eigenschaftswoerterliste (EWL) [The Adjective Check List].* Goettingen: Hogrefe.

Janke, W. et al. (1988). *Pharmaka-Verbrauchs-Fragebogen (PVF)[Questionnaire on drug consumption].* Wuerzburg: University, Institute for Psychology.

Janke, W. et al. (1976). *Fragebogen zur Ausgangslage (FAL) [The Initial state Questionnaire].* Wuerzburg: University, Institute for Psychology.

Kallus, K. W. (1991). Initial State and the Law of Initial Value. *Journal of Psychophysiology, 5*, 109-110.

Kallus, K. W. (1995). *Der Erholungs-Belastungs-Fragebogen (EBF) [The Recovery-Stress-Questionnaire].* Frankfurt a.M.: Swets Test Services.

Kerr, J. S., Sherwood, N. & Hindmarch, I. (1991). Separate and combined effects of social drugs and psychomotor performance. *Psychopharmacology, 104*,113-119.

Kenemans, J. L. & Verbaten, M. N. (1998). Caffeine and visuo-spatial attention. *Psychopharmacology, 135*, 353-360.

Lehmacher, W. (1986). *Verlaufskurven und Crossover. Statistische Analysen von Verlaufskurven im Zwei-Stichproben-Vergleich und von Crossover-Versuchen [Trends and Crossover].* Berlin: Springer.

Lieberman, H. R., Wurtman, R. J., Emde, G. G., Roberts, C. & Coviella, I. L. (1987). The effects of low doses of caffeine on human performance and mood. *Psychopharmacology, 92*(3). 308-312.

Lorist, M. M., Snel, J., Kok, A. & Mulder, G. (1994). Influence of caffeine on information processing stages in well rested and fatigued subjects. *Psychopharmacology, 113*(3-4), 411-421.

Reyner, L. A. & Horne, J. A. (1997). Suppression of sleepiness in drivers: combination of caffeine with a short nap. *Psychophysiology, 34*, 721-725.

Reyner, L. A. & Horne, J. A. (2000). Early morning driver sleepiness: effectiveness of 200mg caffeine. *Psychophysiology, 37*(2), 251-256.

Schuhfried, G. (1994). *Cognitrone: Test zur Erfassung der Aufmerksamkeit [Cognitrone: a match to sample test].* Moedling: Schuhfried.

Schuhfried, G. (1996). *Wiener Reaktionstest (RT) [Vienna Choice Reaction Test].* Moedling: Schuhfried.

Schuhfried, G. (1995). *Pilot Spatial Test*. Moedling: Schuhfried.

Schulz, H. (1988). Spatial working memory in rats: the effect of the dipeptide gamma-L-glutamyl-taurine and haloperidol. *Acta Physiologica Hungaria, 72*(3-4), 263-7.

Tewes, U. (1991). *HAWIE-R: Hamburg-Wechsler Intelligenztest fuer Erwachsene – Revision [The Hamburg-Wechsler intelligence test for adults].* Bern: Huber.

Vohra, B. P. & Hui, X. (2000). Improvement of impaired memory in mice by taurine. *Neural. Plast, 7*(4), 245-259.

Warburton, D. M. (1995). Effects of caffeine on cognition and mood without caffeine abstinence. *Psychopharmacology, 119*(1), 66-70.

Warburton, D. M. (2001). An evaluation of a caffeinated taurine drink and mood, memory and information processing in healthy volunteers without caffeine abstinence. *Psychopharmacology, 158*(3), 322-328.

International Journal of Applied Aviation Studies, 3, 2

Traditional versus Electronic Information Delivery: The Effect on Student Achievement

Cass D. Howell, Todd V. Denning
and William B. Fitzpatrick
Embry-Riddle Aeronautical University
600 S. Clyde Morris Blvd
Daytona Beach, FL 32114

Abstract

The Department of Aeronautical Science at Embry-Riddle Aeronautical University recently moved into a state-of-the-art instructional building. Along with the new building came teaching opportunities previously unobtainable, or at least difficult to achieve on any scale. One of these occasions involved a requirement for students in a Flight Safety class to obtain instructional material (journal and magazine articles) independently from a variety of on-line sources, versus the traditional method of in-class distribution of handouts. Anecdotal evidence seemed to indicate that students tasked with retrieving and utilizing the publications themselves were less proficient in correctly answering test questions drawn from the material than were students who were provided the handouts of the same publications in class. A study was devised to test this observation, involving test and control groups of students in four Flight Safety classes over two semesters. Results indicated that, while there may have been significant differences in correct response rates to individual questions between the two groups, in the aggregate, there was no statistically significant difference in performance at the .05 level between test and control groups that was attributable to an electronic versus traditional mode of delivery of these specific course materials.

Requests for reprints should be sent to Beverly Laughead, FAA Academy, AMA-502-OU, P.O. Box 25082, Oklahoma City, OK 73125.

Introduction

Nature of the Problem

The acquisition of an advanced technology instructional building, plus the availability of a new academic intranet software service, Blackboard, has made alternative methods of instruction possible on a scale never before possible at Embry-Riddle Aeronautical University's Department of Aeronautical Science. This parallels the evolution of electronic delivery at most schools and colleges across practically the entire spectrum of subjects. However, does the availability of this new technology translate to increased learning?

The Purpose of the Study

The purpose of the study was to find out if changing a specific element of instructional material from traditional to electronic delivery had a measurable effect on student learning. To keep this small study to an appropriate size, it was decided to focus on one component of the available technology, i.e., the ability to provide supplemental materials, in this case magazine and journal articles, by electronic postings. Based on non-scientific observations, it was the authors' impression that students were not as successful in correctly answering questions originating from material that they were required to download electronically as were students who obtained the same material from handouts in class. Since the trend, at least at this university, (and presumably many more) is to rely more and more on various forms of electronic delivery, it was deemed to be a matter of some importance to ascertain the effects these modes of instruction had on student performance. Depending on the results, this information could guide future instruction, or at least suggest how any negative findings could be addressed.

Review of the Literature

As is readily evident to even the most casual observer, non-traditional methods of instruction, many of them via electronic delivery, have flourished in the last few years. (As used here, "electronic/digital delivery" includes the spectrum from a simple modification such as converting lecture notes to Power Point, to the use of sophisticated interactive websites that operate sans professors.) Even as early as 1995, higher education institutions had spent 20 billion dollars equipping with computers and associated software and support systems (Katz, Tate, & Weimer, 1995, cited in Jones & Paolucci, 1999). A few schools are making sweeping commitments to electronic access and delivery. Massachusetts Institute of Technology, perhaps foremost in this group, is in the process of publishing on line the lecture notes, assignments and reading lists for all of its 2000 on-campus courses (Massachusetts Institute of Technology, 2003) and Farleigh Dickinson University now requires all entering students to take at least one course on-line each year (Young, 2002). Many other schools, seeking to reach new students and to increase their revenue bases, have made great efforts to develop on-line versions of their traditional courses to reach students virtually anywhere on the globe. The U.S. military, whose members have been in almost constant motion for the last few years, has embraced electronic distance education to a vast degree. For example, the U.S. Army recently began its first civilian education Internet portal, eArmyU, and plans to enroll up to 80,000 students in the next two years. The program is supported by 23 colleges throughout the nation, who offer Army students 90-degree programs. The Army will spend over 450 million dollars in the first five years of the project to purchase the services and facilitate the colleges' course delivery (Arnone, 2002).

More recently, "hybrid" courses, which incorporate a melding of classroom and distance education, have become more popular (Young, 2002). These courses seek to blend the best of traditional classroom instruction with the flexibility of virtual classroom instruction into a more efficient instructional model. Bourne, editor of the *Journal of Asynchronous Learning Networks*, says in the *Chronicle of Higher Education*, (Young, 2002, p. 33), "I would guess that somewhere in the 80-90 percent range of classes could sometime become hybrid."

Nevertheless, how effective are these new approaches in facilitating learning? Despite the embrace of academics and the rapid expansion of electronic forms of delivery, there is not a quantity of evidence demonstrating that these methods of instructional delivery are equal to or better than the more traditional. Jones and Paolucci (1999, p. 17) agree saying, "…research supporting the massive adoption of technology simply doesn't exist to the extent that these widespread trends are justifiable." They estimate that "…since 1993 less than 5% of published research was sufficiently empirical, quantitative, and valid to support conclusions with respect to the effectiveness of technology in educational learning outcomes" (p. 18). Those studies that have been done show decidedly mixed results. For example, Kendall (2001) reported high student satisfaction overall and higher grades for those who accessed WebCT based course notes and posted messages most frequently in a study of 46 communications students. However, a study by Gay and Grace-Martin reported in the *Chronicle of Higher Education* (Carlson, 2001) found while students who visited more websites during computer science and communications classes made higher grades, those communications students who spent more time on-line outside of class were less successful academically. A Canadian study by Goldberg (1997) indicated that students in the hybrid version of a computer class out-performed both students who took the course exclusively on-line as well as those who completed it in a traditional lecture based format. Contrasting this is Byers' (2001, p. 360) assessment that "…there is no significant difference in direct learning between technology based and face-to-face learning…."

It would appear from personal experience and a survey of the available literature, which includes many "how to" articles, that emerging technology has been embraced and incorporated into the teaching and learning process largely for reasons other than enhanced academic achievement. While some of these reasons are compelling in their own right (lower costs, ease of availability, better graphics, etc.), the academic value of the various forms of electronic delivery is either assumed, or in some cases, hoped for (Willet, 2002, Wallace & Mutooni, 1997, are program/ technology descriptions that illustrate this trend.). A conclusion can be drawn that the temptation is to use whatever technology is available without regard to (or awareness of) learning efficacy.

Methodology

The research hypothesis for this study was as follows: "There is a significant difference between the success rates of Flight Safety students who obtained class materials electronically versus those who were provided the same materials by hand-outs in class." The dependant variable was defined as the aggregate correct response rate for test questions drawn from the subject material, and the independent variable were the two methods of obtaining the material, i.e., electronic or hand-out.

Data collection was based on regularly occurring academic tests that were part of a three credit-hour university course, Flight Safety. Students in this course are typically juniors and seniors, have limited yet significant flight experiences, and have usually earned FAA commercial and instrument certificates and ratings. Their goal, in most cases, is to become airline pilots with U.S. and international carriers.

As part of the course requirements, students were to read a selection of articles related to the objectives of the course. The test group was directed to electronic versions of the articles, posted on either Blackboard, an educational software program to which the University subscribes, or on the World Wide Web at several websites. In both cases, specific addresses were provided during class meetings. The control group was provided the articles by traditional means, i.e., individual handouts in class. The test and control groups were different sections of the same Flight Safety course. Both of the two sections were provided identical instructions and syllabi from the same professor, used the same text, and even met in the same classroom. The sections met sequentially, i.e., "back-to-back."

The study was conducted over the course of an academic year. During the initial semester (Fall) the first section (01) was the test group, and the second section (02), was the control group. In the Spring semester this was reversed, so that the first section was the test group and the second section was the control group. This was done to minimize the possibility of skewed results from individual students in the second class obtaining test information from students in the first class. The groups averaged about 30 students per section, with very slight variations from late enrollments and drops. Class Grade Point Averages (GPAs) were calculated for each test/control pair, with results indicating a close match in each case. Grade Point Averages were therefore determined to not be a factor in the research design. Test questions specific to the designated articles were developed. These were included on two regular multiple-choice evaluations that were given during the conduct of the course. Upon completion of the tests, the relevant responses were isolated and analyzed per the procedures described in the next section.

Results

Raw scores were collected from four separate groups. The raw scores were converted into percentage of incorrectly answered questions on each test and then analyzed by method of teaching (i.e., traditional versus electronic groups). Table 1 shows the proportions for each group by test. Proportions were calculated by dividing the number of individual wrong answers on each question by the size of each control group. Each of the four tests had a control condition (i.e., traditional teaching method) and a treatment condition (i.e., electronic teaching method). Therefore, each test has two scores representing the mean percentage answered incorrectly.

For the traditional group, the average percentage of incorrectly answered questions was 20% (SD=15). In the electronic group, the average percentage of incorrectly answered questions was 29% (SD=19).

A Levene's Test for Equality of Variance was conducted and the results indicate that the test scores for the four separate test groups do vary approximately the same. Table 2 shows no significant difference was found between the means of the four separate test groups. Because the 95% confidence interval was used between the groups, we can conclude that no significant differences were found between the groups.

Table 1
Test Scores for Traditional and Electronic Teaching Methods

Question Number	Test1		Test2		Test3		Test4	
	Trad.	Elec.	Trad.	Elec.	Trad.	Elec.	Trad.	Elec.
1	.055	.147	.030	.125	.212	.200	.364	.300
2	.333	.352	.060	.264	.454	.200	.06	.150
3	.250	.323	.242	.437	.272	.300	.09	.150
4	.055	.176	-	-	.181	.150	.52	.500
5	.111	.176	-	-	.030	.150	-	-
6	.333	.705	-	-	.454	.850	-	-
7	.194	.235	-	-	.181	.400	-	-

Note. Test scores are given in percentage of population with wrong answers

Table 2

Inferential Statistics by test using Independent Samples *t*-test

Test Number	Significance	95% confidence interval	
		Lower bound	Upper bound
Test One	.509	-.299	.076
Test Two	.726	-.475	.146
Test Three	.451	-.308	.175
Test Four	.305	-.355	.322

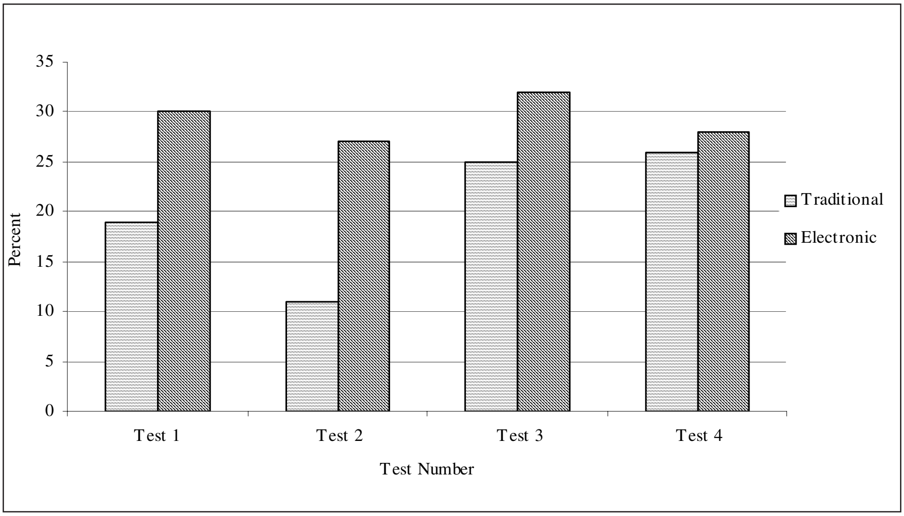

Figure 1. Graphic display of Test Scores

Conclusions

Despite the higher overall incorrect response rate (29% versus 20%), student achievement was not significantly adversely affected by the requirement to obtain course materials electronically on their own. This finding is consistent with Russell (1999), who determined that there was no significant difference in the academic performance of students in on-line and traditional courses.

It should be noted that there may well have been significance in correct response rates on individual questions; however, due to small sample sizes this was not feasible to examine and draw conclusions. The most relevant aspect is that, in the aggregate, students are equally successful in performance regardless of the method of obtaining the instructional information. It may be speculated that today's students are sufficiently computer literate, and, more importantly, possess enough motivation to take the extra steps required to obtain and utilize the electronic material. This would be expected with easy access (computer labs on campus, web access in the residence halls, high rate of personal computer ownership), need for electronic research in other courses, and the generally high degree of adeptness expected of a generation of students who have always had access to computers as part of their teenage experience onward. One might wonder if this research would have had different findings if it had been conducted, say, five years ago; however, that would be looking backward. The view towards the future seems to be that we can be reasonably confident that electronic delivery, at least as far as this small study can be extrapolated, is a viable and efficient mode of instruction as compared to traditional delivery means.

References

Arnone, M. (2002). Army's huge distance-education effort wins may supporters in first year.The Chronicle of Higher Education: 22, A33-35.

Byers, C. (2001). Interactive assessment: An approach to enhance teaching and learning. *Journal of Interactive Learning Research: 1*, 359-374.

Carlson, S. (2001). Wireless technology is a double edged sword, researchers conclude. *The Chronicle of Higher Education: 32*, A55.

Goldberg, M. (1997). WebCT and first year: student reaction to and use of a web based resource in first year computer science. Proceedings of *Integrating Technology into Computer Science Education, Uppsala, Sweden, 1997.*

Jones, T., & Paolucci, R. (1999). Research framework and dimensions for evaluating the effectiveness of educational technology systems on learning outcomes. *Journal of Research on Computing in Education, Fall 1999, 1:* 17-32.

Katz, E., Tate, R., & Weimer, W. (1995). Survey of telematics for education and training: Vol 2, United States [Online document]. European Association of Distance Teaching Universities, 1995. Available: http://194.70.69.3/ethos/survey/vol2/techsch.htm

Kendall, M. (2001). Teaching online to campus based students: The experience of using WebCT for the community information module at Manchester Metropolitan University.*Education for Information: 19*, 325-346.

Massachusetts Institute of Technology. (2003). MIT OPENCOURSEWARE. Retrieved April 2, 2003, from http://ocw.mit.edu/index.html

Russell, T. L. (1999). *The No Significance Phenomenon.* Raleigh, NC: North Carolina State University, Office of Instructional Communications.

Wallace, D. & Mutooni, P. (1997). A comparative evaluation of world wide web-based and classroom teaching. *Journal of Engineering Education, July 1997:* 211-219.

Willet, H. (2002). Not one or the other or both: hybrid course delivery using WebCT. *The Electronic Library: 5*, 413-419. Retrieved February 19, 2003 from http://emeraldinsight.com0264-0473.htm

Young, J. (2002). "Hybrid" teaching seeks to end the divide between traditional and online instruction. *The Chronicle of Higher Education*, *28:* A33-34.

International Journal of Applied Aviation Studies, 3, 2

Advanced Collegiate Flight Automation Training: What Is Needed and At What Cost?

John P.Young and Richard O. Fanjoy
Purdue University
Department of Aviation Technology
West Lafayette, IN 47907-3398

Abstract

The prominence of new flight technology instrumentation and problems associated with its use has prompted flight program educators to consider a variety of training aids and methods to prepare flight graduates for employment. The expense associated with latest generation training aircraft, flight training devices, and materials has led most collegiate aviation program administrators, however, to defer such training and related costs to the aviation industry. The authors recommend a tailored approach to advanced collegiate flight training that will take advantage of cost effective program components and generate the potential for interaction with aviation industry partners. Several issues related to funding and operation of advanced training resources are presented. In addition, three levels of advanced technology implementation are suggested to effectively address the needs of collegiate aviation students, within the operating constraints of the host school. Finally, opportunities for cooperative training arrangements with aviation industry partners are discussed. Such arrangements have had and will continue to have lasting benefit to both collegiate aviation and the air transportation industry.

Over the last two decades, advances in computer technology have led to amazing sophistication in aircraft and training systems resources. During the same period, such advances have led to many unforeseen challenges for crewmembers whose training historically has focused on psychomotor and procedural competencies. Airline pilot surveys and an evolving trend in flight automation-related accidents/incidents have highlighted the need for increased at-

Requests for reprints should be sent to Beverly Laughead, FAA Academy, AMA-502-OU, P.O. Box 25082, Oklahoma City, OK 73125.

tention to latest generation flight technology (Fletcher et al., 1997; Funk & Lyall, 1999; Hughes & Dornheim, 1995; Sherman & Helmreich, 1997). The costs associated with modern flight training equipment and media, however, have led much of the air transportation industry to outsource the majority of initial, aircraft-specific training to high-volume, resource-intensive flight training academies (Warwick, 2003). In addition, airline training sections have found great utility in commercially produced computer-based training materials for a wide variety of continuing and upgrade-training segments (Doherty, 2003). Collegiate aviation faculties also have an interest in presenting training that reflects the latest technology to best prepare their graduates for a very competitive job market. Advanced flight training resources currently used in collegiate programs vary widely from lecture materials and video to computer-based training programs to modern regional jet flight training devices (FTDs). Since most collegiate programs operate within a fairly restrictive funding base, support for the latest generation of flight training resources is limited. Despite funding shortfall for advanced technology flight training resources, collegiate program administrators attempt to maintain an appropriate level of training, tailored to a school's particular resource limitations. This paper will address issues associated with advanced flight training resources as well suggested levels of implementation that address the particular needs and financial constraints of each collegiate flight training unit. By doing so, the authors hope to present important programmatic considerations of interest to flight program administrators and educators.

Aviation Program Survey

Young and Fanjoy (2002) completed a survey of four-year flight programs in the U.S. to identify resources currently used to teach glass cockpit familiarization. Most program administrators indicated that this type of training was very important to meeting their curriculum goals. However, only 51 percent of the schools surveyed stated they taught this newer technology to their students. While most schools indicated they would like to initiate this type of training, resource costs and an already saturated curriculum were major barriers to doing so.

Results of the survey indicated that faculty who presented glass cockpit technology (usually during the upper division course work) preferred hands-on, interactive materials or devices to teach this subject area. Desktop trainers, interactive computer programs, flight management system (FMS) trainers, and flight training devices were perceived to be more effective than more traditional means of presenting material, including lecture, video, and reading assignments. Several of the cockpit training aids reported by collegiate programs were donated or in joint use with airlines and other training organizations. Virtually all survey respondents indicated that cost was the primary reason that prevented more effective teaching in this curriculum area. Schools were very interested in the potential of low-cost alternatives to more expensive flight training devices and full flight simulators.

Issues Related to Technology Procurement

The wide variety of issues associated with the purchase of advanced flight training equipment frequently leads to several "show-stopper" questions. Should the eventual employer or the collegiate flight program be responsible for advanced automation training? Is the overall cost for advanced flight training enhancements out of the range of most collegiate flight programs? What burden sharing is appropriate for students in support of new flight training resources? Can new training resources be effectively funded through rental to outside users? What type of advanced training resources best mesh with the program's objectives and funding limitations? Each of these questions will be considered in turn.

Training Responsibility

The objective of most collegiate flight programs is to produce a graduate with 200 to 300 flying hours who needs some additional experience and training before being considered for duty with a commercial airline. Although exposure to advanced flight systems is desirable, the authors' survey indicates that many administrators believe such training falls within the responsibility of the employing airline. Roessingh et al. (1999) conducted 58 interviews with pilots and training instructors to evaluate transition training for pilot upgrade to automated aircraft from traditional round-dial instrumentation. Findings from that study suggested that insufficient time is allocated to transition training courses in general, and aircraft difference training in particular to adequately prepare pilots for operational flight duty.Although many collegiate flight faculty members would like to add advanced flight technology training to their curricula, they are prevented from doing so by staff availability, limited physical resources, and austere budgets that frequently make such considerations prohibitive (Young & Fanjoy, 2002). In addition to funding and staffing, there are a number of other factors that impact a school's potential to conduct advanced automation training. Some collegiate flight programs have invested time and money in advanced technology resources only to experience hidden costs or find that such resources did not integrate well with program goals. Extraordinary student fee increases to support activity in a new full flight simulator or flight-training device, for example, may price the overall flight program out of the reach of most students. On the other hand, if the school accepts donated industry resources, such equipment may only provide "showroom pieces" that are too expensive to update or maintain. Finally, internet-based training has been embraced by many airlines as a relatively lost cost option, but the cost of access to college flight students may be unreasonable. Although some level of advanced flight training technology is appropriate for most programs, many issues must first be addressed to insure new program additions provide a good fit for the department in question.

Budget Tradeoffs

When flight training administrators conduct annual budget deliberations, aircraft and the resources to maintain them are assigned a very high priority.The significant costs associated with aircraft, ongoing maintenance, and associated insurance forces many collegiate aviation departments to consider outside flight training providers. A similar philosophy has led many airlines to use independent flight training centers for initial and recurrent training (Warwick, 2003). Although simulators are much cheaper and safer to operate than actual aircraft, modern flight simulators, in some cases, have become more expensive than an actual aircraft. Barber (1997) noted that modern flight simulators can be priced as much as ten times the cost of the aircraft they represent. Unfortunately, the end users/system operators must "foot the bill" for such expensive technology. In consideration of expensive simulation assets and the impracticality of student travel to distant advanced flight training centers, collegiate programs frequently are limited in the level of advanced flight training they can provide. Although many college programs employ relatively simple light airplane training devices, few can afford an advanced full flight simulator (FFS) or flight training device (FTD) for their flight student population. Initial funding for either can easily amount to several million dollars. Even when the considered resource is a donated FFS or FTD, shipping and installation costs alone can easily amount to over $100,000 (W. Bauman, personal communication, May 20, 2003). A collegiate program administrator must consider the long-term utility of high priced flight resources as well as hidden costs of operation. A flight department, for example, could accept a donated advanced flight simulator only to find that corresponding costs of maintenance and operations consumed an inordinately large share of the budget. If, on the other hand, the school elected to purchase a state-of-the-art glass cockpit flight training device, student costs to support that resource might be unacceptable. Such concerns make capital outlays for advanced technology, either in the form of aircraft upgrades, simulation equipment, or extensive software/internet support, unlikely for all but the most fiscally solvent collegiate programs. Fortunately for the collegiate consumer, innovative companies have begun to market lower-cost training equipment alternatives (Fitzsimmons, 2003).

Burden-Sharing

Student burden sharing of advanced training costs is a fact of life. With the current trend in rising tuition across the country, additional fees for aircraft and flight simulator training can easily place a collegiate aviation education out of the reach of most college students. Accordingly, it is imperative that additional costs of modern training resources be carefully considered and distributed through multiple users, rather than simply tacked on to existing student fees. The impact of training resources on student fees can be minimized through research grants, shared used by other campus departments, and through external fees for outside users. A proposed new FTD or aircraft upgrade, for example, could find utility in support of a major grant for an engineering department. Desired part task aircraft trainers might be developed or evaluated jointly by computer science, education, or other academic units at the university. In addition, burden-sharing arrangements should be explored with industrial partners. The benefit of such an association, should be clearly identified for a prospective partner, particularly in consideration of the currently depressed aviation industry. Opportunities for a mutually beneficial partnership do exist. A ready source of future hires and training methods research are just two of the benefits that might be offered. A change in the operating environment, however, may negate the benefit to the industrial partner at any time. A downturn in the already volatile airline industry or change of aircraft type can lead to termination of sponsorship. Such concerns are every day fare in the commercial sector but may not be intuitive for a collegiate flight program.

Resource Rental

Funding for new flight training resources may be offset through resource rental to outside users for either refresher or initial flight training. A college flight program may be able to provide such services to individuals at a cost well below the fees charged by a major flight training center. Dry leasing of equipment by industry partners can also provide significant support for collegiate flight program resources and should be actively courted. Such arrangements, however, can have negative aspects for both partners. An industrial partner may require a level of scheduling flexibility, for example, which does not correlate well with weekly student use. In addition, the industrial partner may want commonly shared resources to be located near a major airline hub for easy employee access. An airline partner may be supportive of a collegiate flight-training program, but will not want to add extra costs to their operation. Accordingly, creative and flexible solutions may be indicated to benefit each partner.

Maintenance Concerns

Equipment maintenance and replacement parts may constitute a significant investment in support of an advanced FFS or FTD. Current industry standards dictate a minimum of two technicians present in support of each simulator during operation. Although salaries for technicians vary widely, the ability to maintain equipment such as modern simulators is highly prized and commands a respectable amount. In addition, a standing parts inventory for a modern FFS may amount to more than $200,000. Individual parts, whether they represent rehosted aircraft equipment or simulated instruments, are very expensive. A typical "round dial" airliner attitude indicator, for example, may cost $20,000 to repair or replace (W. Bauman, personal communication, May 20, 2003). Although an individual component may be commonly used to support of a wide variety of simulators in a typical flight training center, a university may only have a single flight training device, increasing the relative cost/value of each spare component.

Resource Alternatives

Some college programs may desire to enter the advanced flight technology market on a smaller scale, with a part-task trainer or computer-based training program that uses an Internet/CD/DVD format. Several vendors currently produce these resources in a wide range of costs and capabilities. Inherent in the purchase consideration is whether the media/equipment can be upgraded and how it integrates with other systems in use by the department. If a particular variant of a flight management system is purchased, for example, will the skills it supports be easily transferable to other program aircraft or classroom instruction? Is "free-play" available to reinforce student understanding of component operation? Can a part-task trainer such as an electronic flight instrument system (EFIS) be integrated with programmed computer-based assets and building space forecast by the department? Does the new resource under consideration have to exactly duplicate an existing aircraft? Recent studies, for example, seem to show that generic training devices provide a significant transfer of training to wide variety of aircraft systems (Johnston, 1998; Lintern & Naikar, 1999; Teunissen, 1999). The cost of a generic FTD can be a fraction of that incurred for a type-specific piece of flight training equipment. In addition, several studies on computer-based training (CBT) reflect great value for that training method, despite its minimal presentation of psychomotor aspects (Dennis & Harris, 1998; Koonce & Bramble, 1998; Moroney, Hampton, and Eggemeier, 1997; Taylor et al, 1999). The Federal Aviation Administration (FAA) and the airline industry have already accepted the findings of many of these studies and have expanded CBT offerings as a prelude to actual FFS or aircraft training (Adams, 2003; Doherty, 2003; Phillips, 2000).

Instructional Design

Most collegiate flight program administrators believe that technology associated with advanced flight systems presents a significant opportunity for addressing important program goals and further refining the high quality of their graduates. Advances in flight instrumentation have led to significant growth in the flight training industry over the last decade, yet the costs of high-end flight training resources seem well out of the reach of most collegiate flight programs. The authors believe there are opportunities to match new resources with program capacity and improve the level of offerings in support of advanced flight instrumentation training. Rather than indiscriminately purchasing flashy new toys, however, program administrators must first reconsider their program objectives and determine whether new technology training is appropriate for their program and if so, which elements are most cost effective. Instructional design methodology can be very useful in making such a determination. Brown (2001) described a proficiency and mission-oriented training model, developed by United Airlines, which sequentially considers training needs, corresponding training objectives, and training devices that are matched to the objectives. Such a model may be very useful to college program administrators who want to update their curriculum offerings to include modern aircraft operations. When program objectives are weighed against competing program/budgetary priorities, appropriate training resources can be identified.

Levels of Glass Cockpit Training

Upon review of vendor contacts and survey information from collegiate aviation schools, the authors suggest three possible levels of training for aviation programs (Fanjoy & Young, in press). These levels vary in complexity and cost. The suggested levels are certainly not comprehensive and school administrators must evaluate curriculum objectives and available resources in determining an appropriate level of advanced flight training resources for its student population. A description of three proposed levels, including advantages, disadvantages, and typical costs, is presented.

Orientation Level

An introduction to new generation aircraft can be accomplished at this level. Basic nomenclature, terms, and cockpit layout can be discussed. The primary training focus is the identification and basic operation of electronic flight instrument system (EFIS) components, such as primary flight displays (PFDs), multi-function displays (MFDs), engine indication and crew alerting system (EICAS), and FMS. Integration of automated flight systems, such as flight directors and autopilots, can also be presented.

The orientation level may be accomplished through traditional teaching media and methods. Lecture, supplemented by handouts, overhead transparencies, video, PowerPoint slides, and/or other similar media, is commonly used to orient the new student to advanced cockpits. In addition, several low cost computer programs are available to introduce glass cockpit and automation technology.Advantages of the orientation level include instructor familiarity with materials, availability of classrooms and resources that support this instructional delivery, and relatively low cost of materials. Disadvantages may include a lack of student interaction with the new technology, the use of "generic" materials that do not support detailed training objectives, and an inability to demonstrate how all glass cockpit components are integrated within the aircraft.

Materials that support the orientation level may be found over the Internet, in textbooks, from partner airlines or training organizations, and through vendors. Some resources may be obtained at no cost while others may be priced up to several hundred dollars. This level of flight automation training seems very affordable for most four-year schools.

Limited Familiarization Level

This level of training provides a more sophisticated means for learners to become engaged with the new technology. During this level of training, students are exposed to more "hands-on" operation of aircraft systems. This may be accomplished through part-task systems media, FMS trainers, virtual cockpits, and other computer-generated displays that facilitate student interaction with airplane systems in either a structured or a "free-play" format. Typical resources that represent this level of training might be system mock-up panels or interactive computer software.

Advantages of this level of training include increased student interaction with specific aircraft systems. If the training is focused on a particular type of aircraft, students can spend the time needed to become proficient in the use of specific systems. This saves valuable training time when the student transitions to an FTD, FFS, or actual aircraft. Students can work in a less stressful environment, at their own pace, using these types of educational devices. The main disadvantages of these resources are increased purchase/lease costs and possible maintenance costs above those used in the orientation level of glass cockpit training. In addition, training is limited to a subset of actual aircraft systems and the student may not have the opportunity to master systems integration required for effective operation of the aircraft in all phases of flight.

Typical materials that support this level of training can be purchased from a number of vendors or obtained from partner airlines or training organizations. The cost for such resources can range from several hundred to several hundred thousand dollars. Survey findings suggest that many collegiate aviation programs are currently interested in resources that will address this training level (Young & Fanjoy, 2002).

Full Integration

Schools operating at this level employ flight training devices, flight simulators, and/or actual aircraft to relate all operations, performance, and integration of systems into one training platform. While the majority of glass cockpit training devices in this category are airplane specific, several vendors have developed "generic" glass cockpit trainers that represent an entire class of modern aircraft (i.e. regional jets). For educational purposes, these devices can be used to present the basic concepts and procedures of new generation aircraft at a lower cost than traditional aircraft simulators. Several airlines, aviation schools, and training organizations are adding newer generation training devices to their inventory to offset training time in actual aircraft (Seidenman, 2002).

The full integration level allows students to manipulate all the controls of the aircraft and to observe the effects on various aircraft systems. In these training devices most, if not all, systems are accurately represented in all possible modes of operation. Students will not only be able to touch and manipulate aircraft system controls but also comprehend the interrelationship of systems to one another. In addition, this type of training device may be flown, unlike lower level resources. This allows the student to experience the operational aspects of systems and flight controls in actual (or simulated) flight conditions.

The advantage of the full integration level is that students acquire an accurate picture of how all systems interrelate. If a student is preparing for flight in an actual aircraft, a high fidelity flight training device or simulator can be used to minimize safety concerns as well as actual flight time required for proficiency. Airline training sections currently use modern full flight simulators to great advantage. In certain training programs, airline flight crews are allowed to complete all type-rating and airline transport pilot certificate training and evaluation requirements in a Level D full flight simulator.The primary disadvantage to this methodology is cost – in both initial outlay and maintenance support funding. Unit costs vary from approximately $200,000 for a "generic" device to $15M or more for an aircraft-specific full flight simulator. In addition, significant space and technical support are required to operate such devices. The high cost per hour to operate these more expensive devices currently put them beyond the reach of most collegiate flight training programs (Fanjoy & Young, in press).

Conclusion

Virtually all transport aircraft are now constructed with glass cockpit instrumentation and enhanced automation. Open literature seems to indicate that general aviation aircraft will soon follow suit. A survey of college educators suggests the importance of training in the operation of modern flight automation systems. Such knowledge is critical to the effective preparation of college aviation graduates for future employment. The authors have identified some key issues and program options that may be useful when considering the objectives and opportunities of a typical collegiate aviation program. Many aviation departments are currently pondering major program adjustments to meet regulatory changes and changing department objectives. In addition, an evolving emphasis on research in aviation education offers the potential to reap additional benefits from newly acquired training resources. Although the proposed levels of advanced flight training resources presented here are not intended to be prescriptive, they do offer a starting point for consideration by members of a training unit's flight faculty.A careful consideration of curricular objectives and available resources may present aviation faculty with an opportunity to provide great service both to their students and the commercial aviation community at large.

References

Adams, R. (2003). The Internet is finally paying off. *The Journal for Civil Aviation Training, 3*, 31-34.

Barber, S. (1997). From PC's to flight simulators: cost effective aircrew training deices for aeronautical decision making and crew resource management. *Proceedings of the Ninth International Symposium of Aviation Psychology*. (pp. 1178-1183). Columbus, OH: The Ohio State University.

Brown, M. D. (2001). A global airline: Training at the speed of change. *Proceedings of the RAeS Flight Simulation Group Conference*, 3.1–3.8.

Dennis, K. A. & Harris, D. (1998). Computer-based simulation as an adjunct to ab initio flight training. *The International Journal of Aviation Psychology, 8* (3), 261-276.

Doherty, M. (2003). Efficiency and effectiveness: Developments in cbt and multimedia learning. *Touchdown 3* (1), 32–35.

Fanjoy, R. & Young, J. (in press). Training levels and methodologies for new technology aircraft in collegiate aviation. *Journal of Aviation/Aerospace Education and Research.*

Fitzsimmons, R. (2003). Flight training devices: Doing more with less. *Touchdown, 3* (1), 28-31.

Fletcher, G.C.L., Dudfield, H.J., Davy, E.C., Crick, J., Gorton, T., Russell, S., et al. (1997). Transition training and the glass cockpit. *Proceedings of the Ninth International Symposium of Aviation Psychology*. (pp. 237-242). Columbus, OH: The Ohio State University.

Funk, K. & Lyall, B. (1999). The evidence for flight deck automation issues. *Proceedings of the Tenth International Symposium of Aviation Psychology*. (pp. 105-110). Columbus, OH: The Ohio State University.

Hughes, D. & Dornheim, M. A. (1995, January 30). Accidents direct focus on cockpit automation. *Aviation Week and Space Technology, 1*, 52-54.

Johnston, N. (1998). Towards evaluation criteria for generic and low fidelity simulations. *Proceedings of the Conference, London, United Kingdom, May 13-14, 1998 (A98-6751 09-09)*, London, Royal Aeronautical Society, 1998, (pp. 1.1-1.12).

Koonce, J. M. & Bramble Jr., W. J. (1998). Personal computer-based flight training devices. *The International Journal of Aviation Psychology, 8* (3), 277-292.

Lintern, G., & Naikar, N. (1999). Specification of training simulators by cognitive work analysis. *Proceedings of the Tenth International Symposium on Aviation Psychology*. (pp. 421-426). Columbus, OH: The Ohio State University.

Moroney, W.F., Hampton, D.W., & Eggemeier,F.T. (1997). Considerations in the design and use of personal computer-based aviation training devices (PCATDs) for instrument flight training: A survey of instructors. *Proceedings of the Ninth International Symposium* on Aviation Psychology. (pp. 87-106). Columbus, OH: The Ohio State University.

Phillips, E. H. (2000, November 27). FAA team focusing on technology, data. *Aviation Week and Space Technology*, 64-65.

Roessingh, J.J.M., Van Gent, R.N.H.W. Fletcher, G., Dudfield, H., Lodge, M., Koehl, F., et al. (1999). Training for the 'glass cockpit': Trends, problems and solutions. *Proceedings of the Tenth International Symposium of Aviation Psychology*. (pp. 985-991). Columbus, OH: The Ohio State University.

Seidenman, P. (2002). Coming in at the deep end. *The Journal for Civil Aviation Training, 8*, 22-25.

Sherman, P.J. & Helmreich, R.L. (1997). Training for use of automation: The value of 'free-play' as a complement to traditional transition training. *Proceedings of the Ninth International Symposium of Aviation Psychology*. (pp. 243-248). Columbus, OH: The Ohio State University.

Taylor, H. L., Talleur, D. A., Emanuel, T. W., Jr., Phillips, S. I., Lintern, G., &

Hulin, C. I., (1999). Transfer of training effectiveness of a personal computer aviation training device. *International Journal of Aviation Psychology, 9* (4), 319-335.

Teunissen, R. P.A. M. (1999). The future of simulation and training. *Proceedings of the Conference, London, United Kingdom, May 19-20, 1999 (A99-40076 11-09), London, Royal Aeronautical Society, 1999*, (pp. 6.1-6.11).

W arwick, G. (2003). Names of the game. *Flight International, 157*, 40-42.

Young, J. & Fanjoy, R. (2002). Glass cockpit technology training: How is it addressed in collegiate flight programs? *Collegiate Aviation Review, 20* (1), 85-94.

International Journal of Applied Aviation Studies, 3, 2

Dynamic Spatial Ability & Ability to Coordinate Information

Predictive Contributions of Dynamic Spatial Ability and the Ability to Coordinate Information for Air Traffic Controller and Pilot Selection

Teresa C. D'Oliveira[1]
Instituto Superior de Psicologia Aplicada
Lisbon, Portugal
Teresa.Oliveira@ispa.pt

Abstract

Air traffic control and piloting are occupations traditionally considered when referring to the predictive value of spatial ability. Recent developments in the spatial ability literature have identified a new ability, dynamic spatial ability. Along with the ability to coordinate information, dynamic spatial ability is considered particularly important to performance in these occupations. This study considered the predictive relevance of dynamic spatial ability and the ability to coordinate information for training results of air traffic controllers and pilots. A total of 86 trainees were administered a battery of paper-and-pencil and computerized spatial ability tests that included specific tasks to assess dynamic spatial ability and the ability to coordinate information. Results obtained revealed the importance of coordinating information, either as a specific skill or as a distinct task. Implications for current selection procedures for air traffic controllers and pilots are considered. Recommendations are also presented regarding the use and development of computerized measures for personnel selection in general.

[1] Part of the work presented in this paper was conducted while the author was developing her PhD at Cranfield University, UK, under the supervision of Dr. John Harris. The author appreciates the collaboration of Earl Hunt, University of Washington, and the materials provided. Research supported by the Fundação para a Ciência e a Tecnologia Lisbon, Portugal. Requests for reprints should be sent to Teresa C. D'Oliveira, ISPA, Rua Jardim do Tabaco, 34, 1149-041 Lisbon, Portugal or Teresa.Oliveira@ispa.pt. An initial version of this paper was presented at the 25th Conference of the European Association for Aviation Psychology (16 to 20 September 2002), held in Warsaw, Poland.

Requests for reprints should be sent to Beverly Laughead, FAA Academy, AMA-502-OU, P.O. Box 25082, Oklahoma City, OK 73125.

Introduction

The idea that spatial ability constitutes a separate domain in human intelligence is probably one of the most uncontroversial positions in the literature (e.g., Boer, 1991; Pellegrino & Kail, 1982). However, the panorama changes drastically when one considers spatial ability itself. Spatial ability has been defined in very different ways with several authors speaking of several spatial abilities and not just one (e.g., Lohman, Pellegrino, Alderton, and Regian, 1987). Those same spatial abilities have been labeled in a variety of ways. Confusion also is present in the literature regarding the immeasurable diversity of spatial ability tests. Eliot and Hauptman (1981) claimed that spatial ability has been referred to in such a variety of ways that it is difficult to be precise about the meaning of the terms.

In spite of this apparent disagreement, the predictive importance of spatial ability in several occupations is well established. Smith (1964) presented the most detailed review on the predictive value of spatial ability.The author presented most initial studies, some dating to the 1920's, where it was concluded that spatial ability made a positive contribution to the performance prediction of several technical training courses, namely mechanics in the Royal Air Force, technical courses of engineering drawing, engineering apprentices, mathematics courses in college, among others. The author also considered the recommendations of the United States Employment Service regarding the occupations that required high levels of spatial ability. In 1957, and as presented by Smith (1964), occupations such as architects, architectural draughtsman, structural draughtsman, aeronautical draughtsman, mechanical draughtsman, industrial designers, machinery and tool designers, sculptors, among others were thought to require performers with very high levels of spatial ability. Smith (1964) argued that several classes of engineering and draughtsmanship are emphasized in this list. McGee (1979) subsequently contended that four job categories – engineering, science, drafting, and designing – accounted for nearly 85% of all jobs listed by the 1957 United States Employment Service.

Today the list of occupations could be enlarged to include others, namely in the aviation domain. Carretta (1987) considered that spatial ability is required for a great variety of jobs in the military domain, especially for pilots. A similar case has been put forward by EUROCONTROL (1996, 2001) in what concerns air traffic controllers (ATCs) while reviewing the selection procedures adopted in many European countries[2].

Integrated in a NATO Aircrew Selection Working Group, Carretta, Rodgers, and Hansen (1993) developed a project that attempted to identify the abilities required for successful performance in piloting fighter aircraft and the instruments to be used in selection procedures that could assess those critical abilities. The most important characteristics identified were in descending order,

[2] The literature makes no distinction between military and civil contexts regarding the predictive value of spatial ability. It is as if the occupations under study are of much greater importance than the context, when discussing the predictive value of spatial ability.

situational awareness, spatial orientation, time-sharing, aggressiveness, divided attention, psychomotor coordination, perceptual speed, selective attention, and visualization. As for the tests proposed to evaluate these characteristics, five out of 21 tests were measures intended to evaluate spatial orientation and visualization.

The US Air Force has an extensive battery of written tests for the selection of pilots – The Air Force Officer Qualifying Test (AFOQT). A detailed description of the AFOQT can be found in Carretta and Ree (1994). This multiple-aptitude battery is composed of 16 paper-and-pencil tests that measure general intelligence (g) and five specific factors: verbal, quantitative, spatial, perceptual speed, and aircrew interest/aptitude[3]. Usually a composite score is calculated and used in the selection procedure. That same composite score also is used as a predictor in validation studies.

Hunter and Burke (1994) also described how the visual domain has been crucial for pilots. This occupation requires individuals to make quick and accurate comparisons, to identify objects embedded in other objects, to establish locational relationships between objects or self and objects, imagining how something will look after changes, etc.

As for ATCs, the spatial domain has also been considered as representing crucial abilities for successful performance. Several concepts are used to refer to the same characteristic of the ATC's job: mental picture, mental model, mental imagery, and more recently, situation awareness. Regardless of the concepts employed, all descriptions seem to emphasize a continuously changing mental representation of the vast amount of information controllers receive.

The development and maintenance of this mental representation is dependent upon a spatial component. Air traffic control takes place in a three-dimensional space (Hopkin, 1995). Controllers have well-defined spatial boundaries within which their responsibility lies. Spatial relations between aircraft, ground, weather patterns, winds, etc are considered (Endsley & Rodgers, 1994). The ability to transpose two-dimensional map information into three-dimensional air space visualization is involved (Isaac, 1995).

On the other hand, the importance of spatial abilities in air traffic control seems to be recognized worldwide (e.g., Isaac & Ruitenberg, 1999; Wing, 1991). The EUROCONTROL survey of the selection procedures adopted for ATCs illustrated this tendency as it was concluded that 71% of the countries involved in

[3] The specific tests included in this battery are: Verbal analogies, arithmetic reasoning, reading comprehension, data interpretation word knowledge, math knowledge, mechanical comprehension, electrical maze, scale reading, instrument comprehension, block counting, table reading, aviation information, rotated blocks, general science and hidden figures (Carretta & Ree, 1994).

[4] Countries included in the survey: Austria, Belgium, Bulgaria, Czech Republic, Denmark, Eurocontrol (Maastricht UAC), Finland, France, Germany, Hungary, Ireland, Italy, Lithuania, Malta, Netherlands, Norway, Poland, Portugal, Romania, Spain, Sweden, Switzerland and United Kingdom.

the survey evaluated spatial abilities[4].

Recently the literature on spatial abilities has identified a new spatial ability: dynamic spatial ability. This new ability is defined as the ability to reason about movement and its identification and evaluation are intrinsically related with the use of computers in personnel selection (Pellegrino & Hunt, 1989, 1991).

According to Pellegrino and Hunt (1989), the ability to deal with moving elements and dynamic spatial relations is separate from the abilities associated with reasoning about static spatial information, which are assessed by conventional paper-and-pencil tests.

Several occupations can be considered where the ability to make decisions regarding spatial events that involve moving objects or representations of moving objects is necessary. Law, Pellegrino, Mitchell, Fisher, McDonald, and Hunt (1993) considered that air traffic controllers and pilots to be good examples of such occupations. ATCs have to consider data regarding relative velocities and position of numerous objects, assess that information, and integrate it. Pilots also are involved in a dynamic context with moving objects and also must integrate data provided from the onboard instruments, the controllers, and their own perspective.

Pellegrino, Hunt, Abate, and Farr (1987) emphasized that it would be important to analyze if this ability to represent and reason about dynamically changing spatial relations will actually better predict performance in jobs that seem to require this ability. The authors also claimed that no studies have examined the utility of this new ability as a predictor of performance. Pellegrino and Hunt (1991) suggested that dynamic spatial ability may be of great practical significance for predicting several visual spatial reasoning activities such as piloting and air traffic control.

When one considers the descriptions of pilot and ATC jobs, it is not only dynamic spatial information (e.g., winds, relations between aircraft) that is involved, but also verbal information, either aural or written, is included. Indeed, in most circumstances, what is relevant is not the different abilities that are involved, but the way these are combined and used. In other words, it is how one coordinates different types of important information. Examples of the involvement of such a skill are presented in several features of the jobs. In what concerns ATCs, flight strips are of crucial importance for an accurate representation of the airspace under control. Pilots receive verbal instructions from controllers during take off, in flight, and in arrival.

Barbarino's (1995) description of the ATC job is quite illustrative of this aspect. The author claimed that the job "largely depends on the cognitive capacity of the human to create a clear, stable two and often three-dimensional picture of the airspace and aircraft within it from aural and visual data" (p. 1262).

Therefore, the ability to coordinate dynamic spatial and verbal information is also of practical importance for these occupations. Such ability was defined by Yee, Hunt, and Pellegrino (1991) and its predictive value has yet to be analyzed.

This study analyzed the potential predictive value of static and dynamic spatial ability and the ability to coordinate information when predicting training results of air traffic controllers and pilots. Two hypotheses were formulated:

H_1: Is dynamic spatial ability a significantly better predictor of training results than static spatial ability?

H_2: Is the ability to coordinate information a significantly better predictor of training results than static spatial ability?

Method

Participants

A total of 86 volunteers (10 female and 76 male) were recruited from two specific professional areas: air traffic control and piloting. Participants were all trainees/recruits in the Portuguese Air Force military training courses of ATC and piloting (Mean age was 22.32 with SD 4.78). Data was collected at the beginning of the training courses.

Instruments

Paper-and-pencil spatial ability tests. Table 1 presents the nine paper-and-pencil administered[5]

Table 1

Test	Author
MacQuarrie – Copying Subtest	MacQuarrie (1925)
MacQuarrie – Blocks Subtest	MacQuarrie (1925)
MacQuarrie – Pursuit Subtest	MacQuarrie (1925)
PMA - Spatial test	Thurstone (1947)
Figures Rotation	Yela (1967)
DAT - Spatial Relations	Bennett, Seashore & Wesman (1947)
Spatial Reasoning	Almeida (1992)
GATB 7	Boss, Cardinet, Maire & Muller (1963)
GATB 12	Boss, Cardinet, Maire & Muller (1963)

[5] With the exception of Yela (1967) and Almeida (1992), complete references for all paper-and-pencil tests, publishers, and commercial availability can be found in Eliot and Smith (1981).

List of paper-and-pencil tests administered and their authors

A main concern underlying the choice of paper-and-pencil tests was that these measures should be in use by Portuguese organizations. The idea was that traditional measures included in this study could actually be included in any personnel selection battery used in Portugal and therefore would not require a study for potential adaptation to the Portuguese population.

Computer administered tests. Dynamic spatial tests are still experimental and were mainly developed for Pellegrino and Hunt's studies. The relative arrival time task used in this study is a marker of dynamic spatial ability (Law, Pellegrino and Hunt, 1993). In this task subjects observed a computer controlled display, containing a black and a white object (referred to as targets in task description). Each object moved horizontally towards its own vertical "wall" line. The objects traveled horizontally across the screen over a period of four seconds, and then disappeared. The task was to determine which "target" would arrive first at its respective "wall" assuming that they continued to move at the same speed. Speed discrimination difficulty (three levels), path length traveled by the winning object (two levels), and colour of the winning object (two levels) were crossed over trials. Decisions required the observer to judge not only how far each target was from its wall when it disappeared from the screen, but also how fast the objects were travelling relative to each other. Feedback was always provided, including errors and responses longer than nine seconds. A detailed description of the task can be found in Yee, Hunt, and Pellegrino (1991).

Two tasks were used in order to evaluate the ability to coordinate information (CI). Both tasks included two components: a visual component that corresponded to the dynamic spatial ability task previously described and a verbal component that regarded a sentence presented in the lower part of the screen describing a possible outcome of the visual component. Participants had to observe the dynamic visual situation and then determine whether the sentence was a true or false description of the situation. Difficulty of the verbal component was manipulated by varying grammatical complexity. Complementary information regarding the verbal component can be found in Yee et al. (1991). Tasks differed in the amount of information to coordinate, with a simple version and a more complex type (e.g., targets did not always start from the same point and therefore the distance traveled would be different).

All computer administered tests presented two performance indicators: accuracy and reaction time.

Criterion Measures

For both ATC and pilot recruits, training performance results were used as criterion measures. Such results are usually expressed in a 0/20 scale with values higher than 10 representing a pass mark.

Procedure

Groups of participants were constituted according to their training classes.

The following administration order was used: paper-and-pencil tests, dynamic spatial ability test, simple task of coordinating information, and complex task of information coordination. Some general instructions regarding the testing session and the variety of tests being administered were given at the beginning of the session. At the end of the session, participants were debriefed and the main objective of the study was discussed. Criterion measures were obtained through the administrative services after participants concluded their training courses.

Results

Table 2 presents means, standard deviations for the entire sample and by occupation.

Table 2

Mean and standard deviations for entire sample and by occupation in each test

Measure	Entire Sample		ATCs		Pilots	
	M	SD	M	SD	M	SD
1.Copying	44.56	15.85	38.16	13.87	47.03	15.97
2.Blocks	12.29	3.84	10.38	3.61	13.03	3.69
3.Pursuit	27.08	7.07	27.08	6.24	27.08	7.41
4.PMA	40.10	9.47	35.33	9.74	41.95	8.76
5.F.Rotation	10.74	4.26	10.33	4.38	10.90	4.24
6.DAT	69.84	14.93	65.75	14.23	71.42	15.00
7.SR	22.78	5.05	21.38	5.66	23.32	4.73
8.GATB 7	27.95	5.87	26.50	6.33	28.52	5.63
9.GATB 12	24.86	5.36	23.92	6.51	25.23	4.85
10.Acc.D.Spatial	75.45	12.06	82.86	7.03	72.58	12.42
11.R.Time D.Spatial	1437.22	381.66	1538.44	403.57	1398.04	368.74
12.Acc.Simple CI	69.39	13.35	76.48	11.71	66.65	13.01
13.R.Time Simple CI	2199.01	362.02	2264.64	399.43	2173.60	346.58
14.Acc.Complex CI	60.95	9.46	66.75	8.50	58.70	8.89
15.R.Time Complex CI	2479.97	565.24	2734.30	469.01	2381.52	571.80

Table 3 presents the correlation matrix of all measures used and the reliability results.

Table 3

Correlation matrix of all measures administered

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	r_{xx}
1.Copying	-	.31	.34	.31	.20	.26	.24	.53	.45	-.17	-.11	-.06	.01	-.15	-.14	-[a]
2.Blocks		-	.24	.40	.24	.41	.33	.33	.38	-.00	-.16	-.06	.05	.05	.05	-[a]
3.Pursuit			-	.29	.07	.20	-.11	.30	.24	-.05	-.21	-.09	-.03	-.05	-.09	-[a]
4.PMA				-	.42	.52	.40	.31	.37	.07	-.15	-.03	-.01	.11	-.05	.94
5.F.Rotation					-	.51	.32	.34	.35	.18	-.17	.17	-.03	.38	.01	.82
6.DAT						-	.53	.35	.48	.25	-.06	.14	.14	.22	.13	.93
7.SR							-	.24	.28	.20	-.06	.21	.04	.29	.03	.85
8.GATB 7								-	.59	-.07	-.20	-.09	-.09	-.04	-.07	.90
9.GATB 12									-	.07	-.19	-.07	-.11	-.13	-.14	.86
10.Acc.D.Spatial										-	.42	.31	.20	.46	.28	.88
11.R.Time D.Spatial											-	.20	.41	.18	.24	.99
12.Acc.Simple CI												-	.28	.70	.36	.91
13 R.Time Simple CI													-	.24	.65	.97
14.Acc.Complex CI														-	.44	.89
15.R.Time Complex CI																.99

1 -Copying; 2 – Blocks; 3 – Pursuit; 4 –PMA; 5 – Fig.Rotation; 6 – DAT; 7 – SR; 8 – GATB 7; 9 – GATB 12; 10 – Accuracy Dynamic spatial; 11 – Reaction Time Dynamic Spatial; 12 -Accuracy Simple CI; 13 – R.Time Simple CI; 14 -Accuracy Complex CI; 13 – R.Time Complex CI.

r_{xx} – Cronbach'alpha

[a] – Tests that require another form of reliability evaluation.

The procedure recommended by Yee et al. (1991) and followed by Ramos, Heil and Manning (2001) was adopted in order to estimate values for the coordinating skill or trait. Such procedure implies that when one intends to demonstrate that a coordinating trait exists, three measures have to be used:

a. trait a (e.g., dynamic spatial task)
b. trait b (e.g., verbal task)
c. trait ab (e.g., combined task)

Two steps should then be followed: (1) Calculate the regression for the prediction of trait ab (e.g. combined task) using traits a and b (e.g. dynamic spatial task and verbal task) as predictors; (tables 4 and 5) (2) Calculate the difference between the predicted and actual trait ab (i.e., combined task). This residual score corresponds to the coordinating skill. This procedure was followed separately both for accuracy and reaction time thus resulting in two additional performance indicators: accuracy of coordination and reaction time of coordination.

Table 4

Summary of multiple regression results for calculating accuracy of coordinating ability (N=86)

Variable	B	SE B	β	p	R^2	p
Acc.Dynamic Spatial	.193	.063	.247	.003		
Acc.Simple CI	.420	.057	.593	.000	.508	.000

Table 5

Summary of multiple regression results for calculating reaction time of coordinating ability (N=86)

Variable	B	SE B	β	p	R^2	p
R.Time Dynamic Spatial	-.061	.133	-.041	.648		
R.Time Simple CI	1.053	.140	.675	.000	.435	.000

Results were analyzed in two different phases. Initially, a forward stepwise multiple regression was conducted using all possible predictors of training results. Results obtained are presented in table 6.

Table 6

Summary of forward stepwise multiple regression for measures administered in training results (N=86)

Variable	B	SE B	β	p	R^2	ΔR^2
Step 1						
Acc.Complex CI	.039	.007	.486	.000	.237*	.237*
Step 2						
Acc.Complex CI	.031	.008	.387	.000		
R.Time Coordinating	.000	.000	.214	.044	.273*	.036*
Step 3						
Acc.Complex CI	.019	.010	.236	.080		
R.Time Coordination	.000	.000	.215	.041		
Acc.Simple CI	.013	.007	.224	.076	.300*	.027

Table 6 - continued

Variable	B	SE B	β	p	R^2	ΔR^2
Step 4						
Acc.Complex CI	.019	.010	.235	.076		
R.Time Coordination	.001	.000	.642	.006		
Acc.Simple CI	.015	.007	.262	.037		
R.Time Complex CI	-.001	.000	-.479	.039	.336*	.036*
Step 5						
Acc.Complex CI	.019	.010	.235	.077		
R.Time Coordination	.001	.000	.643	.006		
Acc.Simple CI	.015	.007	.264	.036		
R.Time Complex CI	-.001	.000	-.473	.041		
Pursuit	.010	.010	.093	.305	.345*	.008
Step 6						
Acc.Complex CI	.022	.011	.270	.048		
R.Time Coordination	.001	.000	.612	.009		
Acc.Simple CI	.013	.007	.231	.72		
R.Time Complex CI	-.001	.000	-.437	.61		
Pursuit	.012	.010	.112	.226		
Blocks	-.022	.019	-.111	.247	.356*	.011
Step 7						
Acc.Complex CI	.020	.011	.247	.069		
R.Time Coordination	.001	.000	.641	.006		
Acc.Simple CI	.013	.007	.224	.078		
R.Time Complex CI	-.001	.000	-.462	.047		
Pursuit	.012	.010	.115	.210		
Blocks	-.034	.021	-.171	.099		
DAT	.007	.005	.152	.132	.374*	.018

Results of the stepwise multiple regression suggested that seven measures are significant predictors of training results: Accuracy and Reaction Time of Complex CI, Reaction time of Coordination, Accuracy of Simple CI, Pursuit, Blocks and DAT.As the adopted regression method tests the unique variance of each predictor, it is not possible to assess the relevance of any particular domain of spatial ability (either static or dynamic) or the relevance of the ability to coordinate information.

In order to test the hypotheses, two hierarchical multiple regressions were conducted. In both regressions, step 1 included all the traditional measures of static spatial ability. In order to test H_1, accuracy and reaction time of the dynamic spatial task were added in the first hierarchical multiple regression as a second step. Results can be observed in table 7.

Table 7

Summary of hierarchical multiple regression for predictors static spatial ability and dynamic spatial ability (N=86)

Model	R	R^2	ΔR^2	ΔF	df1	df2	p
1	.297	.088	.088	.815	9	76	.604
2	.328	.107	.019	.803	2	74	.452

*Significant at p< .05

Model1: Copying, Blocks, Pursuit, PMA, Fig. Rotation, DAT, SR, GATB 7, GATB 12

Model2: Copying, Blocks, Pursuit, PMA, Fig. Rotation, DAT, SR, GATB 7, GATB 12, Accuracy dynamic spatial, Reaction time dynamic spatial

Results obtained suggested that both static (R^2=.088, n.s) and dynamic spatial ability (R^2=.107, n.s) are not significant predictors of ATCs and pilots training results, therefore infirming H_1.

In order to test H_2, accuracy and reaction time of both complex and simple CI and accuracy and reaction time of coordinating skill were included as predictors in step 2. Results obtained can be observed in table 8.

Table 8

Summary of hierarchical multiple regression for predictors static spatial ability and the ability to coordinate information (N=86)

Model	R	R^2	ΔR^2	ΔF	df1	df2	p
1	.297	.088	.088	.815	9	76	.604
2	.634	.402	.314	6.122	6	70	.000*

*Significant at p< .05

Model1: Copying, Blocks, Pursuit, PMA, Fig. Rotation, DAT, SR, GATB 7, GATB 12

Model2: Copying, Blocks, Pursuit, PMA, Fig. Rotation, DAT, SR, GATB 7, GATB 12, Accuracy simple CI, Reaction time simple CI, Accuracy complex CI, Reaction Time complex CI, Accuracy Coordination, Reaction Time Coordination

Results suggested that measures included in step 2 (i.e., measures evaluating the ability to coordinate information) are significant predictors of training results of ATCs and pilots (R^2=.402, p<.05) and that these measures are significantly better predictors of training results than static spatial ability ($F(6,70)$=6.122, p>.000), thus confirming H_2.

Discussion

Recent developments in the spatial ability literature have identified a new spatial ability: dynamic spatial ability. This ability is thought to be particularly important in occupations such as air traffic control and piloting. At the same time, these occupations also exhibit another interesting characteristic: both required the occupant to make decisions after considering different types of information such as verbal and dynamic spatial information. The relevance of these abilities for the prediction of training results of ATCs and pilots was addressed in this study. In particular, this research attempted to analyze if both dynamic spatial ability and the ability to coordinate information would be significantly better predictors of training results than static spatial ability, therefore contributing to potential improvements in current selection procedures adopted for these occupations.

An initial stepwise multiple regression suggested that the best predictors of training results for ATCs and pilots were accuracy for complex and simple CI, reaction time for complex CI and reaction time for the coordination skill or trait. Results from the hierarchical multiple regressions also seem to confirm this initial result, i.e., significant predictors of training results are mainly associated with the ability to coordinate information.

All in all, results in this research appear to highlight the importance of integrating and coordinating both spatial and verbal information to successful training results. Contrary to the literature, traditional forms of evaluating spatial ability (i.e., paper-and-pencil tests) did not reveal themselves as significant predictors of training results. Results from the hierarchical multiple regression also seem to point in this direction. In fact, in order to present statistical significance the spatial component has to have dynamic features and be associated with verbal information.

Several issues may be considered regarding this study. First, as it has been suggested in the literature, spatial ability is an important predictor of training results in ATCs and pilots. However, no distinction is made in the literature regarding the predictive value of the different domains or areas that can be identified in the spatial domain. Therefore, it could be expected that all measures of spatial ability would be significant predictors of training results in some way. Results in this study suggested that this is not the case. Further research should address this issue and try to evaluate the predictive relevance of different static spatial abilities. When considering the literature, namely the work developed by Pellegrino and his colleagues, spatial ability is a significant predictor when considered along with other types of abilities. The efficiency with which different types of information are combined and a decision is made seems to be more important than individual abilities. Evaluating and responding to a spatial situation requires more than individual abilities; a global response is more than the aggregation of distinct components. A general decision is the result of coordination and integration of information. Therefore, selection procedures should actually mirror this job characteristic. Results from the second hierarchical multiple regression suggested that when this characteristic is incorporated in selection procedures, significantly better prediction is achieved.

Another important issue to be addressed is the diversity of methods that can be used in personnel selection and the advantages associated with the introduction of computers into these procedures. The main reason underlying the use of computers in training programs is the increased fidelity in the way work situations are portrayed (i.e., higher fidelity of scenarios) without compromising safety.The introduction of computers in personnel selection in the aviation domain is not a new idea and it has been put forward by several studies. Main advantages associated with the use of computers in personnel selection are: time required for administration and scoring procedures is reduced, increased reliability of scoring procedures, standardization of administration procedures, etc (Bartram & Bayliss, 1984; Burke & Normand, 1987; Hunt & Pellegrino, 1985; Skinner & Pakula, 1986; among others). That is, advantages are mainly associated with a more efficient management of the selection process. Disadvantages have also been suggested: financial investment required (not only for initial setup, but also for continuous system upgrading), staff training, and in the case of measures such as simulations, the time required for participants to gain enough practice with the system. What seldom is considered in the literature are the benefits resulting from the introduction of computers in personnel selection in terms of content, that is, in terms of what is evaluated. Fleishman (1988) considered that computers could introduce substantial innovations in psychological evaluation, namely it could lead to the identification of distinct human abilities. The predictive importance of a coordination skill for training results of ATCs and pilots suggested in this study illustrates this issue. A review and appraisal of the implications of computerized psychological assessment in terms of how human abilities are today conceptualized is now overdue.

Several limitations should be considered when analysing results of this study. Although participants involved in this study were starting their military training, range restriction issues may have applied. In particular, it is known that static spatial measures are used in the selection battery administered to pilot candidates. However, the same is not true for ATCs. Results regarding the relevance of static spatial ability should therefore be interpreted with care. Further research should address potential differences between ATCs and pilots in regard to spatial ability. When considering mean results for ATCs and pilots presented in table 2, it is possible to identify a pattern of results: in what concerns static measures, pilots usually have better results than ATCs. However, when considering computer-administered measures, ATCs respond with higher accuracy and a slower reaction time than pilots. Although no specific analyses were conducted to evaluate statistical significance of these differences due to sample size issues, further research should explore the existence of different patterns of results, namely in terms of response strategy. On the other hand, it also would be interesting to analyse in detail issues related with mental rotation processes in both jobs. When such psychological processes are considered, different types of mental rotation processes seem to be involved in these occupations. In air traffic control, for example, a typical mental rotation problem would be when a target of an aircraft is displayed on radar with a heading 180°, the controller has to consider the pilot perspective and give instructions considering a "mirror image" (i.e., the controller's right is the pilot's left). In pilots, mental rotation issues involve both psychological and physical processes, i.e., when mental rotation issues are considered in pilots, usually they are associated with physical rotation of the aircraft. In other words, in pilots, mental rotation issues always involve a physical rotation of the self (unsuccessful outcomes of these processes result in spatial disorientation issues). In ATCs mental rotation issues never involve this characteristic, as rotation is only a psychological process. Although this difference seems clear when considering job descriptions of both occupations, it has not been fully addressed by the literature.

Finally, it is worth mentioning the important effects that may arise in terms of face validity from the use of computerized measures. By portraying work situations in a more reliable way or by simulating specific job tasks, computerized measures may promote the selection process within the current jobholders and contribute to a better understanding of job characteristics within job candidates.

References

Almeida, L. S. (1992). *Bateria de provas de raciocínio diferencial.* Braga: Universidade do Minho.

Barbarino, M. (1995). Task and job description the core of human resources management in EATCHIP. In R. S. Jensen & L. A. Rakovan (Eds.). *Proceedings of the Eighth International Symposium on Aviation Psychology*, Vol. 2, (pp.1259-1264). Columbus OH: Ohio State University.

Bartram, D. & Bayliss, R. (1984). Automated testing: past, present and future. *Journal of Occupational Psychology, 57*, 221-237.

Boer, L. C. (1991). Spatial ability and orientation of pilots. In R. Gal & A. D. Mangelsdorff (Eds.). *Handbook of Military Psychology* (pp. 103-114). Chichester: John Wiley & Sons.

Burke, M. J. & Normand, J. (1987). Computerized psychological testing: overview and critique. *Professional Psychology: research and practice, 18*, 42-51.

Carretta, T. R. (1987). *Spatial ability as a predictor of flight training performance.* (Report no. FHRL-TP-86-70) Brooks Air Force Base, Texas: Air Force Human Resources Laboratory,Air Force Systems Command.

Carretta, T. R. & Ree, M. J. (1994). Pilot-candidate selection method: sources of validity. *International Journal of Aviation Psychology, 4*, 103-117.

Carretta, T. R., Rodgers, M. N. & Hansen, I. (1993). *The Identification of Ability Requirements and Selection Instruments for Fighter Pilot Training.* (Report no. AL/HR-TP-1993-0016) Armstrong Laboratory, Brooks Air Force Base, Texas: Air Force Human Resources Laboratory,Air Force Systems Command.

Eliot, J. & Hauptman, A. (1981). Different dimensions of spatial ability. *Studies in Science Education, 8*, 45-66.

Eliot, J. & Smith, I. M. (1983). *An International Directory of Spatial Tests.* Windsor, UK: NFER-Nelson.

Endsley, M. C. Rodgers, M. D. (1994). *Situation Awareness Information Re-*

quirements for En Route Air Traffic Control. (Report no. DOT/FAA/AM-94/27). Washington D.C.: U. S. Department of Transportation, Federal Aviation Administration.

Eurocontrol (1996). *Guidelines for selection procedures and tests for ab initio trainee controllers.* Brussels: Eurocontrol.

Eurocontrol (2001). *Guidelines for selection procedures and tests for ab initio trainee controllers.* Brussels: Eurocontrol.

Fleishman, E. A. (1988). Some new frontiers in personnel selection research. *Personnel Psychology, 41*, 679-701.

Hopkin, V. D. (1995). *Human Factors in Air Traffic Control.* London: Taylor & Francis.

Hunt, E. & Pellegrino, J. W. (1985). Using interactive computing to expand intelligence testing: a critique and prospectus. *Intelligence, 12*, 77-100.

Hunter, D. R. & Burke, E. F. (1994). Predicting aircraft pilot-training success: a meta-analysis of published research. *International Journal of Aviation Psychology, 4*, 297-313.

Isaac, A. R. (1995). Short-term memory and advanced technology: the use of imagery in air traffic control. In N. Johnston, R. Fuller & N. McDonald (Eds.). *Aviation Psychology: training and selection* (Proceedings of the 21st conference of the European Association for Aviation Psychology, vol. 2, pp. 107-111). Aldershot: Avebury Aviation.

Isaac, A. R. & Ruitenberg, B. (1999). *Air Traffic Control: human performance factors.* Aldershot, UK: Ashgate.

Law, D. J., Pellegrino, J. W. & Hunt, E. B. (1993). Comparing the tortoise and the hare: gender differences and experience in dynamic spatial reasoning. *Psychological Science, 4*, 35-40.

Lohman, D. F., Pellegrino, J. W., Alderton, D. L. & Regian, J. W. (1987). Dimensions and components of individual differences in spatial abilities. In S. H. Irvine and S. E. Newstead (Eds.) *Intelligence and cognition: contemporary frames of reference* (pp. 253-312). Dordrecht, Netherlands: Martinus Nijhoff Publishers

McGee, M. G. (1979). Human spatial abilities: psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin, 86*, 889-918.

Pellegrino, J. W. & Hunt, E. B. (1989). Computer-controlled assessment of static and dynamic spatial reasoning. In R. F. Dillon & J. W. Pellegrino (Eds.). *Testing: theoretical and applied perspectives* (pp. 174-198). New York: Praeger.

Pellegrino, J. W. & Hunt, E. B. (1991). Cognitive models for understanding and assessing spatial abilities. In H. A. H. Rowe (Eds.). *Intelligence: reconceptualization and measurement* (pp. 203-225). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pellegrino, J. W., Hunt, E. B., Abate, R & Farr, S. (1987). A computer-based test battery for the assessment of static and dynamic spatial reasoning abilities. *Behaviour Methods, Instruments & Computers, 19*, 231-236

Pellegrino, J. W. & Kail, R. (1982). Process analyses of spatial aptitude. In R.J. Sternberg (Ed.). *Advances in the Psychology of Human Intelligence*, (vol. 1, pp. 311-365). Hillsdale, NJ: Lawrence Erlbaum.

Ramos, R. A., Heil, M. C. & Manning, C. A. (2001). *Documentation of validity for AT_SAT computerized test battery* (Report no. DOT/FAA/AM-01/6) Washington DC: U.S. Department of Transportation, Federal Aviation Administration.

Skinner, H. A. & Pakula, A. (1986). Challenge of computers in psychological assessment. *Professional Psychology: research and practice, 17*, 44-50.

Smith, I. M. (1964). *Spatial Ability: its educational and social significance.* London: University of London Press.

Wing, H. (1991). Selecting for air traffic control: the state of the art. In J. A. Wise, V. D. Hopkin & M. L. Smith (Eds.). *Automation and systems issues in air traffic control*, (NATO ASI series vol. F73, pp. 409-427), Berlin: Springer-Verlag.

Yee, P. L., Hunt, E. & Pellegrino, J. W. (1991). Coordinating cognitive information: task effects and individual differences in integration information from several sources. *Cognitive Psychology, 23*, 615-680.

Yela, M. (1967). *Rotação de Figuras Maciças – RFM*. Adaptação portuguesa de 1985. Lisboa: Secção de estudos de Testes da Cegoc-TEA.

International Journal of Applied Aviation Studies, 3, 2

A Study of Safety Culture at an Aviation Organization[1]

Manoj S. Patankar
Saint Louis University
Parks College of Engineering and Aviation
3450 Lindell Boulevard
Saint Louis, MO 63156-0907

Abstract

Safety culture research in aviation indicates that national, organizational, and professional cultures influence the overall safety culture in any given organization. Such research has been a result of the somewhat divergent developments in the fields of crew resource management and maintenance resource management. Therefore, different research groups, with somewhat different research methods, have been involved in the studies of flight operations personnel and maintenance personnel. This study is unique in that it presents an analysis of the extant safety culture in one aviation organization—using both flight operations personnel as well as maintenance personnel—and using a common questionnaire. The survey reported in this article is based on a survey of a stratified sample of flight operations personnel, maintenance personnel, and other employees to measure individual attitudes regarding safety and their opinions regarding the factors that may contribute toward their accident-free safety record. In the case of this particular partner company, the success factors were identified as follows: emphasis on compliance with the Standard Operating Procedures, collective commitment to safety, individual sense of responsibility toward safety, and the high level of employee-management trust.

[1] This study was made possible due to the proactive management at the partner company. It is the author

Requests for reprints should be sent to Beverly Laughead, FAA Academy, AMA-502-OU, P.O. Box 25082, Oklahoma City, OK 73125.

Introduction

In the past two decades, the related fields of Crew Resource Management (CRM) and Maintenance Resource Management (MRM) have grown somewhat independently.Although the early MRM programs were based on the extant CRM programs, both CRM and MRM programs have matured along separate tracks. Initially, both programs were based on the need to improve interpersonal communication, teamwork, awareness about human performance limitations, and understanding of certain organizational factors that may contribute toward accidents. Majority of the CRM research has been conducted by Professor Helmreich and his team from the University of Texas at Houston, and majority of the MRM research has been conducted by Professor Taylor and his team starting at University of Southern California and culminating in Santa Clara University. While Helmreich used a survey questionnaire called Cockpit Management Attitudes Questionnaire (and later Flight Management Attitudes Questionnaire), Taylor used a survey questionnaire called MRM/Technical Operations Questionnaire. Although some of the items in these questionnaires are similar, they have evolved significantly throughout the last 15-20 years of research. Helmreich's questionnaires were designed for pilots and Taylor's questionnaire was designed for maintenance personnel. As such, both researchers conducted their research among different populations using different survey questionnaires.

This article presents a study that used one questionnaire—called the Organizational Safety Culture Questionnaire (OSCQ)—to assess the safety attitudes and opinions among flight operations personnel, maintenance personnel, and other employees at one partner organization. The OSCQ is a composite of items that have been used in the flight operations, maintenance, and organizational culture domains, and a few additional items. The new items were used to assess perceptions regarding professionalism and specific factors that may contribute toward the organization's safety record.

This study, unlike other studies in CRM and MRM, was not conducted in conjunction with, or in response to, any training or intervention programs. This was simply an exploratory study to determine the factors that may contribute to accident-free safety record of the partner organization.

Literature Review

Safety Culture

Culture, in general, is accepted as a set of shared beliefs, norms, attitudes, and practices within a certain population (Pidgeon & O'Leary, 1995). When these parameters were studied with respect to safety attitudes and behaviors among pilots, Helmreich and Merritt (1998) classified culture, or rather safety culture, into three areas: national culture, organizational culture, and professional culture. Differences in national cultures are typically categorized in terms of collectivism or individualism (cf. Hofstede, 1984). Accordingly, western Europeans tend be more individualistic in their attitudes and behaviors compared to their Asian and Latin American counterparts. Differences in organizational cultures have been measured in terms of factors such as safety compliance, hazard communication practices, and employee-management trust (Ciavarelli & Figlock, 1997; and Taylor, 1995). Organizations with more positive safety cultures tend to have better communication among their employees, higher levels of assertiveness, and higher levels of employee-management trust. From the perspective of differences in professional cultures, Helmreich and Merritt (1998) reported that the pilots were more individualistic than the surgeons. Subsequently, Taylor and Patankar (1999) reported that mechanics were more individualistic than pilots. Taylor and Patankar categorized maintenance professionals according to the type of their licensure: either based on the U.K.-Civil Aviation Authority's (CAA's) system of Aircraft Maintenance Engineer or based on the U.S.-Federal Aviation Administration's (FAA's) system of Aircraft Mechanic. Taylor and Patankar discovered that the CAA-based system tends to be more collectivistic than the FAA-based system.

Measuring Instruments

Traditionally, survey questionnaires such as the Cockpit Management Attitudes Questionnaire (Helmreich, Foushee, Benson, & Roussini, 1986), Maintenance Resource Management/Technical Operations Questionnaire (Taylor, 1995), and Command Safety Assessment Questionnaire (Ciavarelli & Figlock, 1997) have been used to study the safety attitudes among either the flight operations personnel or the maintenance personnel.

The Cockpit Management Attitudes Questionnaire (CMAQ). The CMAQ was originally developed to measure pilot attitudes and opinions regarding safety in the cockpit. When Helmreich and Merritt first discovered that the responses received from Taiwan were significantly different from their predominantly American sample, they added items to specifically measure differences that may exist based on differences in national cultures. Thus, some of Hofstede's items were included in the CMAQ (later changed to Flight Management Attitudes Questionnaire—FMAQ) by Helmreich and Merritt. CMAQ/FMAQ, in its latest form contains 86 items and has been administered at numerous airlines worldwide (Helmreich & Merritt, 1998).

The Maintenance Resource Management/Technical Operations Questionnaire (MRM/TOQ). The MRM/TOQ was originally developed from the CMAQ, but has undergone several revisions, at times based on a series of factor analyses and at times to incorporate items related to interpersonal trust. In its latest form, the MRM/TOQ contains 15 items on a five-point agreement scale. Respondents are asked to express their degree of agreement in a series of statements. The MRM/TOQ has been administered globally to over 5000 aircraft maintenance personnel (Taylor and Thomas, 2001).

The Command Safety Assessment Questionnaire (CSAQ). The CSAQ, in its latest form, is a 62-item questionnaire (for civilian use) that incorporates "many of the measurable components of high-reliability organizations" and has been used extensively in the naval aviation domain (with 57 items). Responses to each item on this questionnaire are rated on a seven-point Likert-type scale. The results from CSAQ studies have been used by the United States Navy to provide feedback to the Navy's human factors Quality Management Board conducting a comprehensive study of aircraft accidents (Ciavarelli & Figlock, 1997).

Methodology

Goal

The goal of this study was to identify factors that contribute to the partner organization's accident-free safety record. A survey methodology was used to conduct this study.

Survey Questionnaires

In order to combine flight operations personnel, maintenance personnel, and other employees in the same study, using a common questionnaire, relevant questions from the CMAQ, MRM/TOQ, and the CSAQ were extracted. A few new questions aimed at testing the organizational attachment and the perception of professionalism between the flight operations personnel and the maintenance personnel were added.

A stratified sample consisting of a statistically meaningful representation of maintenance, flight operations, and other employees was selected. Within each professional group, all the participants received the Organizational Safety Culture Questionnaire (see Appendix) via the subject company's mail distribution system. The participants filled-out the survey and mailed them directly to the Principal Investigator at Saint Louis University (SLU). The research team at SLU then coded the survey data.

At the subject company, there are approximately 497 flight operations employees (required sample size = 217), 177 maintenance employees (required sample size = 123), and 48 other employees (required sample size = 44). The requisite sample sizes, per Gay and Airasian (2003, p. 113), for each group are indicated in the parentheses.

The data collected through the Organizational Safety Culture Questionnaire was analyzed as follows:

a. *Factor Analysis*: A factor analysis of the questionnaire items was performed to determine the factor groupings or scales with which each item may associate. Past research (Taylor, 1995; Ciavarelli & Figlock, 1997; and Helmreich et al., 1986), with similar items, indicates that scales such as trust, power distance, hazard communication, etc. have emerged. Since the present questionnaire is somewhat different from those used in the past research, additional or different scales were likely.

b. *Comparison Across Groups*: Mean scores on each of the scales derived from factor analysis were compared across professional groups. Analysis of Variance (ANOVA) tests were performed to test the significance of the differences that may exist across these groups.

Inclusion Criteria

The Organizational Safety Culture Questionnaire was sent to all the employees in flight operations, maintenance, and rest of the company. The distribution, response, and required sample size are summarized in Table 1. Considering the small sample size, all the responses were included in the analysis.

Table 1
Distribution, Responses, and Required Sample Sizes

Job Category	Surveys Distributed	Responses Received	Required Sample Size
Pilots	497	227 (46%)	217
Maintenance	177	124 (70%)	123
Other Employees	48	48 (100%)	44
Total	722	399 (55%)	196*

*The total required sample sized is based on total surveys distributed, not the sum of the preceding required sample sizes.

Results

Description of the Samples

Out of the 722 surveys distributed among maintenance, flight operations, and other employees, 399 surveys were returned. Seventy percent of maintenance personnel returned their surveys, 46% of flight operations personnel returned their surveys, and 100% of the other employees returned their surveys. In the response sample, 57% were flight crew personnel, 31% were maintenance personnel, and 12% were other employees.

Analysis of Responses to the Organizational Safety Culture Questionnaire

A factor analysis was conducted across the items in the Organizational Safety Culture Questionnaire. As a result of such analysis, eight scales emerged: pride in company, professionalism, safety, supervisor trust and safety, effects of my stress, need to speak-up, safety compliance, and hazard communication. These scales are consistent with those reported by Taylor (1995), Ciavarelli and Figlock (1997), and Helmreich et al. (1986).

Responses to questions regarding the perceptions of professional image of pilots and mechanics across the three groups indicate that both mechanics as well as other employees seem to think that pilots practice high standards of safety. While other employees seemed to have similar opinion about mechanics, the pilots did not think that the mechanics practiced high standards of safety.

Of the various scales identified in this study, only Pride in Company, Safety Opinions, and Supervisor Trust and Safety, are discussed in detail in this article because they are most relevant to the goal of this study. Each scale was tested for reliability using Cronbach's Alpha: Pride in Company tested at 0.88 (n=397), Safety Opinions tested at 0.79 (n=400), and Supervisor Trust and Safety tested at 0.85 (n=389).

Pride in Company

The following items were included in this category: "Working here is like being a part of a large family," "I am proud to work for this company," "I am aware of my company's mission, values, and core ideology," and "My company is the 'Best in the business'." Figure 1 shows the overall rating by the three employee groups on a scale of 1-5 (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree).

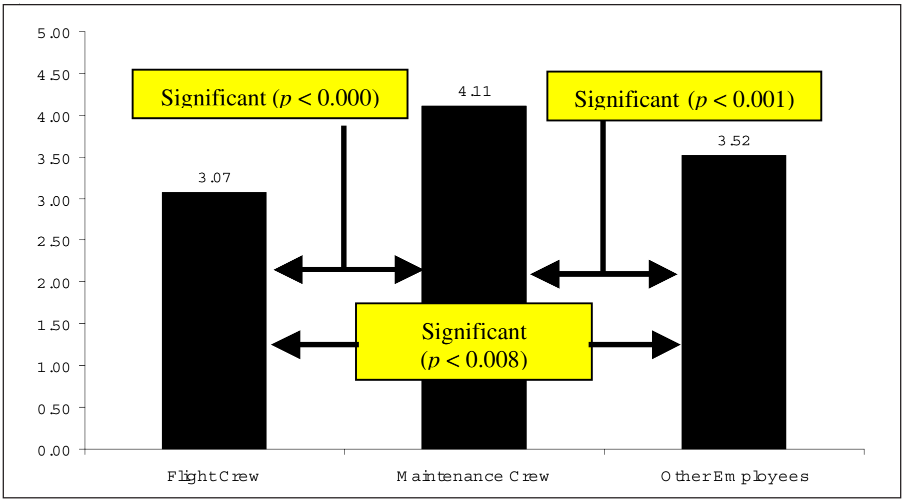

Figure 1. Comparison across the three employee groups on "Pride in Company"

Analysis of Variance (ANOVA) test on *Pride in Company* revealed that the differences in the three employee groups' ratings were statistically significant ($p<0.000$). Thus, there was a better than 99.99% probability that these ratings were not due to a random or chance error.

A further analysis of differences in ratings between groups, using the Scheffe test, revealed that there were statistically significant differences between flight crew and maintenance ($p<0.000$), between maintenance and other employees ($p<0.001$), and between flight crew and other employees ($p<0.008$). The maintenance crew is most proud to work for this company.

Safety Opinions

The following items were included in this category: "Safety in this organization is largely due to good luck" (scored on reverse scale), "Safety in this organization is largely due to adherence to the standard operating procedures," "Safety in this organization is largely due to our collective commitment to safety," "Safety in this organization is largely due to the efforts of a few key individuals" (scored on reverse scale), "Safety in this organization is largely due to positive changes resulting from our past experience with incidents and/or accidents," and "Safety is my responsibility." Figure 2 shows the overall rating by the three employee groups on a scale of 1-5 (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree).

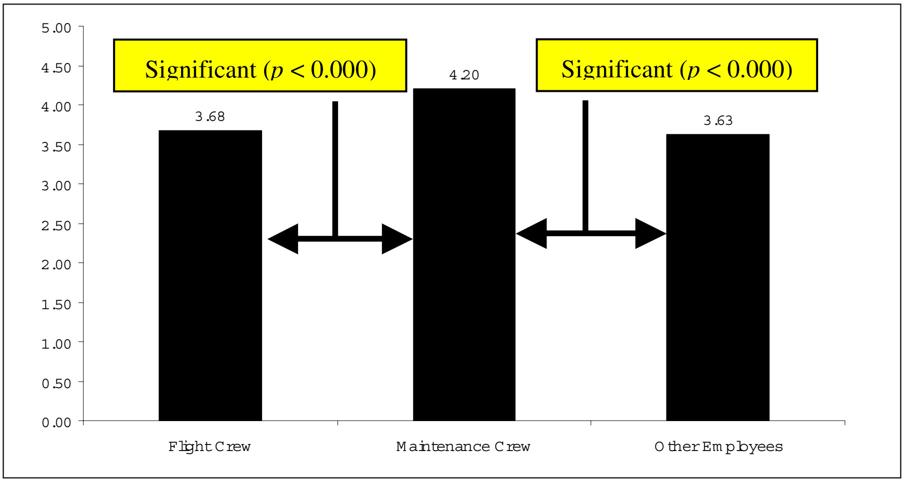

Figure 2. Comparison across the three employee groups on "Safety Opinions"

Analysis of Variance (ANOVA) test on *Safety* revealed that the differences in the three employee groups' ratings were statistically significant ($p<0.000$). Thus, there was a better than 99.99% probability that these ratings were not due to a random or chance error.

A further analysis of differences in ratings between groups, using the Scheffe test, revealed that there were statistically significant differences between flight crew and maintenance ($p<0.000$) and between maintenance and other employees ($p<0.000$). The maintenance crew rated the organizational safety at the highest level.

Supervisor Trust and Safety

The following items were included in this category: "I feel comfortable going to my supervisor's office to discuss problems or operational issues," "I know the proper channels to route questions regarding safety practices," "My supervisor can be trusted," "I am not comfortable reporting a safety violation because people in my organization would react negatively toward me" (scored on reverse scale), "My supervisor listens to me and cares about my concerns," "My supervisor protects confidential or sensitive information," "Within my organization, good communications flow exists up an down the organizational chain of command," and "My suggestions about safety would be acted upon if I expressed them to my supervisor." Figure 3 shows the overall rating by the three employee groups on a scale of 1-5 (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree).

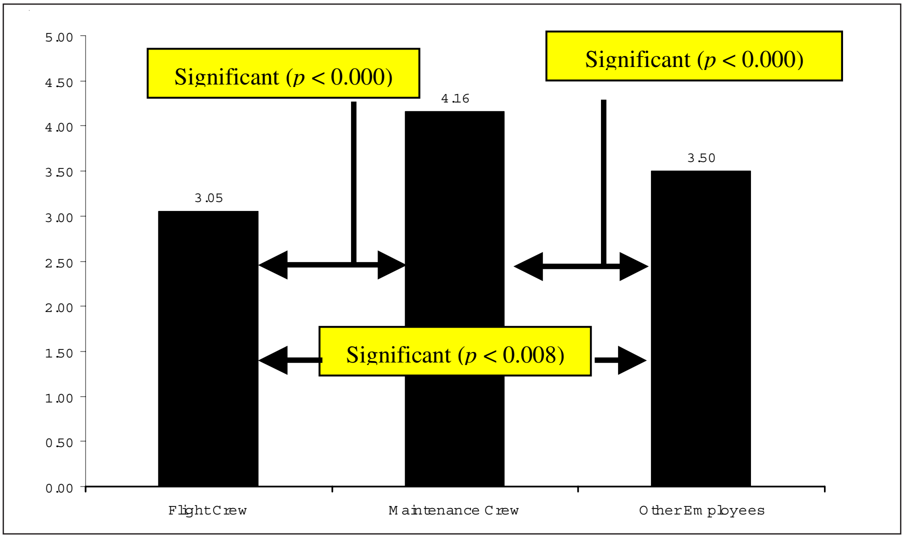

Figure 3: Comparison across the three employee groups on "Supervisor Trust and Safety"

Analysis of Variance (ANOVA) test on *Supervisor Trust and Safety* revealed that the differences in the three employee groups' ratings were statistically significant ($p<0.000$). Thus, there was a better than 99.99% probability that these ratings were not due to a random or chance error.

A further analysis of differences in ratings between groups, using the Scheffe test, revealed that there were statistically significant differences between flight crew and maintenance ($p<0.000$), between maintenance and other employees ($p<0.000$), and between flight crew and other employees ($p<0.008$). The maintenance crew rated the supervisor trust at the highest level.

Furthermore, it is extremely important to note that since the U.S. national average on this scale for maintenance personnel is 3.50/5.00 (n=2938, five companies) (Taylor & Thomas, 2001), the subject company's maintenance personnel have the highest level of supervisor trust in the industry.

Factors Contributing to the Company's Accident-free Safety Record

The ultimate goal of this study was to identify the characteristics that may contribute toward the accident-free safety record of the partner company. Responses to the following five items were analyzed: "Safety in this organization is largely due to good luck," "Safety in this organization is largely due to adherence to the standard operating procedures," "Safety in this organization is largely due to our collective commitment to safety," "Safety in this organization is largely due to the efforts of a few key individuals," and "Safety is my responsibility". Figures 4 - 8 show the overall rating by the three employee groups on a scale of 1-5 (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree).

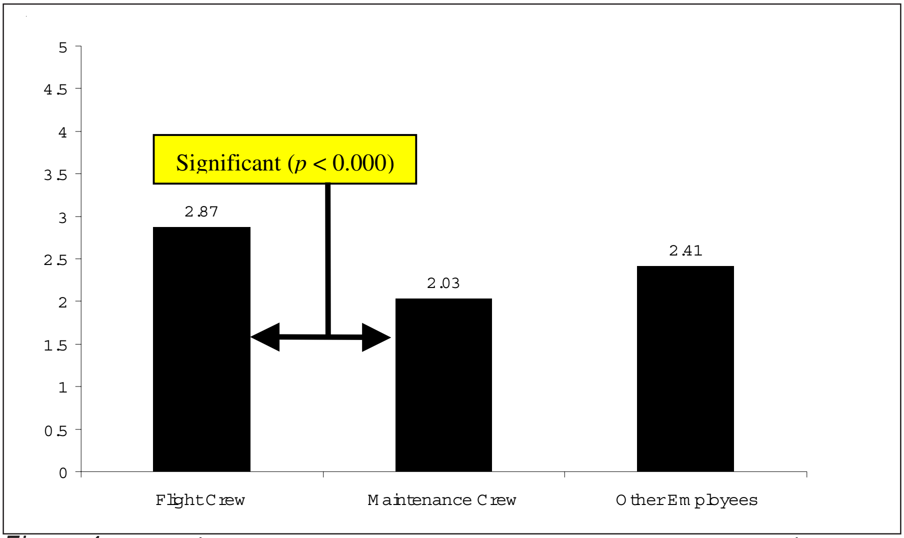

Figure 4. Comparison across the three employee groups on "Safety is due to luck"

Analysis of Variance (ANOVA) test on *Safety is due to Luck* revealed that the differences in the three employee groups' ratings were statistically significant ($p<0.000$). Thus, there was a better than 99.99% probability that these ratings were not due to a random or chance error.

A further analysis of differences in ratings between groups, using the Scheffe test, revealed that there were statistically significant differences between flight crew and maintenance ($p<0.000$) The flight crew seemed to think that safety was due to luck, more than the other employee groups.

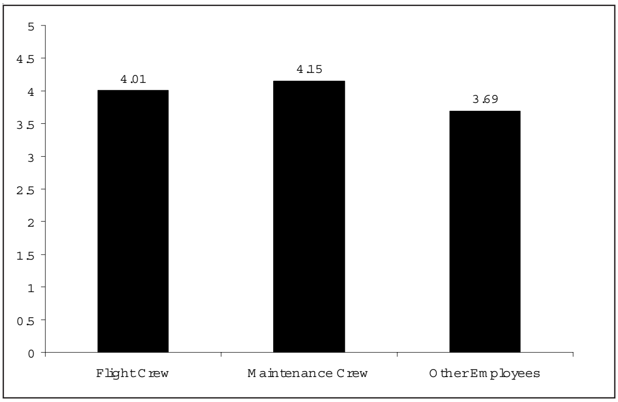

Figure 5. Comparison across the three employee groups on "Safety is due to Standard Operating Procedures"

Analysis of Variance (ANOVA) test on *Safety is due to Standard Operating Procedures* revealed that the differences in the three employee groups' ratings were statistically significant (p<0.022). Thus, there was a better than 97% probability that these ratings were not due to a random or chance error.

A further analysis of differences in ratings between groups, using the Scheffe test, revealed that there were statistically significant differences between maintenance and other employees (p<0.022) and flight crew and other employees (p<0.022). The flight crew and the maintenance personnel seemed to attribute safety to standard operating procedures, more than the other employees did.

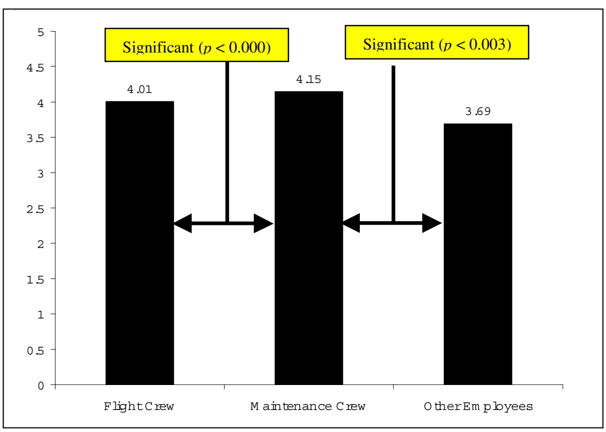

Figure 6. Comparison across the three employee groups on "Safety is due to collective commitment to safety"

Analysis of Variance (ANOVA) test on *Safety is Collective Commitment to Safety* revealed that the differences in the three employee groups' ratings were statistically significant (p<0.000). Thus, there was a better than 99.99% probability that these ratings were not due to a random or chance error.

A further analysis of differences in ratings between groups, using the Scheffe test, revealed that there were statistically significant differences between flight crew and maintenance ($p<0.000$) and maintenance and other employees ($p<0.003$). The maintenance personnel seemed to attribute safety to a collective commitment, more than the other employee groups.

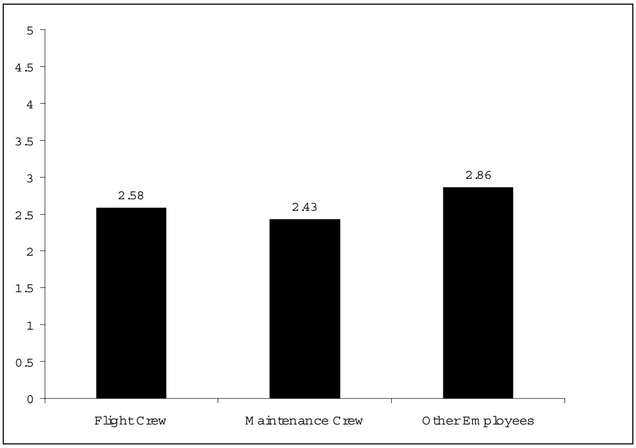

Figure 7. Comparison across the three employee groups on "Safety due to efforts of key individuals"

Analysis of Variance (ANOVA) test on *Safety is due to Efforts of Key Individuals* revealed that the differences in the three employee groups' ratings were <u>not</u> statistically significant. It is safe to say that all employee groups disagreed with this item; therefore, they regarded safety as a team effort.

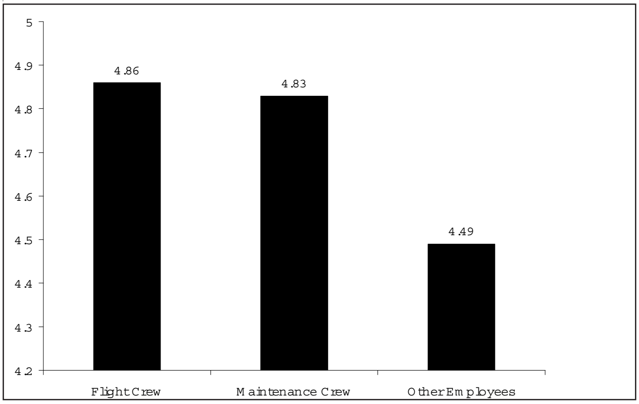

Figure 8. Comparison across the three employee groups on "Safety is my responsibility"

Analysis of Variance (ANOVA) test on *Safety is My Responsibility* revealed that the differences in the three employee groups' ratings were statistically significant ($p<0.001$). Thus, there was a better than 99.99% probability that these ratings were not due to a random or chance error.

A further analysis of differences in ratings between groups, using the Scheffe test, revealed that there were statistically significant differences between flight crew and other employees ($p<0.001$) and between maintenance and other employees ($p<0.005$). Both the flight crew and the maintenance personnel seemed to accept more responsibility for safety. Such ownership is consistent with their legal responsibilities derived from their respective FAA certificates. Also, it is normal for pilots to feel the most responsible for safety because they risk their own lives when safety is compromised.

Discussion

The results of this research indicate that the employees are proud to work for this company, they trust that their management will make decisions in the interest of safety, they emphasize consistent use of Standard Operating Procedures, and they think that safety is truly a result of their collective efforts. The flight crew and the maintenance crew also exhibit a very high sense of personal responsibility toward the safety of flight.

The most important element that distinguishes this company from most other airlines and repair stations studied by Taylor and Patankar (2001), however, is that the maintenance crew in this company exhibit a very high sense of trust in their management: they trust that their managers will make decisions in the interest of safety. Goglia, Patankar, and Taylor (2002) reported that up to about one-third of the mechanics in the United States do not trust that their managers will act in the interest of safety. Therefore, when the mechanic-to-management trust is high, it is highly commendable. Conversely, a lack of such trust is a strong indicator of serious organizational problems.

As more and more aviation companies strive to implement a confidential error-reporting program such as the Aviation Safety Action Program, the level of employee-management trust is likely to play a vital role in the success of such programs. Therefore, before trying to implement any error-reporting programs, it would be prudent to assess the organizational readiness to accept such programs. The organizational safety culture survey presented in this study would be effective in measuring the extant safety climate, especially on *Pride in Company*, *Safety Opinions*, and *Supervisor Trust and Safety.*

Conclusion

Strong emphasis on compliance with the Standard Operating Procedures, collective commitment to safety, individual sense of responsibility toward safety, and the high level of employee-management trust are the factors that contribute toward the accident-free safety record of the aviation organization reported in this article.

References

Ciavarelli, A., & Figlock, R. (1997, June). Organizational factors in Naval Aviation Accidents. Paper presented at *The Center for Risk Mitigation Conference*, University of California, Berkeley.Assessing Organizational Accident Risk.

Gay, L. R., & Airasian, P. W. (2003). *Educational Research: Competencies for analysis and application* (7th ed.). Merrill, Prentice Hall: Englewood Cliffs, NJ.

Goglia, J. Patankar, M., & Taylor, J. (2002). Lack of error mitigation tools: the weakest link in maintaining airworthiness? In *Proceedings of the 55th Annual International Air Safety Seminar*, November 4-7, 2002, Dublin, Ireland: Flight Safety Foundation, pp. 253-272.

Helmreich, R., & Merritt, A (1998). *Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences*. Aldershot, Hants: Ashgate Publishing Ltd.

Helmreich, R.L., Foushee, H.C., Benson, R., & Russini, R. (1986). Cockpit management attitudes: exploring the attitude-performance linkage. *Aviation, Space and Environmental Medicine*, 57, 1198-1200.

Hofstede, G. (1984). *Culture's consequences: International differences in work related values (Abridged Edition)*. Beverly Hills, CA: Sage.

Patankar, M., & Taylor, J. (2001). Analyses of organizational and individual factors leading to maintenance errors. In *Proceedings of the 2001 SAE World Aviation Congress & Exposition*. [SAE Technical Paper Number 2001-01-3005]. Seattle, WA.

Pidgeon, N., & O'Leary, M. (1995). Organisational safety culture and aviation practice. In N. McDonald, N. Johnston, and R. Fuller (Eds.), *Applications of psychology to the aviation system*, 47-52. Aldershot, U.K.: Ashgate Publishing Limited.

Taylor, J.C. (1995) Effects of Communication & Participation in Aviation Maintenance. In *Proceedings of the Eighth International Symposium on Aviation Psychology*. Columbus, Ohio, Ohio State University, pp. 472-477.

Taylor, J., & Patankar, M. (1999). Cultural Factors contributing to the successful implementation of macro human factors principles in aviation maintenance. In R. Jensen (Ed.) *Proceedings of the Tenth International Symposium on Aviation Psychology* at Ohio State University, Columbus, OH.

Taylor, J., & Patankar, M. (2001). Four generations of MRM: evolution of human error management programs in the United States. *Journal of Air Transportation World Wide 6*(2), 3-32. Omaha, NE: University of Nebraska at Omaha.

Taylor, J, & Thomas, R. (2001). Toward Measuring Safety Culture in Aviation Maintenance: The Structure of Trust and Professionalism. Technical Report. Santa Clara, CA: Santa Clara University.

Appendix

ORGANIZATIONAL SAFETY CULTURE QUESTIONNAIRE

I. BACKGROUND INFORMATION: Today's Date: ___/___/___

For Flight Crew:

1. Crew Position: Captain ____ FO _____ SO _____

2. Status: Line Pilot _____ Instructor _____
Check Pilot _____ Management _____

3. Base City or Station: ______

4. Fleet (A/C type and series): _______

5. Years of experience in this aircraft: ________

6. Years as a pilot in this airline:

7. Past Experience or Training: (# of years: fill in below)
Military: ______ (Years of service)
Civilian (Private School): ______ (Years of service)
Other Airline [Use Name (years) format]:
____

8. Gender: Male ______ Female _____

9. Year of birth: 19___

10. Citizenship:

For Maintenance Crew:

1. Job Title: ________________________________

2. Years in Maintenance at your company: _____

3. City or Station: ______

4. Present Shift: ______

5. Gender: Male _____ Female _____

6. Year of birth: 19___

7. Past Experience or Training: (# of years: fill in below)
Military: _____ (Years) Trade School: _____ (Years)
College: _____ (Years) Other Company [Use Name (years) format]:
____

8. Mark type of employment: Regular Employee _______ Contractor _______

9. Where do you work? Line Hangar QC Planning Shop Stores Engineering Appearance Other [Circle one]

For All Other Employees:

1. Job Title: ________________________________

2. Years in your department at your company: _____

3. City or Station: _______

4. Present Shift: _______

5. Gender: Male _____ Female _____

6. Year of birth: __________

7. Past Experience or Training: (# of years: fill in below)
Military: _____ (Years) Trade School: _____ (Years)
College: _____ (Years) Other Company [Use Name (years) format]:
____

8. Mark type of employment: Regular Employee _______ Contractor _______

9. Where do you work? Line Hangar QC Planning Shop Ramp Customer Service Store Engineering Appearance Other [Circle one]

FLIGHT/MAINTENANCE OPERATIONS ATTITUDES:

1	2	3	4	5
Strongly Disagree	Slightly Disagree	Neutral	Slightly Agree	Strongly Agree

Using the above scale, please circle the number that best describes your opinion.

1 2 3 4 5	1.	Our company's safety practices are consistent with the published corporate values and mission.	1 2 3 4 5	9.	Team members should avoid disagreeing with one another.
1 2 3 4 5	2.	Our company provides adequate safety equipment.	1 2 3 4 5	10.	Our internal hazard identification and management program is effective.
1 2 3 4 5	3.	I feel comfortable going to my supervisor's office to discuss problems or operational issues.	1 2 3 4 5	11.	It is important to avoid negative comments about the procedures and techniques of other team members.
1 2 3 4 5	4.	My colleagues use the appropriate safety equipment at all times.	1 2 3 4 5	12.	We should always provide both written and verbal turnover to the oncoming shift
1 2 3 4 5	5.	I am encouraged by my supervisors and coworkers to report any unsafe conditions I may observe.	1 2 3 4 5	13.	Start of shift team meetings are important for safety and for effective team management
1 2 3 4 5	6.	Our mechanics practice the highest maintenance standards.	1 2 3 4 5	14.	We receive adequate amount of safety training.
1 2 3 4 5	7.	Our drivers practice the highest standards of professionalism.	1 2 3 4 5	15.	Our ramp agents practice the highest standards of professionalism.
1 2 3 4 5	8.	The people that I work with comply with the company's Standard Operating Procedures/General Maintenance Manual/Published Company Procedures.	1 2 3 4 5	16.	Our pilots practice the highest standards of professionalism.

III. GENERAL SAFETY CULTURE: Questions for ALL Employees

1	2	3	4	5
Strongly Disagree	Slightly Disagree	Neutral	Slightly Agree	Strongly Agree

Using the above scale, please circle the number that best describes your opinion.

1 2 3 4 5	17.	Even when I am fatigued, I perform effectively during critical phases of work.	1 2 3 4 5	34.	My organization uses an internal audit and hazard reporting system to catch any problems that may lead to an accident.
1 2 3 4 5	18.	I know the proper channels to route questions regarding safety practices.	1 2 3 4 5	35.	My supervisor protects confidential or sensitive information.
1 2 3 4 5	19.	Having the trust and confidence of my coworkers is important.	1 2 3 4 5	36.	I am aware of my company's mission, values, and core ideology.
1 2 3 4 5	20.	A truly professional team member can leave personal problems behind when working.	1 2 3 4 5	37.	My company is the "Best in the Business."
1 2 3 4 5	21.	All employees should make the effort to foster open, honest, and sincere communication.	1 2 3 4 5	38.	I am adequately trained to conduct all of my job duties and responsibilities.
1 2 3 4 5	22.	My supervisor can be trusted.	1 2 3 4 5	39.	Within my organization, good communications flow exists up and down the organizational chain of command.
1 2 3 4 5	23.	A debriefing and critique of procedures and decisions after a significant task is completed is an important part of developing and maintaining effective coordination.	1 2 3 4 5	40.	Communication between pilots and mechanics should be encouraged.
1 2 3 4 5	24.	Personal problems can adversely affect my performance.	1 2 3 4 5	41.	Safety in this organization is largely due to good luck.
1 2 3 4 5	25.	My coworkers valued consistency between words and actions.	1 2 3 4 5	42.	Safety in this organization is largely due to adherence to the standard operating procedures

1 2 3 4 5	26.	Working here is like being a part of a large family.	1 2 3 4 5	43.	Safety in this organization is largely due to our collective commitment to safety.
1 2 3 4 5	27.	I am proud to work for this company.	1 2 3 4 5	44.	Safety in this organization is largely due to the efforts of a few key individuals.
1 2 3 4 5	28.	My organization closely monitors quality and corrects any deviations from established quality standards.	1 2 3 4 5	45.	Safety in this organization is largely due to positive changes resulting from our past experience with incidents and/or accidents.
1 2 3 4 5	29.	In my organization, deviations from published procedures are rare.	1 2 3 4 5	46.	Safety is my responsibility.
1 2 3 4 5	30.	I am not comfortable reporting a safety violation because people in my organization would react negatively toward me.	1 2 3 4 5	47.	I trust my subordinates to choose safety over performance, regardless of my presence.
1 2 3 4 5	31.	We are proactive in hazard identification and management.	1 2 3 4 5	48.	I trust my superiors to choose safety over performance.
1 2 3 4 5	32.	I am aware of a self-reporting system for incidents.	1 2 3 4 5	49.	My subordinates trust me.
1 2 3 4 5	33.	My supervisor listens to me and cares about my concerns.	1 2 3 4 5	50.	My suggestions about safety would be acted upon if I expressed them to my supervisor.

IInternational Journal of Applied Aviation Studies, 3, 2

The Effects of Leadership Style and Primary Task Workload on Team Performance and Follower Satisfaction

James P. Bliss
and
Corey K. Fallon
Old Dominion University
Psychology Department, MGB 244B
Old Dominion University
Norfolk, VA 23513
(757) 683-4222
jbliss@odu.edu

Abstract

Previous research has indicated that the majority of errors made in aircraft cockpits are the result of faulty information transfer among members of the flight crew. This study examined the problem by manipulating leadership style and studying its effects on performance and satisfaction under conditions of high and low workload. The participants in this study were Old Dominion graduate and undergraduate psychology students between the ages of 18 to 42. Forty-one dyads reacted to true and false alarms while performing a primary psychomotor task. After the dyads completed the experiment, the researchers administered a survey to collect data about copilots' ratings of worker satisfaction. Statistical analysis revealed that dyads in the participative leadership condition reacted to the alarms significantly more appropriately than those in the autocratic condition did, and that participants in the low workload group reset primary task gauges significantly faster than those in the high workload group. The post experimental survey indicated that, in general, copilots were more satisfied with participative leadership. Also, copilots in the low workload group reported significantly more successful team interaction. The survey and alarm reaction appropriateness results could warrant training pilots to exercise participative leadership in times of task ambiguity.

Requests for reprints should be sent to Beverly Laughead, FAA Academy, AMA-502-OU, P.O. Box 25082, Oklahoma City, OK 73125.

The Effects of Leadership Style and Primary Task Workload on Team Performance and Follower Satisfaction

In past research, psychologists have linked many of the causes of human error in flight crews to communication lapses in the cockpit. These breakdowns often reflect issues such as automation mistrust (Parasuraman & Riley, 1997) or interpersonal problems among members of the crew (Foushee, 1984). Some psychologists have stressed the importance of pilot leadership to resolve these issues and restore communication before error occurs. It was traditionally believed that leadership in the cockpit should be highly task oriented with very little two-way communication. However, several studies have shown that when a communication breakdown occurs, a leadership style that fosters two-way communication may enhance flight crew performance (Smith, 1979; Hines, 1998). The benefits are especially notable during situations of high mental workload where the breakdown in information transfer is most common (Billings, 1991). In addition, a communicative leadership style may also create a more pleasant work environment.

Because of increased automation in the cockpit, modern flight crews do more observing than flying (Koonce, 1999). Such monitoring, however, is not an easy task. Flight crews must be cognizant of many environmental aspects, including the aircraft's location, altitude, and velocity, the surrounding terrain, current and projected weather conditions, and system states, including the proper settings and functioning of equipment (Endsley, 1999). Problems often arise because information about one or more of these factors is not effectively shared within the flight crew. Such communication breakdowns can occur because of social issues such as lack of assertiveness or interpersonal conflicts among members (Jentsch & Smith-Jentsch, 2001; Wiegmann & Shappell, 2001). Communication disruptions may also occur between members of the flight crew and the aircraft's hardware and software interface elements. Such problems are especially common when members of a flight crew must interact with unreliable flight equipment (Wiegmann & Shappell, 2001).

Regardless of the source of disruption, a breakdown in information transfer often leads to disorder and ambiguity in the cockpit, which in turn renders flight crews susceptible to human error. One study showed that failures in information transfer account for more than 50% of all errors inside the cockpit (Wickens, 1995). Furthermore, researchers have suggested that these problems are most likely to occur during periods of high mental workload (Bliss & Dunn, 2000; Foushee, 1984).

According to Lysaght, Hill, Dick, Plamondon, Linton, Wierwille, Zaklad, Bittner, and Wherry (1989), when the number of task elements rises and the complexity within each element increases, the load of information placed on each flight crew member's cognitive processes also increases. Human error is likely to occur in the cockpit during periods of high mental workload because each crewmember's mental resources are limited as to the amount of information they can process (Lysaght et al., 1989). If mental workload becomes too great, members of the flight crew will no longer process information at the desired rate and task performance will degrade. In the case of alarm research, empirical studies have shown that reaction time during ambiguous situations is significantly slower under conditions of high mental workload (Bliss & Dunn, 2000).

Many researchers believe that training crew members to optimize team interdependence and communication is the best way to prevent disorder and restore the lines of communication following crises (Sexton & Helmreich, 2000; Wickens 1995). However, communication among members of the crew can take many forms and may affect performance in many ways. The style of communication present in the cockpit is often dictated by airline policy and directed by the leadership style of the pilot. By studying different pilot leadership styles, psychologists can determine the type of communication that works best during situations of disorder and ambiguity, and can make recommendations about how to manage human error (Helmreich, Merritt, & Wilhelm, 1999).

In the early 1970s, Vroom and Yetton developed a theory of leadership in work environments that proposed two main styles of leadership: participative and autocratic (Chemers, 2000). These forms of leadership are collectively included within Normative Decision Theory.According to that theory, the participative leader promotes two-way communication and allows subordinates to have equal weight in the decision making process. In contrast, the autocratic leadership style involves very little interaction between the leader and subordinates. An autocratic leader remains focused on the team's tasks and makes all of the team decisions without much input from other team members (Chemers, 2000).

Normative Decision Theory states that the effectiveness of each leadership style is dependent on the situation (Chemers, 2000). The autocratic style is more effective in situations where the task is clear and all members of the team have a high degree of control over their responsibilities. In these situations very little communication is needed in the decision making process, allowing the autocratic leader to make quick decisions (Vroom, 2000). The participative style is more effective in situations where problems with information transfer have contributed to an ambiguous environment and the leader does not have much control over the situation, yet must make very important decisions (Vroom, 2000). The increased two-way communication characteristic of this style helps to clarify the situation and creates a supportive work environment that may increase worker satisfaction and commitment to the team (Chemers, 2000). One potential drawback of the participative style is that the increased amount of participation may result in a much slower decision making process. Overall, studies testing the validity of Normative Decision Theory have been favorable (Chemers, 2000; Vroom & Jago, 1978). In one study, teams that properly instituted the use of this theory increased their number of successful decisions by 20% (Vroom & Jago, 1978). Also, although the effectiveness of leadership styles may be situation specific, several studies indicate that subordinates generally prefer the participative style regardless of the situation (Field, 1990; Foels, Driskell, Mullen & Salas, 2000).

Early research concerning leadership style focused on industrial work environments. However, recent research in this area has been conducted in nontraditional work environments. One work environment that has received particular attention is the aircraft cockpit. The cockpit differs from industrial working environments in both setting and size. Flight crews are often much smaller than traditional work groups, typically consisting of only a pilot and copilot. Also, cockpit dyads tend to be more cohesive. They function more as a team to perform highly specialized tasks.

Several studies of leadership in the cockpit have indicated that the type of communication a participative leader exhibits may be more effective for minimizing errors made in the cockpit (Foushee, 1982, 1984; Prince, Chidester, Bowers, & Cannon-Bowers, 1992; Sexton & Helmreich, 2000). In one study Nicholas and Penwell (1995) examined leadership styles in several work environments including aviation. They found that the task management style of effective leaders is predominantly participative and a strict autocratic style "does not lend itself to effective operation of complex, technical machinery" (Nicholas & Penwell, 1995, p.70).

One reason for using the participative style of leadership in the cockpit may be to diagnose and understand communication disruptions when they occur.An element that often causes disorder and uncertainty for flight crews is unreliable avionics (Bliss, 1997). This includes the settings and functioning of avionics, including alarm systems (Endsley, 1999). A breakdown in information transfer may occur if communication between the human flight crew and the avionics is disrupted due to an unpredictable or unreliable piece of flight equipment (Wiegmann & Shappell, 2001). According to Muir's Machine Trust Theory, the uncertainty and disorder created by unreliable technology weakens the operator's trust (Lee & Moray, 1992). Anecdotal evidence has shown the applicability of Muir's theory, revealing that alarm system mistrust in the cockpit can decrease the frequency, speed, and appropriateness of pilot responses (Bliss, 2003b).

Mistrust of automated systems in multi-crew cockpits can be further aggravated by psychosocial problems between flight crewmembers (Wiegmann & Shappell, 2001). For example, the first officer's lack of assertiveness or an interpersonal conflict between members can create further disruptions in information transfer and more potential for human error (Jentsch & Smith-Jentsch, 2001).

If practiced in situations of machine mistrust, the participative leadership style may serve two useful purposes. First, the increased communication may help bring clarity and order to a situation rendered unclear by machine unpredictability. Second, by promoting two-way communication and assertiveness, participative leadership may prevent psychosocial problems from further confusing an already chaotic situation. Such benefits are likely to become increasingly important as further advances in aviation technology occur (Wickens, 1995).

Goal of this research

Several studies in the past have examined the effects of participative and autocratic leadership on flight crew performance (Foushee, 1982, 1984; Prince et al., 1992; Sexton & Helmreich, 2000). However, no existing research has examined the relationship between leadership style and team performance during periods of machine mistrust. The purpose of this study was to examine the effects of leadership style and mental workload on performance of a primary psychomotor task and a secondary alarm response task by dyads.

As part of this investigation, we examined the effects of leadership style and mental workload on subjective ratings of leader satisfaction by followers. Leader satisfaction has been studied in several work environments (Chemers, 2000; Dubin, 1965; Field, 1990; Foels et al., 2000). However, to date there have been no investigations in situations similar to a cockpit.

The two independent variables in the study were leadership style and mental workload. Leadership style had two levels, autocratic and participative. The workload independent variable also had two levels, high workload (three computer subtasks), and low workload (two computer subtasks).

We examined the effects of the independent variables on the speed and appropriateness of alarm reactions, and on primary task performances. We also examined the effects of workload and satisfaction on several dependent measures of copilot satisfaction. These variables included self report ratings by copilots on measures of overall team effectiveness, drive to participate in the secondary task, perception of pilot cooperation, perception of pilot accuracy, the degree of enjoyment experienced working with the pilot, and the degree to which the pilot was perceived to be a likable person.

Hypotheses

We hypothesized an interaction between leadership style and workload on reaction appropriateness to the secondary alarm task. We believed that participatively led dyads in the high workload group would yield significantly more appropriate reactions to true and false alarms compared to dyads that were autocratically led in this group. This hypothesis is consistent with Normative Decision Theory, Muir's Machine Trust theory, several studies of cockpit leadership, and research concerning mental workload (Chemers, 2000; Foushee, 1982, 1984; Lee & Moray, 1992; Lysaght et al., 1989; Prince, Chidester, Bowers, & Cannon-Bowers, 1992; Sexton & Helmreich, 2000).

We also hypothesized main effects for workload and leadership style on alarm reaction time. We believed that dyads following the autocratic leadership style would react significantly faster to alarms than those following the participative leadership style. We also believed that dyads in the low workload group would react significantly faster to the alarms than those in the high workload group. These predictions are consistent with the Normative Decision Theory and studies of alarm reactivity in conditions of varying mental workload (Bliss & Dunn, 2000, Chemers, 2000; Dubin, 1965; Vroom, 2000).

We hypothesized a main effect for leadership on the copilot's worker satisfaction variables. We predicted that copilots would be significantly more satisfied with the participative leadership style than with the autocratic style regardless of the workload condition. This hypothesis is consistent with Normative Decision Theory and several studies of worker satisfaction (Chemers, 2000; Dubin, 1965; Field, 1990; Foels, et al., 2000).

Method

Design

This experiment was conducted using a 2 X 2 mixed design. The between-groups independent variable, workload, was determined by the number of subtasks (2 or 3) on the primary task workstation monitor.The within-groups variable, leadership style (autocratic or participative) was counterbalanced across two experimental sessions, so that dyads followed each leadership style while performing the primary tasks and reacting to alarms.

Participants

Convenience sampling resulted in a sample consisting of volunteer graduate and undergraduate psychology students from Old Dominion University. Participants were offered class credit and a monetary bonus ($10.00 for the team with the best experimental task performance) as an incentive for participating.

A power analysis revealed that 40 dyads would yield an experimental power of 0.80 at p = .05 (Cohen, 1988). The researchers obtained results from 40 dyads comprised of one pilot and one copilot per dyad. Of the 40 dyads, 21 were in the low mental workload group and 19 were in the high mental workload group. Attempts were made to equalize the sex balance across dyads: In the low workload group, nine dyads included both sexes (in five cases the male was the pilot; in four cases the female was the pilot), seven included both males, and five included both females. In the high workload group, ten dyads included a male and a female (in four cases the male was the pilot, in six cases the female was the pilot), five included both males, and four included both females. In addition, the within subjects variable (leadership style) was counterbalanced so that half of the dyads followed the autocratic strategy in the first session and the other half were administered the participative strategy first. Participants ranged from 18 to 42 years old, with the average age of approximately 23 years.

Two important considerations are how well the participants knew each other prior to participating, and how often they normally participate in teamed activities. To determine these considerations, we surveyed team members before they participated. We found that 65% of team members did not know each other at all before participating, 16% of team members barely knew each other, 5% knew each other fairly well, 2% knew each other quite well, and 12% knew each other extremely well. We also found that 59% of participants regularly participated in team activities outside of the experiment, whereas 41% did not.

Materials

The laboratory space used for this study was designed to simulate a cockpit environment. The space consisted of two primary task IBM-compatible computer workstations positioned side by side. A Macintosh computer hosting an alarm response task separated the workstations.

The Multi-Attribute Task (MAT) battery (Comstock & Arnegard, 1992) was hosted on the workstations as the primary experimental task. The MAT battery includes three separate subtasks (dual-axis compensatory tracking, gauge monitoring, and resource allocation) that require cognitive and spatial abilities, simulating aircraft piloting demands. Depending on primary task workload condition assignment, teams were presented either two concurrent subtasks (tracking and monitoring) or three (tracking, monitoring, and resource allocation). Within a workload condition, each 14" VGA primary task computer monitor displayed the same number of subtasks. The MAT battery has been used frequently as a primary and secondary task in alarm trust research, and is thoroughly described in other sources (Bliss, 2003a; Getty, Swets, Pickett, & Gonthier, 1995).

Participants performed the MAT tasks using the mice and keyboards to make responses. At the same time, auditory and visual alarms were presented on a Macintosh PowerPC personal computer with a 14" VGA monitor located between the MAT workstations. The auditory stimulus was the fire bell digitized from a Boeing 757/767 simulator, followed by a male voice announcing "temperature, temperature." The stimulus lasted four seconds. The visual stimulus was a yellow panel with the word "TEMPERATURE!" on it. When an alarm occurred, participants were to determine whether the MAT TEMP1 and TEMP2 gauges had both fluctuated out of tolerance within 15 seconds prior to the alarm. If so, the alarm was true. Participants were then to reset the MAT gauges and press the F12 key (marked "R" for "RESPOND") on the Macintosh keyboard, in that order. If the gauges had fluctuated out of tolerance more than 15 seconds prior to the alarm, the alarm was false. Participants were then to press the F9 key (marked "C" for "CANCEL") on the Macintosh keyboard and resume the MAT tasks. Alarm stimuli were presented on a variable interval time schedule (average interstimulus interval was 70 seconds), at 60 dB(A) (ambient sound was 45 dB(A). Other than the timing of the gauge fluctuations, there was no way to tell true alarms from false alarms, as they looked and sounded identical.

Alarm reaction measures included speed to react (in seconds), appropriateness of reactions (how often participants appropriately responded to true alarms and canceled false alarms), and response frequency (the percentage of alarms participants responded to within each experimental session).

In addition to the computer tasks, participants also completed two forms. Prior to participation, they completed a background information form including questions about hearing and vision abilities, computer experience and their degree of familiarity with their teammate. After participation, all copilots completed a four-page opinion questionnaire. The questionnaire included thirteen Likert scale items to assess aspects of copilot satisfaction such as how much the copilot liked the pilot, how much the copilot enjoyed working with the pilot, the copilot's impression of the team's effectiveness, the copilot's impression of the pilot's cooperation level, and how driven the copilot was to perform the experimental tasks. Calculation of the questionnaire's internal reliability indicated that the driven variable was not consistent with the five other measures and was therefore excluded from future analyses. Excluding the driven variable, Cronbach's Alpha revealed that the opinion questionnaire had an internal consistency reliability of Alpha = .81.

Procedure

Upon arrival at the research laboratory, the two participants were required to complete an informed consent form and the background questionnaire. Their team was then randomly assigned to the high or low workload condition, and participants were randomly assigned to the role of pilot or copilot.

Once the experimenter assigned the team to a condition and the team members to their roles, the experimenter presented the primary task experimental instructions. To ensure comprehension, the instructions were delivered in writing and were also read aloud to the participants. Teams in the high workload condition received instructions for three primary subtasks: tracking (keeping a ball aligned with a set of crosshairs), gauge monitoring (monitoring moving gauges for excessive fluctuations), and resource allocation (keeping fluid present in a number of on-screen containers). Those in the low workload condition received instructions for only the tracking and gauge monitoring tasks.

The experimenter explained how to respond to true alarms and cancel false alarms. The experimenter explained that during the actual experiment, alarms would be occurring on the Macintosh computer.The teams had to determine if the alarm was true or false and react accordingly.The experimenter explained that alarms related to the monitoring subtask of the MAT. If an alarm was true, it meant that the two TEMP gauges had fluctuated out of tolerance within 15 seconds prior to the alarm.To properly respond to a true alarm both participants had to reset their gauges by pressing the F1 and F3 keys on the IBM keyboard, and then press the F12 key on the Mac keyboard, marked "R" for respond. The experimenter explained that if the team's decision to respond was correct, the team would receive an increase of 1.5 points to their alarm score. If the decision were not correct, 1.5 points would be taken away from their alarm score. If an alarm was false, the two gauges had fluctuated more than 15 seconds prior to the alarm. When a false alarm occurred, the alarm was to be canceled. One of the teammates had to press the F9 key on the Mac keyboard, marked "C" for CANCEL. Once again, if the team was correct on their decision to cancel, 1.5 points was added to the team score. If the team is incorrect 1.5 points was deducted from the team score. Gauge fluctuations occurred in a congruent fashion on both primary task computers.

After the instructions, the teams were allowed to practice the primary subtasks in isolation for one minute, and were then given a 10-minute practice session with the MAT and the alarms. The experimenter stressed that speed and accuracy were equally important on all tasks. The team began the experiment with an alarm score of 50 points.

Participants were instructed to follow a specific alarm reaction decision-making strategy for each session. The strategy followed either an autocratic or participative leadership style. In the autocratic condition, the participant assigned to the role of pilot was solely responsible for making the alarm response decisions. He or she could not rely on the copilot for any opinions or advice about how to respond. Once the pilot made a decision, he or she commanded the copilot to press the appropriate response key on the alarm response keyboard. The copilot was instructed not to offer information unless specifically asked by the pilot. The participant assigned to this role was also instructed to follow the pilot's commands without question. The autocratic strategy was designed to limit communication and participation.

In the participative condition, both participants had an equal weight in the alarm response decision process. They were permitted to communicate freely with each other throughout the session and when an alarm occurred they were required to reach a mutual verbal agreement before reacting. This leadership strategy was designed to promote communication and participation. Failure to follow the prescribed decision making strategies resulted in a point deduction for the team.

After answering questions, the experimenter started the first 10-minute task session (either autocratic or participative). Once the first session was complete, the experimenter administered leadership strategy instructions for the second session. As a within groups variable, leadership style was manipulated and counterbalanced across sessions so that each team received both sets of instructions. After the team completed both experimental sessions, the experimenter administered the opinion questionnaire to gather subjective ratings of copilot satisfaction. Participants were then debriefed and dismissed.

Results

After calculating descriptive statistics to ensure that the dependent measures were distributed normally, the researchers calculated three 2 x 2 mixed ANOVAs to determine the effects of workload and leadership style on alarm reaction appropriateness and alarm reaction time.

The omnibus ANOVA for alarm reaction appropriateness showed no interaction between leadership style and workload, $p > .05$. However, the main effect for leadership style showed that the participative leadership style led to more appropriate alarm reactions than the autocratic style, $F(1,38) = 4.667$, $p = .037$, ?$^{??}$= .109 (see Figure 1). The main effect for workload was not significant, $p > .05$.

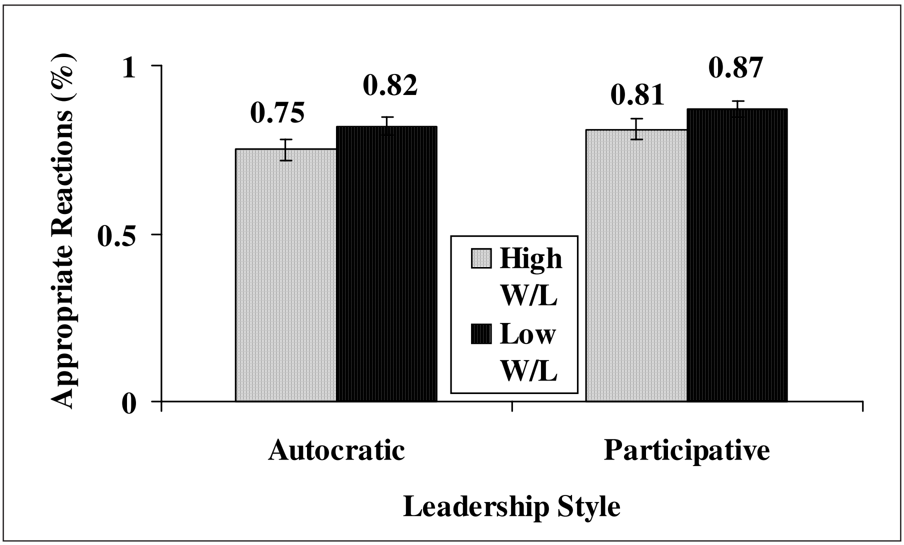

Figure 1. Alarm Response Appropriateness as a Function of Primary Task Workload and Leadership Style.

Next, the researchers computed an ANOVA to determine whether alarm reaction time changed as a function of leadership and workload. The interaction, main effect for workload, and main effect for leadership were not significant, $p > .05$.

We also computed mixed ANOVAs to determine the effects of leadership style and primary task workload on measures of primary task performance. Specifically, we tested whether MAT tracking error, monitor reset time, and resource allocation pump activation frequency varied as a function of leadership style or primary task workload. Of those three ANOVAs, the only significant effect we observed showed that monitor reset time was longer under conditions of high primary task workload, $F(1,33) = 10.472$, $p = .003$, partial □$^{□}$= .241

(See Figure 2). All other interactions and main effects were not significant, $p > .05$.

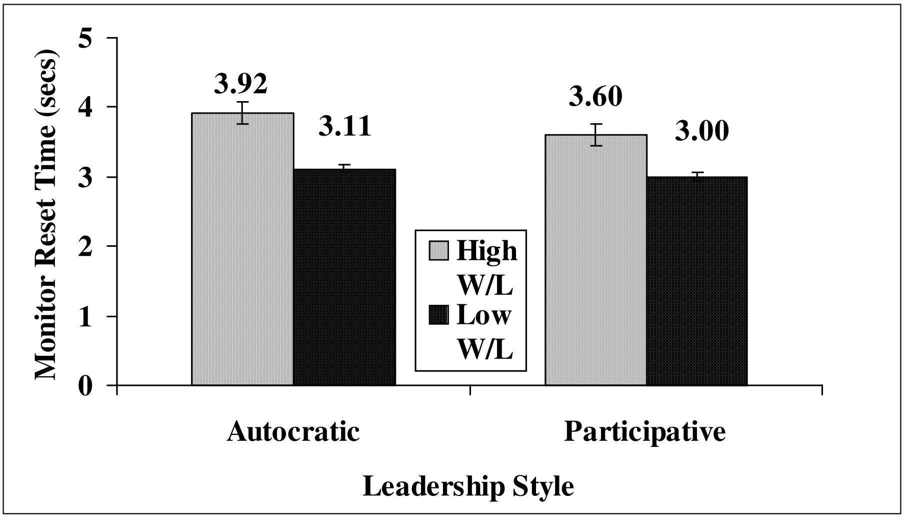

Figure 2. Gauge Monitor Reset Time as a Function of Primary Task Workload and Leadership Style.

Copilot Satisfaction

We computed a series of mixed ANOVAs to assess the effects of primary task workload and leadership style on three copilot satisfaction measures. ANOVAs were chosen to investigate the influence of leadership-workload interactions on the copilot satisfaction variables. The three variables chosen for analysis were the three most reliable dependent variables in the measure, as determined by a Cronbach's alpha test. The variables were the extent to which the team worked together successfully (Teamwork), the extent to which the copilot enjoyed working with the pilot (Enjoy) and the extent to which the copilot found the pilot to be a likable person (Likeability).

Researchers computed a 2 x 2 ANOVA to examine the effects of primary task workload and leadership style on teamwork. The test did not find a significant interaction $p > .05$. The test did, however, show significant main effects for leadership style, $F(1,78) = 23.85$, $p<.001$, partial □□= .23 and workload, $F(1,78) = 6.28$, $p =.014$ partial □□= .08. Copilots reported significantly more teamwork during the participative condition when compared to the autocratic condition and copilots in the low workload group reported significantly more teamwork than those in the high workload group (see Figure 3; note that due to inverse scaling, a lower number indicates a higher teamwork rating).

The 2 x 2 ANOVA for Copilot Enjoyment did not find a significant interaction, nor did the test find a main effect for workload, $p > .05$. Like Teamwork a significant main effect was found for leadership, $F(1,78) =24.05$, $p<.001$, □□= .24.

Copilots enjoyed working with their pilots significantly more during the participative condition than during the autocratic condition (see Figure 4; note that due to inverse scaling, a lower number indicates a higher enjoyment rating).

A third ANOVA was computed to examine the effects of primary task workload and leadership style on Likeability. The results followed the same pattern as the copilot enjoyment variable. Neither the interaction nor the main effect for workload were statistically significant, $p > .05$. However, a main effect for leadership style was significant, $F(1,78) = 22.94$, $p < .001$, ??² = .23. Copilots reported liking their pilot significantly more during the participative condition then during the autocratic session (see Figure 5; note that due to inverse scaling, a lower number indicates a higher likeability rating).

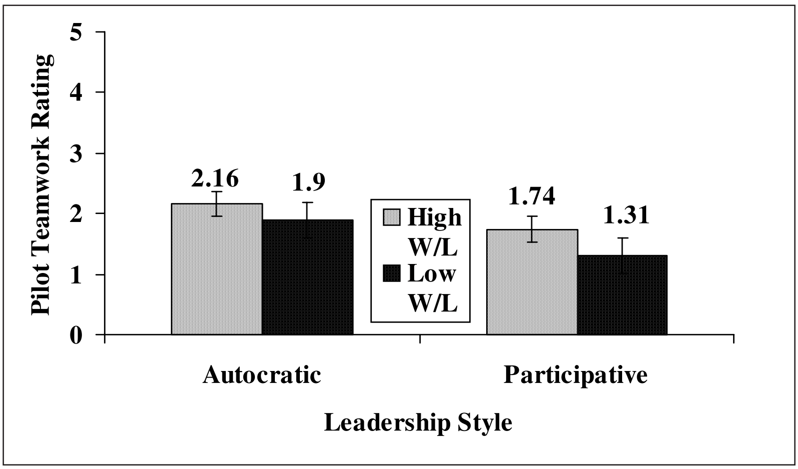

Figure 3. Copilot Perception of the Level of Teamwork as a Function of Primary Task Workload and Leadership Style.

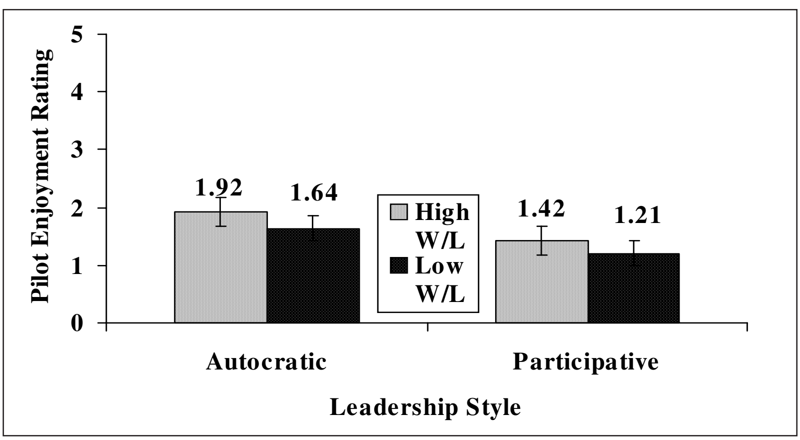

Figure 4. Copilot Perception of Pilot Enjoyment as a Function of Primary Task Workload and Leadership Style.

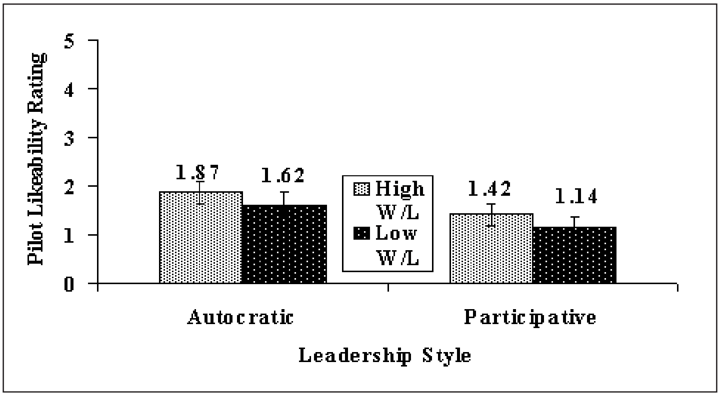

Figure 5. Copilot Perception of Pilot Likeability as a Function of Primary Task Workload and Leadership Style.

Discussion

Our results marginally supported the first hypothesis. Although we did not find an interaction between leadership style and workload for alarm reaction appropriateness, we did observe a main effect: improved alarm reaction appropriateness for participative leadership. This finding is consistent with theories extolling the benefits of participative leadership, such as Normative Decision Theory (Chemers, 2000; Foushee, 1982, 1984; Prince et al., 1992; Sexton & Helmreich, 2000). It is also consistent with recent research by Bliss (2003a) showing better performance by teams that were forced to have more social interaction. The current findings are important because they suggest that participative leadership can reduce human error in the cockpit during periods of machine mistrust. It is interesting that primary task workload did not interact with leadership style. However, because workload was a primary task variable, participants may have devoted separate cognitive resources to the primary and alarm reaction tasks, so that variability in one did not affect the other (Wickens, Sandry, & Vidulich, 1983).

Our findings did not support the second hypothesis. Although we did see an effect of workload on MAT gauge resetting time, there was no significant effect for alarm reaction time as a function of either leadership style or primary task workload. There may be several reasons for this.

It is possible that reaction time was not sensitive to our manipulation of leadership. To maintain experimental rigor, we assigned specific and well defined leadership roles to each team member in the participative condition. In an actual cockpit, such behaviors would likely occur more naturally. Furthermore, it was evident that even though our instructions were clear, there were participative teams that did not communicate much, and autocratic teams that did. Yet, as we observed from alarm reaction appropriateness and the copilot satisfaction questionnaire responses, participants definitely distinguished between the two leadership styles, preferring the participative style. To better study reaction time, other researchers may elect to observe leadership more naturally as it occurs; however, to do so would preclude using it as an experimental independent variable.

One advantage of specifying team duties before the experimental sessions was that participants might have exhibited less confusion during the actual task. In actual flight performance situations, pilots and copilots probably make leadership style decisions in real time, taking more time and perhaps leading to more errors. However, in this experiment, the participants did not have to devote energy to that decision, because it was made for them. As a result, perhaps there was less variability of reaction time within and across the leadership styles.

Another possibility was that the experimental tasks were not complex enough to reveal differences for reaction time (or response frequency). In an actual cockpit, there are a variety of actions that crews must perform following an alarm signal. However, in this experiment, the actions were comparatively straightforward: following an alarm, participants were to estimate whether gauges had fluctuated within 15 seconds prior to the alarm, and cancel or respond to the alarm accordingly. It is clear that reaction appropriateness was better when the leadership style was participative. However, this advantage may generalize to other performance variables only if the alarm reaction task is more challenging. Future researchers may elect to systematically increase the complexity of the alarm task to determine whether this might be true.

The third hypothesis was generally supported by the results. A main effect for leadership style was revealed for the three measures of copilot satisfaction. According to these results, copilots believed that during the participative condition their team exhibited significantly more teamwork, their pilot was a more likable partner and was more enjoyable to work with. These findings support decades of research on worker satisfaction and leadership that demonstrate a strong preference by the worker for participative leadership style in traditional work environments (Chemers, 2000; Dubin, 1965; Field, 1990; Foels et al., 2000). However, the current findings are unique because they indicate that leadership affects worker satisfaction within a more specialized team environment.

The results also suggest a possible link between copilot satisfaction and alarm reaction appropriateness. As team reaction appropriateness improves under participative leadership, copilot satisfaction increases. The low level of follower satisfaction found under autocratic leadership may have contributed to the poorer performance in this condition. Dissatisfied copilots may have created psychosocial problems within the team. For example, when under autocratic leadership copilots may have been less assertive during team interaction, further limiting communication. Some researchers have suggested that problems such as these can lead to human error in the cockpit (Wiegmann & Shappell, 2001; Jentsch & Smith-Jentsch, 2001). This potential relationship should be examined in future studies.

Analysis of the survey data also revealed an unpredicted finding. Copilots in the low workload group reported significantly higher levels of teamwork than those in the high workload group. One possibility for this finding is that during periods of high primary task workload, copilots were consumed by their individual tasks and felt less free to work together with their pilots. This explanation, however, is not supported by any degradation in alarm reaction performance between groups.

Future Research

We believe that the findings reported here have important practical implications for flight crew training programs. For example, trainers might be well advised to consider decision accuracy and copilot satisfaction by encouraging pilots to adopt a participative style of leadership in situations of task uncertainty.

The current study provides some theoretical support for the Normative Decision Theory by revealing that copilots are more satisfied with the participative leadership style, and that this satisfaction may be reflected in alarm reaction appropriateness. This study broadens the applicability of Normative Decision Theory by showing that it can be successfully applied to a specialized teamwork environment. However, there may be limitations to its applicability. For example, the researchers found that autocratic decisions were not made significantly faster than those in the participative condition were. Future research is recommended to continue to clarify the conditions where the theory applies and does not apply.

Limitations of this Study

As noted previously, a primary goal of this research was to maintain experimental control. To do so, we chose to test college students as they performed a contrived task in an experimental laboratory. Because the research reported here was conducted in a laboratory, its generalizability to operational environments is likely limited. Our decision to test college students may have led dyad members to interact with each other in ways that are not typical for flight crewmembers. For example, they may have occasionally chosen to disregard our instructions concerning the details of participative and autocratic leadership. In contrast, actual pilots would likely follow such instructions more closely.

The experimental tasks, although designed to exercise skills used in aviation, are comparatively simplistic. For example, the fear experienced by participants, and the stated consequences of poor performance do not match those faced by pilots, and the complexity of the MAT battery pales in comparison to some piloting demands.

Finally, the alarm reaction scenario assumed that participants were sufficiently motivated to react in realistic ways. This contrived paradigm was necessary for ethical reasons. However, the data presented here likely understate what is possible in an actual piloting situation.

Because of the above limitations, we recommend that future researchers attempt to replicate the current experiment using high fidelity flight simulators and aviators as participants. By doing so, it may be possible to study the influence of such realistic distracters as flight attendant calls, air traffic control calls, and cross-traffic on operator performance. Future research should also be conducted to confirm the validity of the questionnaire items used to measure copilot satisfaction. The questionnaire was specifically designed for the current research and had never been used in previous research. Although we calculated internal consistency reliability using Cronbach's Alpha (r=.81), we cannot be sure that the questionnaire items are a valid measure of copilot satisfaction. One possible procedure might be to assess the face validity of the measure by conducting focus groups of actual copilots. In addition, researchers could assess concurrent validity of the questionnaire by comparing it to other reliable and valid measures of worker satisfaction.

We also recommend that future researchers consider performing longitudinal studies of leadership acceptability in conditions of alarm mistrust. Authors such as Breznitz (1984) have stressed that alarm mistrust becomes more pronounced as task operators become more experienced with the alarm system. Similarly, the benefits and limitations of the leadership styles examined here may become more pronounced after a longer task performance period.

References

Billings, C.E. (1991). *Human centered aircraft automation: A concept and guidelines*. (NASA Technical Memorandum No. 103885). Moffett Field, CA: National Aeronautics and Space Administration.

Bliss, J.P. (1997). Alarm reaction patterns by pilots as a function of reaction modality. *The International Journal of Aviation Psychology, 7*(1), 1-14.

Bliss, J.P. (2003a). Collective mistrust of alarms. *International Journal of Applied Aviation Studies, 3*(1), 13-38.

Bliss, J.P. (2003b). An investigation of alarm related incidents in aviation. *International Journal of Aviation Psychology, 13*(3), 249-268.

Bliss, J. P., & Dunn, M. C. (2000). Behavioural implications of alarm mistrust as a function of task workload. *Ergonomics, 43*, 1283-1300.

Breznitz, S. (1984). *Cry-wolf: The psychology of false alarms*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chemers, M. M. (2000). Leadership research and theory: A functional integration. *Group Dynamics: Theory, Research and Practice, 4*, 27-43.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Comstock, J.R., & Arnegard, R.J. (1992). *The multi-attribute task battery for human operator workload and strategic behavior research*. (Technical Memorandum No. 104174). Hampton, VA: National Aeronautics and Space Administration.

Dubin, R. (1965). Supervision and productivity: Empirical findings and theoretical considerations. In R. Dubin, G. C. Homans, F. C. Mann & D. C. Miller (Eds.), *Leadership and productivity* (pp. 27-41). San Francisco, CA: Chandler.

Endsley, M. R. (1999). Situation awareness in aviation systems. In D. J. Garland, J. A. Wise, & V. D. Hopkin (Eds.), *Handbook of Aviation Human Factors* (pp. 311-326). Mahwah, NJ: Lawrence Erlbaum Associates.

Field, R. H. (1990). A test of the Vroom-Yetton model using manager and subordinate reports. *Journal of Applied Psychology, 7*, 362-366.

Foels, R., Driskell, J. E., Mullen, B., & Salas, E. (2000). The effects of democratic leadership on group member satisfaction: An integration. *Small Group Research, 6*, 676-701.

Foushee, C. H. (1982). The role of communications, sociopsychological, and personality factors in the maintenance of crew coordination. Aviation Space & Environmental Medicine. 53, 1062-1066.

Foushee, C. H. (1984). Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. *American Psychologist, 39*, 885-893.

Getty, D.J., Swets, J.A., & Pickett, R.M., & Gonthier, D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time, *Journal of Experimental Psychology: Applied, 1*, 19-33.

Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The evolution of crew resource management training in commercial aviation. *The International Journal of Aviation Psychology. 9*, 19-32.

Hines, W.E. (1998). *Teams and technology: Flight crew performance in standard and automated aircraft*. Unpublished doctoral dissertation. The University of Texas at Austin.

Jentsch, F., Smith-Jentsch, K. A. (2001). Assertiveness and team performance: More than "just say no". In Salas, E., Bowers, C. A., & Edens, E. (Eds.), *Improving teamwork in organizations: Applications of resource management training*. (pp. 73-94) Mahwah, NJ: Lawrence Erlbaum Associates.

Koonce, J. M. (1999). A historical overview of aviation human factors. In D. J. Garland, J. A. Wise, & V. D. Hopkin (Eds.), *Handbook of Aviation Human Factors* (pp. 311-326). Mahwah, NJ: Lawrence Erlbaum Associates.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics. 35*, 1243-1270.

Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Linton, P.M., Wierwille, W .W., Zaklad, A.L., Bittner,A.C., & Wherry, R.J. (1989). *Operator workload:*

Comprehensive review and evaluation of operator workload methodologies. (Technical Report No. 851). Alexandria, VA: United States Army Research Institute for the Behavioral and Social Sciences.

Nicholas, J. M., & Penwell, L. W. (1995). Proposed profile of the effective leader in human spaceflight based on findings from analog environments. *Aviation, Space, and Environmental Medicine. 66*, 63-72.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*, 230-253.

Prince, C., Chidester, T. R., Bowers, C., & Cannon-Bowers, J. (1992). Aircrew coordination: Achieving teamwork in the cockpit. In R. Swezey & E. Salas (Eds.), *Teams: Their training and performance.* Newark, NJ: Ablex.

Sexton, J. B., & Helmreich, R. L. (2000). Analyzing cockpit communications: The links between language, performance, error, and workload. *Journal of Human Performance in Extreme Environments. 5*, 63-68.

Smith, H.P. R. (1979). *A simulator study of the interaction of pilot workload with errors, vigilance, and decisions*. (NASA Technical Memorandum 78482). Moffett Field, CA: NASA Ames Research Center.

Vroom, V. H., (2000). Leadership and the decision-making process. *Organizational Dynamics, 28*, 82-94.

Vroom, V. H., Jago, A. G. (1978). On the validity of the Vroom-Yetton model. *Journal of Applied Psychology. 63*, 151-162.

Wickens, C.D. (1995). Aerospace techniques. In J. Weimer (Ed.), *Research Techniques in Human Engineering* (pp. 114-118). Upper Saddle River, NJ: Prentice Hall.

Wiegmann, D. A. & Shappell, S. A. (2001). Human error perspectives in aviation. *The International Journal of Aviation Psychology, 11*, 341-357.

Wickens, C. (1995). Aerospace techniques. In Weimer. J. (Ed.), *Research Techniques in Human Engineering.* (pp. 114- 118). Upper Saddle River, N.J.: Prentice Hall PTR.

Wickens, C.D., Sandry, D.L., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors, 25* (2), 227248.

International Journal of Applied Aviation Studies, 3, 2

FAA Pilot Knowledge Tests: Learning or Rote Memorization?

Stephen M. Casner
National Aeronautics and Space Administration
Ames Research Center
Moffett Field, CA

Karen M. Jones
and
Antonio Puentes
San José State University Foundation
San José, CA

and

Homi Irani
Embry Riddle Aeronautical University

Abstract

The FAA pilot knowledge test is a multiple-choice assessment tool designed to measure the extent to which applicants for FAA pilot certificates and ratings have mastered a corpus of required aeronautical knowledge. All questions that appear on the test are drawn from a database of questions that is made available to the public. The FAA and others are concerned that releasing test questions may encourage students to focus their study on memorizing test questions. To investigate this concern, we created our own database of questions that differed from FAA questions in four different ways. Our first three question types were derived by modifying existing FAA questions: (1) rewording questions and answers; (2) shuffling answers; and (3) substituting different figures for problems that used figures. Our last question type posed a question about required knowledge for which no FAA question currently exists. Forty-eight student pilots completed one of two paper-and-pencil knowledge tests that contained a mix of these experimental questions. The results indicated significantly lower scores for some question types when compared to unaltered FAA questions to which participants had prior access.

Requests for reprints should be sent to Beverly Laughead, FAA Academy, AMA-502-OU, P.O. Box 25082, Oklahoma City, OK 73125.

Introduction

A requirement for most every FAA pilot certificate or rating is the *knowledge test* that each applicant must pass to demonstrate mastery of the aeronautical knowledge required to safely exercise the privileges of the certificate or rating being sought. Required by the U. S. Code of Federal Regulations, knowledge tests present applicants with a series of multiple-choice questions designed to assess applicants' knowledge of aeronautical topics such as aerodynamics, weather, regulations, navigation, performance, and flight planning. The questions that appear on every knowledge test are drawn from an *item bank*, a fixed database of questions created by the FAA for each pilot certificate or rating. For example, the item bank for the Private Pilot Airplane knowledge test currently contains 915 questions. Each Private Pilot Airplane knowledge test presented to an applicant consists of 60 questions selected from the 915-question item bank.

Since the mid-1980s, the FAA has made the item banks available to the public. This allows every pilot applicant to access, in advance, all of the questions that can potentially appear on every knowledge test. This situation has led to great controversy for two reasons. First, pilots now have the opportunity to limit their study to just those questions that appear in the item bank. Since the item bank does not contain questions pertaining to all knowledge required to fly safely, it seems possible for pilot applicants to skip over some knowledge and still achieve a high score on the knowledge test. Second, pilots have the opportunity to simply memorize the questions and answers, rather than develop an understanding of the knowledge that the questions aim to test.

The problems associated with neglecting to learn material that is required but that is not formally tested are evident and need no further discussion. The negative outcomes associated with rote memorization strategies have been documented by more than a century of psychological research. In the case of the FAA knowledge tests, the crudest memorization strategies allow students to avoid looking at questions at all. Flash cards that contain an FAA question number on the front and the correct answer letter designation on the back allow students to engage in the most primitive paired-associate learning (Calkins, 1896; Mandler, 1970). Other strategies link the letter designation for the correct answer to the content of the question. For example, the FAA Private Pilot knowledge test contains a flight-planning question using an airport named Addison. The answer to the question was traditionally "A." Other psychologists have demonstrated the effectiveness of using simple mental images to successfully associate pairs of key words or phrases (Kothurkar, 1963; Paivio, 1965; Bower, 1972). For example, 8 minutes is the answer to an FAA flight-planning question involving a flight to an airport called Redbird. The answer is easily recalled after visualizing a red bird flying in a figure-eight pattern. Strategies like these are particularly useful for test questions that otherwise require time-consuming and effortful calculations (e.g., weight and balance, flight planning, aircraft performance, etc.)

Even in the cases in which students read the questions and memorize answers, the distinction between rote memorization and understanding has been demonstrated by other psychological studies (Tverksky, 1973; Craik and Lockhart, 1972; Kieras and Bovair, 1984). Students who achieve deeper levels of understanding are more successfully able to solve similar problems, recall solution steps, and use their knowledge in novel ways. Later research demonstrated the same phenomena in the domain of flight training. This research led to the same conclusions: when complex skills and concepts are to be learned, rote memorization is a poor substitute for understanding (Telfer, 1993; Moore and Telfer, 1990; Telfer, 1991). Consequently, these ideas are emphasized in the FAA's *Aviation Instructor's Handbook* (FAA, 2001).

Indirect Evidence of Question Memorization

The FAA has been concerned with the validity of the knowledge tests for some time. Their principal concerns are, if applicants are focusing their study on questions and answers, that student understanding may be negatively impacted and that the test scores being awarded to pilot applicants may not accurately reflect their knowledge.

Flouris (2001) made an initial attempt to relate differences in study methods to differences in knowledge test scores. Flouris compared the scores of students who completed a formal ground school at Auburn University and students who studied on their own. Flouris stressed that the Auburn ground schools stress learning for understanding and try to dissuade students from question memorization. Flouris found no significant difference between the two groups. Since the true study practices of both groups of pilots remain unknown, it is impossible to know what kind of preparation led to what test scores or the real level of mastery of either group at the time of the tests. Flouris' study cast further suspicion on the extent to which the FAA tests accurately measure knowledge and motivated future research.

A compelling piece of evidence in support of the question-memorization hypothesis was gathered by the FAA. This evidence is the amount of time that test takers spend completing knowledge tests. Today, all FAA knowledge tests are administered by computer. The computer systems record the amount of time that test takers require to complete each exam and the amount of time required to complete each question. Figure 1 shows timing data gathered by the FAA in 2002 for the knowledge tests required for the private and commercial pilot certificates as well as the instrument airplane rating (FAA, 2003).

The data show that many applicants complete the test in far less time than would be required for the average human to even read the questions and answers on the test.

Test	Ave. Completion Time (minutes)	Minimum Time (minutes)
Private Pilot – Airplane	73	2
Instrument Rating – Airplane	79	2
Commercial Pilot - Airplane	91	11

Figure 1. Average and minimum completion times for the Private, Commercial, and Instrument Airplane knowledge tests [Data provided by the FAA (FAA, 2003)].

Figure 2 shows timing data for a few individual test questions. These questions require applicants to work through complex calculations needed for flight planning under specified conditions. The data show that questions that would typically require several minutes to work through are being answered in merely a few seconds.

After departing GJT and arriving at Durango Co., La Plata Co. Airport, you are unable to land because of weather.

How long can you hold over DRO before departing for return flight to the alternate, Grand Junction Co., Walker Field Airport?

Total usable fuel on board, 68 gallons.
Wind and velocity at 16,000, 2308-16
Average fuel consumption 15 GPH.

A) 1 hour 33 minutes.
B) 1 hour 37 minutes.
C) 1 hour 42 minutes.

Mean Time: 569 seconds
Minimum Time: 6 seconds

An airplane descends to an airport under the following conditions:

Cruising altitude	6,500 ft
Airport elevation	700 ft
Descends to	800 ft AGL
Rate of descent	500 ft/min
Average true airspeed	110 kts
True course	335
Average wind velocity	060 at 15 kts
Variation	3 W
Deviation	+2
Average fuel Consumption	8.5 gal/hr

Determine the approximate time, compass heading, distance, and fuel consumed during the descent.

A) 10 minutes, 348, 18 NM, 1.4 gallons.
B) 10 minutes, 355, 17 NM, 2.4 gallons.
C) 12 minutes, 346, 18 NM, 1.6 gallons.

Mean Time: 195 seconds
Minimum Time: 5 seconds

Figure 2. Average and minimum completion times for flight planning questions [Data provided by the FAA (FAA, 2003)].

The FAA data clearly suggested that memorization was at work. For at least some test takers, the data clearly demonstrated that the test was not measuring the applicants' abilities to exercise the required knowledge and skills.

The evidence for question-memorization gathered by the FAA is compelling. However, there remains one credible refutation of the question-memorization hypothesis: it may be possible that pilot applicants satisfactorily learn the material and then memorize the questions in the item bank in order to expedite the taking of the test. In other words, in can be argued that the FAA's timing data demonstrated the presence of question-memorization, but it did not demonstrate the absence of understanding.

To directly address the concerns about the validity of the test, we conducted an experiment using private pilot students recruited from local flight schools. The experiment aimed to measure more directly what pilot applicants know at the time of their FAA knowledge test.

Method

Our approach to measuring pilot understanding of the required aeronautical knowledge was to administer an experimental knowledge test to private pilot applicants who had just prepared for and completed the FAA knowledge test. Our experimental knowledge test contained questions that varied from FAA questions in systematically increasing ways. Some experimental questions made cosmetic changes to existing FAA questions. Other experimental questions asked pilot applicants to use their aeronautical knowledge in different ways. The goal of our experimental test was to discover if pilot applicants' knowledge of the required aeronautical knowledge was based on understanding or on a more superficial review of the questions appearing in the FAA item bank. It was hypothesized that if applicants understood the required material, their performance on the experimental questions would be comparable to their performance on FAA questions.

We distinguished two types of questions that appear in the FAA item banks. *Skills questions* require the applicant to exercise a procedure they had learned such as interpreting a chart, working through a flight planning, or airplane performance calculation problem. *Knowledge questions* require the applicant to recall a fact or to reason about a concept they had learned.

To create our experimental item banks, we began with knowledge and skills questions that we randomly sampled from the FAA item banks. We left some of these FAA questions in their unaltered form to use as a control. We modified the remaining FAA questions to generate our treatment questions. Our selection and modification of the FAA knowledge and skills questions resulted in six different experimental questions types described below.

Unaltered Skills Questions

Figure 3 shows an example of an unaltered FAA skills question. Our purpose in including questions of this type was to establish a baseline measure for how well test takers performed on FAA skills questions, without having to ask them to reveal their score on the FAA test.

```
UA/OV KOKC-KTUL/TM 1800/FL120/TP BE90//SK BKN018-TOP055/OVC072-TOP
089/CLR ABV/TA M7/WV 08021/TB LGT 055-072/IC LGT-MOD RIME 072-089

The wind and temperature at 12,000 feet MSL as reported by a pilot are

A. 080 at 21 knots and -7 C.
B. 090 at 21 knots and -9 C.
C. 090  at 21 MPH and -9  F.
```

Figure 3. Unaltered skills question

Different Data Skills Questions

Figure 4 illustrates our only modification of the FAA skills questions. This modification substituted different data (a different pilot report (PIREP), in this case) for the original data appearing in the question.

```
UA/OV KMOD-KOAK/TM 2209/FL060/TP PA28//SK OVC022-TOP050/
CLR ABV/TA M8/WV 28026/TB MOD 035-060/IC LGT-MOD RIME 035-060

The wind and temperature at 6,000 feet MSL as reported by a pilot are

A. 280 at 26 knots and -8 °C.
B. 280 at 26 knots and 8 °C, measured.
C. 220 at 9 MPH and -8 °F.
```

Figure 4. Different figure skills question

Unaltered Knowledge Questions

Figure 5 shows an example of an unaltered FAA knowledge question. These questions were intended as a control to establish a baseline score for each applicant on FAA knowledge questions.

```
How should an aircraft preflight inspection be accomplished for
the first flight of the day?

A. Thorough and systematic means recommended by the manufacturer.
B. Quick walk around with a check of gas and oil.
C. Any sequence as determined by the pilot-in-command.
```

Figure 5. Unaltered knowledge question

Shuffled Knowledge Questions

Figure 6 illustrates the simplest of the modifications we made to FAA knowledge questions. This modification shuffled the order in which the answer choices appear in the question. This modification was designed to detect reliance on the crudest of memorization strategies that associate questions with memorized letters and answers (e.g., "A" for Addison).

How should an aircraft preflight inspection be accomplished for the first flight of the day?

A. Quick walk around with a check of gas and oil.
B. Any sequence as determined by the pilot-in-command.
C. Thorough and systematic means recommended by the manufacturer.

Figure 6. Shuffled knowledge question

Reworded Knowledge Question

Figure 7 illustrates our second and slightly more sophisticated modification of the FAA knowledge questions. This modification slightly reworded both the question and the answer.The rewordings used for these questions were limited to simple rearrangements of sentence structures and substitutions of non-critical words. In no case was a technical word or phrase (e.g., angle of attack) changed. This modification was designed to further test for memorization strategies based on simple question and answer recognition.

For the first flight of the day, an aircraft should be preflighted using

A. the procedure recommended by the manufacturer.
B. a walk around with a check of gas and oil.
C. a systematic procedure determined by the pilot-in-command.

Figure 7. Reworded knowledge question

Different Knowledge Question

Figure 8 illustrates our third and most sophisticated modification of the FAA knowledge questions. This modification extracted the target concepts tested by each question and asks a slightly different question about the same concepts. Questions of this type were created by a group of three certified flight instructors who worked until consensus was reached on two key points: (1) that the modified question tested the same concepts as the original question; and (2) that the modified question was not more difficult than the original question. The aim of this type of question was to measure applicants' performance when question-memorization strategies were no longer possible.

Figure 8 shows a modified question about wing flaps, along with all of the existing FAA questions about wing flaps. The question in Figure 8 illustrates our attempt to write questions that "fell between the cracks" of the existing FAA questions.

Our Different Question About Wing Flaps
One purpose of wing flaps is to
A - permit touchdown at higher airspeeds.
B - maintain a lower angle of attack at slower airspeeds.
C - permit safe flight at slower airspeeds during approach and landing.

Two Existing FAA Questions About Wing Flaps
One of the main functions of flaps during approach and landing is to
A - decrease the angle of descent without increasing the airspeed.

B - permit a touchdown at a higher indicated airspeed.
C - increase the angle of descent without increasing the airspeed.

What is one purpose of wing flaps?
A - To enable the pilot to make steeper approaches to a landing without increasing the airspeed.
B - To relieve the pilot of maintaining continuous pressure on the controls.
C - To decrease wing area to vary the lift

Figure 8. Different knowledge question

Apparatus

Two different paper and pencil knowledge tests were used for data collection. Each knowledge test contained equal numbers of randomly-sampled and ordered questions drawn from the item bank of question types described above. The first knowledge test contained 50 questions and measured performance for the first five question types. The second knowledge test contained 20 questions and compared performance for the third and sixth question types described above.

The number of questions appearing in our experimental test was a compromise between the statistical power needed and the expected fatigue and motivation levels among our student pilot participants.

Participants

A total of 48 student pilots from local flight schools agreed to participate in the study. We attempted to recruit every private pilot applicant that was scheduled to take the FAA Private Pilot knowledge test at every local flight school that offered testing services. Every applicant that was both qualified and willing participated in the study. Each pilot completed exactly one of the two different experimental knowledge tests.

Participants were not told any details about the experiment in advance; however, they were told that they did not need to undertake any special studying or preparation for the experiment.

Student pilots who completed the 50-question experimental knowledge test received payment equal to the cost of their FAA knowledge test. Pilots who completed the 20-question experimental knowledge test received a NASA Aviation t-shirt.

Procedure

Our experimental knowledge tests were completed by student pilots during a scheduled appointment. Most student pilots took our experimental knowledge test directly after taking their FAA Private Pilot Airplane knowledge test, although some completed the test up to three days after completing the FAA test. Unlike the FAA knowledge test, we did not pose a time limit for the experimental test. When completing the test, participants were allowed to use the same materials permitted by the FAA for their knowledge tests (e.g., blank paper, calculators, pencils).

All participants were informed that their responses would remain anonymous and no names were recorded.

Results and Discussion

Figure 9 shows the scores for the two knowledge tests that compared the six question types.

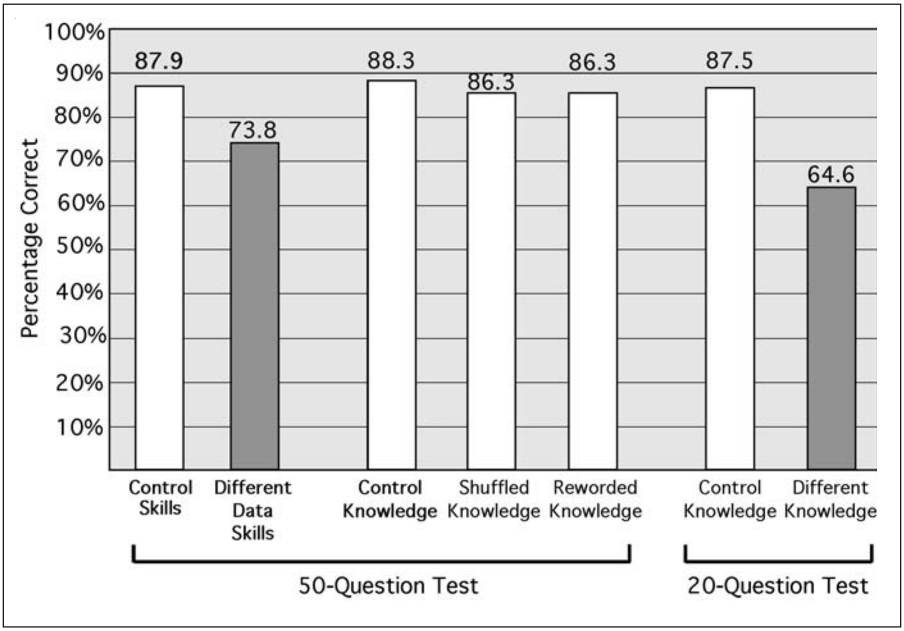

Figure 9. Average scores for two tests that compared the six different question types.

There was no difference between average scores for question types 3, 4, and 5: the unaltered, shuffled, and reworded knowledge questions. These results seem to rule out our worst fear: that participants relied solely on the crudest of memorization strategies in which learners used superficial cues available in the questions and answers. Shuffling and rewording questions had little effect on participants' ability to answer questions correctly.

There was a significant difference [$F(1, 23) = 15.4$, $p < .001$] between average scores for question types 1 and 2: the unaltered skills questions and the skills questions for which different data had been substituted for the original data. The means and standard deviations for the two question types were 87.9% (0.10) and 73.8% (0.18), respectively.This result seemed to confirm our suspicions about the generality of the problem-solving skills being learned by students.

Few instructors or evaluators would disagree – a student whose chart interpretation or flight-planning skills are limited to only those airports appearing in FAA test questions is operating under an obvious knowledge deficiency. Any test instrument that rewards such a pilot with a passing score is therefore also deficient.

There was also a significant difference [$F(1, 23) = 31.2$, $p < .0001$] between average scores for question types 3 and 6: unaltered knowledge questions and knowledge questions that participants did not have the opportunity to see in advance. The means for the two question types were 87.5% (0.12) and 64.6% (0.12), respectively. The less-than-passing scores may indicate fairly serious knowledge deficiencies, since these questions tested aeronautical knowledge that was also tested by existing FAA questions.

Conclusion

Our results supported the FAA's concerns that releasing test questions in advance may: (1) negatively affect the way students learn and ultimately understand required aeronautical knowledge; and (2) reduce the validity of the knowledge test as an assessment tool. If the design and execution of our study are accepted as reasonable, our data suggested that some pilot applicants may leave the test center with two things: (1) aeronautical knowledge that is deficient; and (2) a notarized document certifying that the student's aeronautical knowledge is not deficient.

Limitations of Our Study

We must admit a number of limitations of our study. First, despite our experts' efforts to control for question content and difficulty, the questions we created for question type 6 may have differed in content from, or may have been more difficult than, existing FAA questions. Second, there is the possibility that some falloff in scores was due to some subjects' unwillingness to work through the unfamiliar questions, accepting a lower score in the interest of time. Regardless, even for these "lazy" subjects, the falloff in scores between the two question types still suggested that question and answer memorization was used for the original FAA example questions, and their underlying knowledge of the material was not revealed by the current FAA test.

Changes Made by the FAA to the Knowledge Test

In July 2003, the FAA made a number of changes to the knowledge tests, based on their own investigations and on the evidence and inputs presented to them by outside parties. Our work was among the research that was reported to the governing branch in the FAA. These changes included:

1. FAA question numbers are no longer associated with questions that are released to the public.
2. Answer choices are now randomly shuffled.
3. Only a few examples of each type of skills question are now made

available. Questions appearing on the tests will be of the same form as the released question, but will use different data.

The first two changes were designed to foil crude memorization strategies that make simple associations between questions numbers and letter designations or order of appearance among answer choices. The results for our shuffled-answers condition did not suggest a presence of this kind of memorization for knowledge questions. However, we did not test our shuffling manipulation on skills questions. It may be the case that test takers sometimes do resort to this type of strategy for more difficult and time-consuming questions.

The third change agreed with our finding that significant numbers of test takers may have focused their study on individual questions rather than mastering the underlying skills being tested. There remains at least one credible threat to skills questions despite this countermeasure. Since skills questions require significantly more time to answer, any one test contains a small number of these questions. Under the current scoring system, all questions have the same value. Test takers may opt to simply guess at the more time-consuming skills questions, accepting a slightly lower overall test score.

One remedy for this problem is to increase the value of skills questions. There are a number of problems with this approach. First, it suggests that aeronautical knowledge areas that require more time to exercise are more important than other areas, a notion that is difficult to defend. Second, skills questions are often missed because of simple mathematical errors. Penalizing test takers several points for a simple math error seems unjustified.

What Depth of Understanding Is Enough?

In arguing for measures that might potentially lead to "deeper" understanding of aeronautical knowledge by students, we must recognize a few natural limitations.

A first limitation we encounter is that some aeronautical knowledge elements are simple facts and that the learning of facts inevitably leads to a process of rote memorization. For example, Class D airspace extends by default to an altitude of 2,500 feet AGL. The FAA offers no particular theoretical reason why this altitude is chosen instead of 2,400 or 2,600 feet. The student is left to learn that 2,500 feet AGL is the correct number. For aeronautical knowledge of this type, we should be sure to include questions in the FAA item bank that exhaustively cover all of the facts to ensure that students are learning them.

A second limitation we encounter is that student understanding of any given topic is naturally limited at any given level of expertise. For example, a new student pilot's understanding of flaps may be little more than a collection of memorized facts. Such a student simply has not yet had the opportunity to think through and experience the use of flaps in different flight situations, to develop theories about how they work, and to validate or invalidate those theories during practical experience. If we admit this limitation, care should be taken to ensure that the FAA item bank includes questions that cover a set of facts about flaps that will allow them to act accordingly, for the time being, in all required flight situations. This collection of facts must serve as an acceptable substitute while the student progresses toward a more mature understanding that only further study, practice, and experience can provide.

Knowledge Testing in the International Aviation Community

It is interesting to compare FAA knowledge testing practices to those used in other countries. The European Joint Aviation Authorities (JAA) requires the applicant to pass seven separate knowledge tests for the Private Pilot License (PPL). These tests cover: air law and operational procedures, human performance and limitation, flight performance and planning, aircraft general and principles of flight, navigation and radio aids, meteorology, and radiotelephony. Some tests contain as few at 15 questions. All tests are multiple-choice and the JAA does not reveal the potential test questions in advance. Questions from old tests and questions similar to JAA test questions are available for student study, as well as a suggested reading list.

Similarly, the Australian Civil Aviation Safety Authority (CASA) requires applicants for the Private Pilot License to pass Basic Aeronautical Knowledge and Private Pilot License knowledge tests. Similar to JAA tests, these tests are multiple-choice and sample questions (not drawn from the existing item bank) are made available for student study.

India requires five written tests: navigation, meteorology, regulations, technical, and radiotelephony.All tests are multiple-choice and the questions are not released to applicants in advance.

Whether aeronautical knowledge topics are tested in one or several tests, the move away from publishing test questions makes the FAA more consistent with testing practices used in other countries.

Acknowledgements

We thank all of the flight schools in the San Francisco Bay Area for their help with this research. Mikkel Grandjean-Thomsen and John Shopland provided valuable insights about knowledge testing in other countries.

References

Bower, G.H. (1972). Mental imagery and associative learning. In L. Gregg (Ed.), *Cognition in learning and memory*, 51-88.

Calkins, M. W. (1896). Association: An essay analytic and experimental. Psychological Review Monograph Supplement 1 (2).

Craik, F. & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, 11, 671-684.

Federal Aviation Administration (2003). Personal communication.

Federal Aviation Administration (2001). *Aviation Instructor's Handbook*. FAA-H-8083-9.

Flouris, T. (2001). The impact of ground schools in a collegiate aviation program on FAA written exam scores. Paper presented at the *University Aviation Association Annual Conference*, Nashville, TN, October 24-27.

Kothurkar, V.K. Effect of stimulus-response meaningfulness on paired-associate learning and retention. *Journal of Experimental Psychology, 1963*, 65, 305-308.

Kieras, D. E., and Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science 8*, 255-273.

Mandler, G. (1970). Incremental acquisition of pairedassociate lists. *Journal of Experimental Psychology, 84*, 185186.

Moore, P. J., and Telfer, R. A. (1990). Approaches to learning: Relationship with pilot performance, *Journal of Aviation/Aerospace Education and Research 1* (1), 44-58.

Paivio, A. (1965). Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior, 4*, 32-38

Telfer, R. A. (1993). *Aviation instruction and training*, R. A. Telfer (Ed.), Aldershot, England: Ashgate Publishing Company.

Telfer, R. A. (1991). How airline pilots learn. Paper presented at the *Annual Conference of the New Zealand Psychological Society*, Palmerston North.

Tversky, B. (1973). Encoding processes in recognition and recall. *Cognitive Psychology*, *5*, 275-287.

International Journal of Applied Aviation Studies, 3, 2

Differential Effects On Self-Efficacy Of Treatment Components For Fear Of Flying

Lucas J. Van Gerwen, Claudia Van de Wal and Philip Spinhoven
Stichting VALK, Leiden University
P.O. Box 110, 2300
AC LEIDEN, The Netherlands
Tel.: 31-71-5273733
Fax.: 31-71-5273796
E-mail: info@valk.org or sticktingvalk@cs.com

René F.W. Diekstra
Municipal Health Department, Rotterdam, The Netherlands

Richard Van Dyck
Vrije University, Amsterdam, The Netherlands

Abstract

According to self-efficacy theory, judgments and expectations concerning performance capabilities are relevant for the initiation, persistence, and modification of anxiety problems and specific fears. This article presents a study on the differential effects on self-efficacy expectancies of various treatment components in a fear of flying protocol. Subjects consisted of 199 patients (38.7% male and 61.3% female) with fear of flying, who applied to a treatment agency to deal with their fear of flying. Fear of flying was assessed with three specific flight anxiety measures before and after treatment, the FAS, FAM, and VAFAS. The treatment process was measured from assessment to follow-up at eight different points in time with two self-efficacy instruments, based on Bandura's work (Bandura, 1977, 1986): the Flight Self-Confidence Scale (FSCS), and the Fear of Flying Coping Scale (FFCS). Results showed that the most effective treatment components for enhancing self-efficacy were (in descending order of importance): (a) exposure, graded practice with an in-therapy flight, (b) relaxation training with breathing exercises, (c) information about flying, and (d) controlling up-

Requests for reprints should be sent to Beverly Laughead, FAA Academy, AMA-502-OU, P.O. Box 25082, Oklahoma City, OK 73125.

setting thoughts. The conclusion is that performance experiences particularly enhanced self-efficacy. Limitations of the study are discussed and suggestions for future research provided.

DIFFERENTIAL EFFECTS ON SELF-EFFICACY OF TREATMENT COMPONENTS FOR FEAR OF FLYING

People tend to engage in behaviors they believe will produce a desired effect and that they believe they can perform. We are more likely to take courses of action that we anticipate will lead to desired goals than those that appear less likely to succeed. We are also more likely to attempt actions and strategies that we think are within our range of capabilities than those that seem to exceed it. Several important psychological theories are based on these premises and concern the role of personal effectiveness and control in psychological development and health. Self-efficacy theory (Bandura, 1977, 1986) is one of the most prominent of these theories. The crux of self-efficacy theory is that initiation of and persistence in behaviors and courses of action are largely determined by (a) outcome value, (b) outcome expectancy, and (c) self-efficacy expectancy, that is, judgments and expectations of behavioral skills and the capacity to successfully implement the selected courses of action. Much of Bandura and his associates' work focused on understanding the role of self-efficacy in the development and treatment of fears and phobias (Bandura, 1986). It showed that low self-efficacy expectancies are an important feature in the continuation of anxiety problems and specific fears.

In previous studies, Bandura and his colleagues worked with patients having specific fears or phobias, both to test the basic assumptions and hypotheses of self-efficacy theory and demonstrate its clinical utility (Bandura, 1986). A self-efficacy model of anxiety primarily concerns anticipating that danger or harm is imminent and expecting an inability to effectively prevent or cope with the anticipated aversive event. Self-efficacy theory also hypothesizes that the key element common to all successful clinical interventions for anxiety disorders is an increase in patients' sense of self-efficacy, enabling them to master the anxiety-provoking situation (Bandura, 1977). According to self-efficacy theory, people with anxiety problems have inaccurate, unrealistic expectations of their own behavior and underestimate or overestimate certain results or consequences. They feel that nothing can be done to control events or feel incapable of performing the actions that could control events and lead to achieving goals (of which others seem capable).

The present study was designed to gather information about the specific contribution of elements of a multi-component, standardized fear of flying treatment program to a belief in self-efficacy to master that anxiety. Humans are biologically driven to explore and master their environment, and they feel good when they explore new situations and deal with them effectively. To explore the environment and new situations by plane is a problem for an estimated 10 - 40% of the general population in industrialized countries (Arnarson, 1987; Dean & Whitaker, 1982; Ekeberg, 1991; Nordlund, 1983). Fear of flying is a significant problem and events like the terrorist attacks on the Twin Towers and Pentagon might even lead to higher numbers of people with flight anxiety. Fear of flying is a heterogeneous phenomenon (Howard, Murphy, & Clarke, 1983; Van Gerwen, Spinhoven, Diekstra, & Van Dyck, 1997) and can be conceptualized both as a situational phobia and the expression of nonsituational phobias. Fear of flying often consists of one or more other phobias (Haug, Brenne, Johnsen, Berntzen, Götestam, & Hugdahl, 1987) and can also be the effect of generalizing one or more natural environment phobias, as described in DSM-IV (APA, 1994).

As the number of air travelers worldwide has increased, interest in anxiety associated with flying also has increased. There has been an increase in facilities around the world for treating passenger flight anxiety, often the result of collaboration between airlines and private entrepreneurs or mental health professionals (Van Gerwen & Diekstra, 2000). These developments have not been matched by an increase in scientific research on flight anxiety treatment programs or analysis of the effective ingredients and components of these programs. The latter is probably due to the fact that most treatment programs have mainly focused so far on their overall effects and less on redrafting or restructuring treatment, based on analysis of the contribution of program components to the overall effect. In actual clinical practices, the common components of fear of flying treatment programs are information on flying, teaching coping strategies (breathing exercises and muscle relaxation) and exposure to test flights. More advanced treatment programs also use components like cognitive restructuring and beside the test flights on commercial airliners also exposure to flight stimuli, such as a visit to a stationary aircraft in a hangar and the use of a flight simulator.There is a growing literature on the use of virtual reality exposure therapy but it is not yet routinely used in actual clinical practice (Van Gerwen, Diekstra, Arondeus, & Wolfger, submitted). The study of virtual reality exposure therapy for fear of flying is promising. Virtual reality exposure can have lasting effects both for the short (see Maltby, Kirsch, Mayers, & Allen, 2002 and also Rothbaum, Hodges, Smith, Lee, & Price, 2000) and long terms, as indicated in the first year-long follow-up study by Rothbaum, Hodges, Anderson, Price, and Smith (2002). It is also promising, because a recent study confirmed that virtual reality exposure could be more effective than imaginary exposure in treating fear of flying (Wiederhold, Gevirtz & Spira, 2001).

This study gathered information about the specific contribution of elements of a fear of flying treatment program in an actual clinical practice. The process measures used to assess treatment component effects were based on Bandura's concepts in his book, entitled *Belief in Self-Efficacy* (Bandura, 1995). According to Bandura, "perceived self-efficacy refers to beliefs in one's capabilities to organize and execute the courses of action required to manage prospective situations" (Bandura, 1977, 1986, 1998).

Two scales for belief in self-efficacy developed by Bandura (the Confidence Scale and the Parental Self-Efficacy Scale) were reviewed, translated into Dutch, adapted for flying situations, and psychometrically examined (Van de Wal, 1993). One of them concerns the ability to control negative thoughts while flying and the other assesses coping and dealing with flight anxiety feelings. This resulted in the Flight Self-Confidence Scale (FSCS) and the Fear of Flying Coping Scale (FFCS), two self-report instruments that are easy to administer.

This study is part of a larger research project on the determinants of fear of flying and the differential efficacy of cognitive-behavioral interventions for flying phobics. This study has four goals: (a) To determine the differential effects of the specific components of the treatment program on belief in self-efficacy. (b) To determine the relationship between increased belief in self-efficacy and decreased fear of flying during treatment. (c) To ascertain whether fear of flying at post-treatment (i.e. three months after an in-therapy flight) can be predicted on the basis of self-efficacy before the start of treatment. (d) To assess whether the self-efficacy score at post-treatment can predict the long-term effects of the fear of flying treatment program.

Method

Participants

Participants were consecutively referred to the VALK Foundation. All suffered from fear of flying and participated in the long-established multi-component fear of flying treatment program developed by the VALK Foundation, located in Leiden, the Netherlands. This foundation is a joint enterprise of the Department of Clinical and Health Psychology at Leiden University, KLM Royal Dutch Airlines, and Amsterdam Airport Schiphol. Trainers employed by VALK are all certified clinical psychologists and/or psychotherapists, some with airline cabin crew or pilot experience. The participant sample (N=199) was collected from 1995 to 1998. The group consisted of 38.7% (N=77) males and 61.3% (N=122) females. Mean age was 41.4 years (SD=10.6). Thirty-four percent had higher education (higher professional or academic training), and 21% medium level (professional) education. Thirty-three percent completed high school education. Twelve percent completed elementary school education or lower professional training. Most participants had flying experience; only 10.4% (N=8) of the male and 9% (N=11) of the female participants had never flown before. The majority of the participants, 94.2% of the women and 85.8% of the men, reported symptoms ranging from severe anxiety to panic in flight situations.

Fifty-six percent were employed and 22% were self-employed, -10% were housekeepers (all women), and two smaller groups did either volun-teer work (3%) or attended a school or university (1%). Eight percent of the participants did not fit into one of the above groups.

Measures

Flight Self-Confidence Scale (FSCS). The FSCS was used to measure belief in self-efficacy, specifically to measure participants' belief in their ability to control thoughts. The FSCS consists of nine self-report items that are answered on an 11-point scale, ranging from 0 (I cannot) to 10 (I definitely can). Participants were instructed to imagine being on a turbulent flight with their seat belts fastened. They were asked to evaluate their confidence in their ability to neutralize any unpleasant thoughts they might have in this situation. Participants with low anxiety should have fewer negative thoughts about flying than participants with high flight anxiety, and low-anxiety participants should express more control over negative thoughts than high-anxiety participants. Some examples of the items are "Thinking of something else," "Stopping the thought," and "Keeping the thought under control." There are two versions of the FSCS, versions A and B, which were randomly presented to participants. Originally, the FSCS consisted of 18 items and the psychometric properties proved to be good (Van de W al, 1993). A one-factor solution was found explaining 53.7% of the variance (eigenvalue 9.67) and reliability measured with Cronbach's Alpha was .92. After dividing the 18-item questionnaire into the two 9-item A and B questionnaires, Cronbach's Alphas were .84 for A and .89 for B. The correlation of A with the original questionnaire was .97, of B with the original questionnaire .96, and the correlation between A and B was .87. It was therefore concluded that the two parallel versions of the FSCS correlated sufficiently to be treated as the same instrument (Van de Wal, 1993).

Fear of Flying Coping Scale (FFCS). The FFCS was also used to measure belief in self-efficacy, but from a different angle than the FSCS. The FFCS measures participants' belief in their ability to cope and deal with feelings. The FFCS consists of five items that are answered on a 7-point scale, ranging from 1 (not good at all) to 7 (very good). The items concern different emotional aspects of fear of flying, and participants have to judge how well they can cope with each. People with confidence in their ability to deal effectively with a threatening situation will approach the situation with self-assurance and calm, while those with serious doubts about their coping skills will anticipate catastrophes and generate a state of affective arousal that will interfere with their ability to function effectively. Some examples of the items are "How well can you stop worrying?," "How well can you cope with tension caused by flying?," and "How well can you keep difficult moments from getting you down?" There are also two different versions of the FFCS, A and B, which were randomly presented to participants. Originally, the FFCS consisted of 10 items and the psychometric properties were good (Van de Wal, 1993). A one-factor solution was found explaining 56.7% of the variance (eigenvalue 5.66), and reliability measured with Cronbach's Alpha was .91. After dividing the questionnaire into an A and B version, Cronbach's Alphas were .83 for A and .82 for B. The correlation of A with the original questionnaire was .97, of B with the original questionnaire .97, and the correlation between A and B was .89. Van de Wal's study (1993) concluded that the two parallel versions of the FFCS correlated sufficiently to be treated as the same instrument.

Flight Anxiety Situations questionnaire (FAS). The FAS has 32 items consisting of three subscales: (a) Anxiety experienced anticipating flying, up to the time the flight actually starts (takeoff is announced) containing 14 items; (b) Anxiety experienced during a flight (from start until landing) containing 11 items; (c) Anxiety experienced in general in connection with airplanes, regardless of personal involvement in a flight containing 7 items (Van Gerwen, Spinhoven, Van Dyck, & Diekstra, 1999). The psychometric properties of the FAS proved to be excellent (Van Gerwen et al., 1999). This study showed an internal consistency of subscales varying from .88 to .97 and test-retest reliabilities ranging from .90 to .92 as measured with Pearson product moment correlation coefficients.

Flight Anxiety Modality questionnaire (FAM). The FAM has 18 items that are designed to measure the following two modalities: (a) Somatic Modality, pertaining to physical symptoms, and (b) Cognitive Modality, related to the presence of distressing cognitions (Van Gerwen et al., 1999). The psychometric properties of the FAM proved to be good to excellent (Van Gerwen et al., 1999). The study showed an internal consistency of .89 for both subscales and test-retest reliability of .79 for the Somatic Modality and .84 for the Cognitive Modality.

Visual Analogue Flight Anxiety Scale (VA FAS). The VA FAS enabled participants to indicate the extent to which they were anxious about flying on a one-tailed visual analogue scale. This scale ranges from 0 "No flight anxiety" to 10 "Terrified." Participants were instructed to "Please indicate how anxious you are about flying at present."

Follow-up information. Follow-up evaluations were carried out after 3, 6 and 12 months, during which participants were asked to state the number of one-way flights they had taken in that period.

Procedure

The design of the present study was uncontrolled. Data were collected in an open study before, during, and after the fear of flying treatment program. Moreover, treatment components were provided in a fixed order.The standard multi-component fear of flying treatment started with an assessment phase (diagnostic interview and questionnaires). After initial contact, participants were invited for assessment at the VALK Foundation during which they filled out questionnaires on fears and phobias in general and fear of flying in particular (see Measures-). In a subsequent semi-structured interview, information was gathered on flying behavior (Flying History Interview). To identify a representative group of fearful flyers, they should have a minimum score of 6 on the VA FAS and be willing to follow and pay a two-day group treatment. Assessment was followed by an individualized preparation phase of two hours of individual therapy for relaxation and breathing exercises and, if necessary, exposure to other feared situations (i.e. heights and elevators) and a two-day group treatment (20 hours, with a maximum of eight participants). The two-day group treatment started with a presentation by a pilot, covering aerodynamics, moving aircraft parts, procedures and performances as well as air traffic control and meteorological aspects like turbulence. A video showed aircraft maintenance, air traffic control, and an explanation of radar and transponder.After this, explicit cognitive interventions were introduced. Participants were taught how to identify and change anxiety-provoking thoughts. In addition, information on anxiety, its physical effects, and what can happen during a panic attack was provided. A cognitive model of panic and anxiety was presented to show how panic could result from catastrophic misinterpretation of certain bodily sensations (Clark, 1986). Coping skills were ranked in terms of perceived efficacy. Some examples of coping skills are controlled breathing, muscle relaxation, the "stop" technique for negative thoughts and distraction from negative thoughts. Participants were informed of the best possible flight preparations. Next, an imaginary flight was taken, which was tape-recorded. For this exercise, an aircraft cabin wall was available with twelve aircraft seats. Then in vivo exposure was used. The group visited a Boeing 747 in a hangar at the airport and performed a relaxation exercise in the stationary plane. Two flights were taken in a flight simulator, an Airbus A-310 cabin simulator, which accurately simulates a night flight, including sound and motion. During the first flight, participants received an explanation of flight sounds and motion. During the second simulator flight, participants were encouraged to practice coping exercises. After the two simulated flights, the group headed for a one-way, one-hour flight in Europe in a commercial airliner Boeing 737 guided by the therapist and the pilot. A briefing was held at the airport of destination and the return flight provided a second opportunity to practice. Three months after the in-therapy flight, a three-hour follow-up session was given to monitor maintenance and progress, at which point post-treatment data were collected. The treatment program was based on a standard treatment manual (Van Gerwen, 1992). For a more detailed description of the treatment program, see Van Gerwen, Spinhoven, Diekstra, & Van Dyck (2002).

The FSCS, FFCS, VAFAS, FAS, and FAM were completed during the diagnostic assessment phase and the follow-up session. Only the FSCS and the FFCS were used during the treatment process. They were completed eight times during the period extending from the diagnostic interview to the follow-up. The eight times the questionnaires were completed were: T1) before the diagnostic interview,T2) after individual therapeutic sessions with relaxation training and breathing exercises, T3) after technical information, information on flying and aviation in general, pilot/cabin crew training, sensory experiences and personal hygiene (nutrition, exercise, etc.), T4) after Rational Emotive Therapy (RET) exercises, for controlling upsetting thoughts, T5) after stress-management with information on anxiety and distraction training, T6) after imaginary exposure, a coping-exercise, T7) after in vivo exposure, graded practice (hangar, flight simulator, in-therapy flight), and T8) three months later at a follow-up session. Each time, participants completed one of the two versions of the FSCS and FFCS. Every participant was given all four questionnaires four times in a random order. There were two versions of the questionnaires in order to keep participants from getting bored and from becoming so familiar with a single version that they would no longer read the items properly. Since randomization was important, all possible combinations of the two versions of the FSCS and FFCS questionnaires were used: FSCS-A + FFCS-A, FSCS-A + FFCS-B, FSCS-B + FFCS-A and FSCS-B + FFCS-B. Moreover, at the 3, 6 and 12-month follow-up sessions, the VAFAS scores were collected and participants were asked to state the number of self-initiated one-way flights they had taken in that period. Table 1 shows the points in time when measurements were taken. The length of treatment, the time between the assessment phase and the test flight was approximately three to five weeks.

The treatment program incorporated all four sources of efficacy information as delineated by Bandura (1986): (a) emotional arousal, teaching patients to relax and feel less anxious when flying, which is why relaxation training and breathing exercises were included; (b) verbal persuasion, providing participants with information, encouraging them to attempt behaviors they fear and challenging their expectations of catastrophe; (c) vicarious experiences, observing live models flying (such as the therapist) and participating in a group; (d) performance experiences, actual practice with feared behaviors, such as leaving home and approaching the airport and, of course, taking a real flight. In vivo experience with the feared object or situation was performed in group therapy sessions.

Table 1
The components of treatment when assessment took place and the measurements that were taken.

Intervention	FSCS	FFCS	VAFAS	FAS	FAM	Flights taken
T1 Diagnostic assessment	X	X	X	X	X	X
T2 Individual training	X	X				
T3 Information on flying	X	X				
T4 RET exercise	X	X				
T5 Stress management	X	X				
T6 Imaginary exposure	X	X				
T7 In vivo exposure	X	X				
T8 3-month follow-up	X	X	X		X	X X
6-month follow-up			X			X
12-month follow-up			X			X

Statistical Analyses

To assess overall training effects and differences between pretreatment and follow-up measurements three months after treatment, the VAFAS, FAS, FAM, FSCS and FFCS were analyzed with paired t-tests. The size of training effects was analyzed by computing Cohen's d. Cohen considers an effect size of 0.20 to be slight, 0.50 to be moderate and 0.80 to be substantial. Cohen's d was calculated using the following formula: M post-treatment – M pretreatment / SD common (Cohen, 1977). For each measurement moment, we calculated the mean of the FSCS and FFCS items, irrespective of the version participants used. Differences between the eight consecutive measurements and the previous measurements were analyzed with paired t-tests and Cohen's d for both the FSCS and FFCS.

To analyze the relationship between changes in belief in self-efficacy and fear of flying, multiple regression analyses were used and standardized residualized gain scores (SRS) were computed. All dependent variables were statistically corrected for pretreatment differences on the dependent variable analyzed. Pearson correlations were computed between the self-efficacy SRSs and the fear of flying SRSs. The predictive value of the baseline self-efficacy scores for the magnitude of the training effect was analyzed by computing Pearson correlations between the pretreatment self-efficacy scores and the fear of flying SRSs. Cohen (1977) considers a Pearson correlation coefficient of 0.10 to be slight, 0.30 to be moderate and 0.50 to be strong.

Finally, four hierarchical stepwise multiple regression analyses were performed to analyze the predictive value of the self-efficacy measurements at the 3-month follow-up for the VAFAS follow-up scores at 6 and 12 months. In the first step, VAFAS scores at the 3-month follow-up were forced into the equation to control for differences in outcome at that measurement moment. In the second step, FSCS or FFCS scores at the 3-month follow-up were entered as independent variables in the regression analyses to check their additional, independent contribution to explaining the variance of VAFAS scores at the six or 12-month follow-up.

Results

Overall effect of treatment

The overall effect of treatment is shown in Table 2. This table provides the paired t-tests and effect sizes of the mean on the VAFAS and the means on the FAM and FAS subscales scale at pretreatment (T1) and at 3-month follow-up (T8), but also, the mean scores on the FAS Total which consists of the 32 items together.All variables showed statistically significant improvement 3 months after treatment. All participants took the in-therapy flight. The effect sizes for improvement on the scales and subscales for fear of flying were very large. In particular, improvement on the most important In-Flight subscale showed a very large Cohen's d of 2.62. Improvement on the VAFAS even showed a Cohen's d of 3.43. In addition, all participants took flights on their own within one year after treatment. The mean number of total one-way flights made after one year of treatment was 9.61 (SD 9.29), four to five round trips.

Table 2

Paired t-tests and effect sizes, means and standard deviations on the self-report flying scales at diagnostic assessment (T1) and at 3-month follow-up (T8).

Assessment instrument	*T1* M	SD	*T8* M	SD	*t–test* t	df	*Effect size* d
VAFAS	*8.4*	*1.4*	*2.4*	*2.0*	*40.37**	*198*	*3.43*
FAS – anticipatory	28.6	11.6	8.5	8.4	23.07*	192	1.98
FAS – in-flight	29.2	9.3	7.4	7.2	30.52*	191	2.62
FAS – generalized	5.2	5.1	1.5	2.6	11.32*	192	0.92
FAS – total	70.7	24.2	19.5	18.7	27.95*	192	2.37
FAM – cognitive	15.0	10.0	4.1	4.4	15.63*	192	1.41
FAM – somatic	18.3	7.4	4.2	5.0	26.49*	192	2.23

*$p < .001$

Belief in self-efficacy substantially increased over the course of treatment. Scores on both the FSCS and the FFCS questionnaires showed a highly significant increase from diagnostic assessment (T1) to the 3-month follow-up (T8). The FSCS mean changed from 3.6 to 8.2 ($p < .001$) and the FFCS mean from 3.1 to 5.2 ($p < .001$).

Differential changes in self-efficacy for the specific components

The FSCS and the FFCS questionnaires differed in the component of the treatment with the biggest increase in self-efficacy scores. The results of the t-tests are given in Tables 3 and 4. With respect to changes in participants' belief in their ability to control negative thoughts (FSCS), individual sessions with relaxation training and breathing exercises (T2), and the exposure component (consisting of a visit to a Boeing 747, a flight simulator and an 'in-therapy' flight) (T7) produced the biggest gain, Cohen's d effect sizes being almost equal (0.82 and 0.81, respectively; see Table 3). With respect to belief in the ability to cope with anxiety feelings (FFCS), the biggest increase was observed after exposure (T7) (see Table 4). Individual sessions with relaxation training and breathing exercises (T2) were the second best intervention with respect to changes on the FFCS.

Table 3

Paired t-tests and effect sizes with respect to the Flight Self Confidence Scale (FSCS).

Intervention	range M	1-11 SD	t-test*** t	Effect size d
Diagnostic assessment	3.6	1.8		

Table 3 - continued

Individual training	5.0	1.7	12.0**	0.82
Information on flying	6.0	1.6	11.6**	0.59
RET exercise	6.8	1.5	13.3**	0.56
Stress management	7.3	1.6	7.4**	0.31
Imaginary exposure	7.2	1.8	1.6 n.s.	-0.08
In vivo exposure	8.5	1.6	14.1**	0.81
3-month follow-up	8.2	1.7	2.8*	-0.19

n.s.= not significant, *p < .01 ** p < .001

*** = t-test performed with the previous intervention assessment

Table 4

Paired t-tests and effect sizes with respect to the Fear of Flying Coping Scale (FFCS).

Intervention	range M	1-7 SD		t-test*** t	Effect size d
Diagnostic assessment		3.1	1.1		
Individual training	3.7	0.9		9.0*	0.62
Information on flying	4.2	0.9		10.7*	0.58
RET exercise	4.6	0.8		10.7*	0.56
Stress management	4.9	0.9		6.9*	0.36
Imaginary exposure	4.9	1.0		0.6 n.s.	-0.03
In vivo exposure	5.7	0.8		13.5*	0.84
3-month follow-up	5.2	1.0		6.8*	-0.50

n.s. = not significant, *p < .001

*** = t-test performed with the previous intervention assessment

On both the FSCS and the FFCS questionnaires, technical information on aviation and sensory experience (T3) was the third most effective component in producing changes. The exercise of controlling and modifying negative thoughts (T4) was the fourth most effective intervention. Information on anxiety and distraction training (T5) also produced a moderate increase in self-efficacy. While most interventions resulted in an increased belief in self-efficacy, the imaginary exposure exercise (T6) did not.

At the follow-up session three months after treatment (T8), there was a significant decrease in scores on both questionnaires. The drop was rather substantial, especially for belief in the ability to cope as assessed with the FFCS, having a Cohen's d of -0.50. However, as is clearly shown in Figures 1 and 2, it remained higher than before exposure. Figures 1 and 2 visualize clearly that belief in self-efficacy (as measured with the FSCS and FFCS) had, in this order of components, relative effects of various sizes and the effects appeared additive except after the imaginary exposure exercise (T6) and three months after treatment (T8).

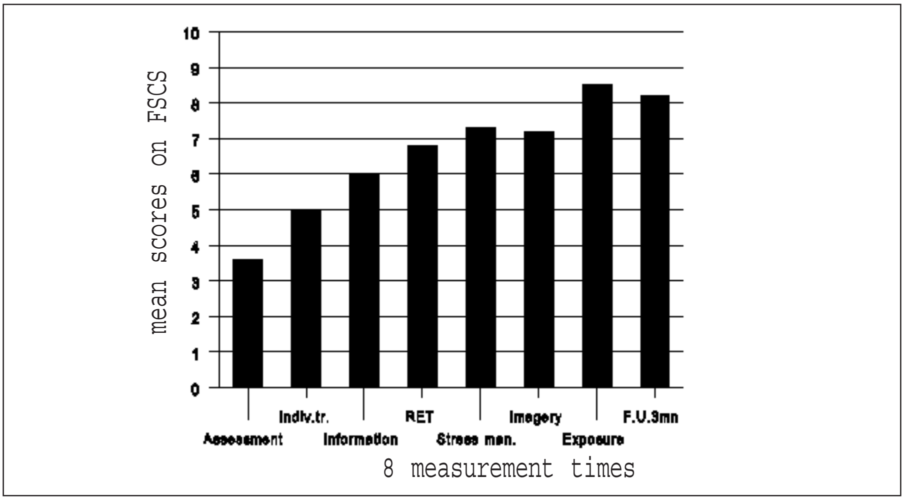

Figure 1. Differential changes in self-efficacy measured with the Flight Self-Confidence Scale (FSCS)

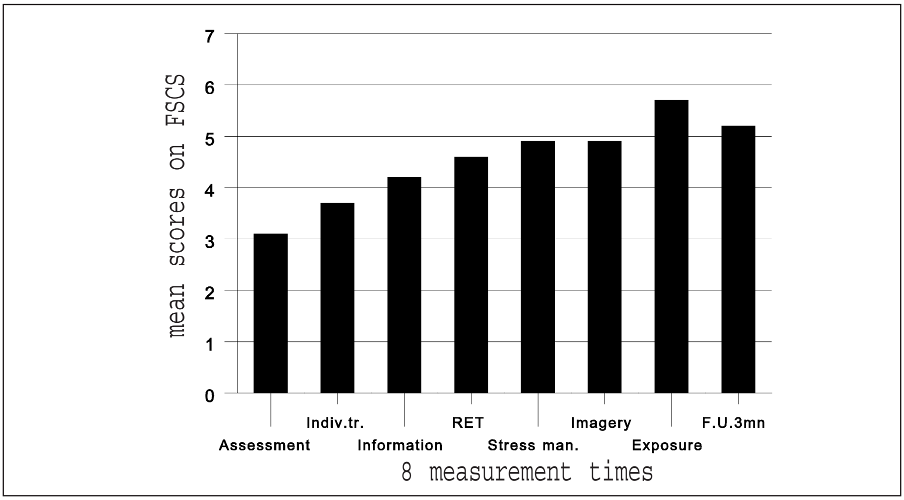

Figure 2. Differential changes in self-efficacy measured with the Fear of Flying Coping Scale (FFCS)

The relationship between increasing self-efficacy and decreasing fear of flying

From the Pearson correlations shown in Table 5, it could be concluded that there was a very strong and consistent relationship between the increase of belief in self-efficacy during treatment, as measured with the FSCS and FFCS, and a decrease in fear of flying, as measured with the FAS, FAM and VA FAS.

Accordingly, it could also be concluded that self-efficacy is strongly congenial with fear, but not identical. Cohen (1977) considered a Pearson correlation of 0.50 to be strong; here a range was found between -.41 to -.68 (see Table 5).

Table 5
Pearson correlations between residualized gain scores for self-efficacy and fear of flying at 3-month follow-up (T8).

	Residualized gain scores (T8): FAS anticipatory	FAS in-flight	FAS	FAS general	FAM total	FAM cognitive	VAFAS somatic
FSCS (T8)	-.57*	-.68*	-.49*	-.64*	-.54*	-.60*	-.50*
FFCS (T8)	-.52*	-.57*	-.41*	-.55*	-.48*	-.52*	-.49*

*p < .01

The initial level of self-efficacy as a predictor of the level of fear of flying after treatment

There was no significant relationship between the FSCS and FFCS scores at diagnostic assessment and FAS, FAM, and VAFAS residualized gain scores at three-month follow-up. The Pearson correlations ranged between .08 for the FAS in-flight with the FSCS to -.12 for the VAFAS and the FFCS. Thus, the initial level of self-efficacy at diagnostic assessment did not predict a decreased fear of flying as measured with the FAS, FAM, and VAFAS at the follow-up session. Consequently, it could only be concluded that changes in flight anxiety were unrelated to belief in self-efficacy before treatment.

Post-treatment self-efficacy scores as predictors of the long-term effect of fear of flying treatment

Fear of flying as measured with the VAFAS showed that with a mean value of 8.4 (SD 1.4) at pre-treatment fear decreased even further to 2.4 (SD 2.0) at 3 months, 2.1 (SD 1.9) at 6 months and 1.9 (SD 1.9) at 12 months after treatment. Analyzed with Wilks' Lambda, the changes on the VAFAS over time between all four measurement points were significant ($F(3,196) = 745.40$, $p < .001$). Pairwise Bonferroni comparisons (alpha individual: $p < .008$) of mean differences between VAFAS1 and VAFAS2 (5.95 SD .15), VAFAS1 and VAFAS3 (6.26 SD .14), and VAFAS1 and VAFAS4 (6.46 SD .14) were all significant ($p < .001$). Pairwise comparisons of the mean differences between VAFAS2 and VAFAS3 (.32 SD .09, $p < .01$) and VAFAS2 and VAFAS4 (.52 SD .12, $p < .001$) were also significant.

Self-efficacy measured at the 3-month follow-up (T8) was significantly associated with VAFAS scores at the 6 and 12-month follow-ups. Correlations were all around .50, ranging from .46 to .51 ($p < .001$). Four separate hierarchical multiple regression-analyses were performed to analyze whether self-efficacy measures predicted fear of flying at the 6 and 12-month follow-ups independent and above the level of fear of flying at post-treatment. In these analyses, the fear of flying level at the 6-month follow-up (VA FAS3) and 12-month follow-up (VA FAS4) were predicted by VA FAS scores at the three-months follow-up in the first step and the FSCS or FFCS scores at the three-months follow-up in the second step of the equation. FFCS scores at post-treatment (T8) explained a significant, additional 2% proportion of the variance of VA FAS scores at the 6-month follow-up (F change (1,196) = 7.878; p = .006) and 2% of the variance of VA FA S scores at the 12-month follow-up (F change (1,196) = 7.821; p < .001). FSCS scores at post-treatment only explained a significant, additional 4% proportion of the variance of VA FAS scores at the 12-month follow-up (F change (1,196) = 13.437; p < .001), over and above the effect of the VA FAS scores at the 3-month follow-up.

Discussion

This paper is one of the few recent studies that provides quantitative data on the process of treating a large group of patients suffering from fear of flying in actual clinical practice. Measures of self-efficacy expectancies were used in the diagnostic assessment prior to treatment, at various stages of treatment, and following treatment.

Belief in self-efficacy substantially increased over the entire course of treatment. Once patients began to experience some success, they might have developed a stronger sense of self-efficacy. According to self-efficacy theory, small successes strengthen patients' sense of self-efficacy and their expectations of additional, bigger successes. Most effective clinical interventions help people experience success as a way of restoring self-efficacy expectancies and a general sense of personal efficacy (Frank, 1961; Goldfried & Robins, 1982).

Although most intervention components produced an increase in self-efficacy, the largest gain was observed after exposure (visit to a Boeing 747, flight simulator and in-therapy flight) (T7) and after relaxation training with breathing exercises (T2). These components were followed in efficacy by technical information on flying (T3), the exercise of controlling and modifying negative thoughts (T4), and information on anxiety (T5). This confirmed that, as the literature suggested, exposure and the learning of coping skills and how to control irrational thoughts are important components of a cognitive-behavioral approach (Maddux, 1991). This is one reason why this fear of flying treatment program was based on these components. The most important sources of efficacy beliefs are people's ability to modulate or control their internal physiological and emotional states (Bandura, 1995) and the amount of confidence people have in- controlling their emotions and cognitions in specific flight situations.

Post-treatment measures (VA FAS, FAS and FAM) were collected three months after treatment instead of immediately after the in-therapy flight to avoid euphoric scores. There is already considerable anecdotal evidence that participants score better directly after an in-therapy flight due to a euphoric mood. At the follow-up session three months after treatment, both the FSCS and FFCS questionnaires showed a significant decrease in belief in self-efficacy scores in comparison to scores obtained directly after the in-therapy flight (see Figures 1 and 2). Until now, no other studies have discussed the possibility of a euphoric mood score directly after an in-therapy flight and no systematic empirical data were available, but we think that increased anxiety and a decreased belief in self-efficacy three months after treatment can be regarded as a correction for this euphoric effect. On the other hand, the long-term effects measured with the VAFAS showed a further decrease of flight anxiety 6 and 12 months after treatment. Consequently, there was no evidence of a fear of flying relapse after three months.

Individual differences in belief in self-efficacy did not predict reduction of fear of flying in the short or long term. Although belief in self-efficacy might not be universally beneficial, it could be argued that participants with a greater belief in self-efficacy will show more improvement in a therapy program where self-control is stressed, such as in the present study (Thompson & Wierson, 2000). Apparently, the present fear of flying treatment program, incorporated all sources of efficacy information such as emotional arousal, verbal persuasion, vicarious experiences and performance experiences (Bandura, 1986) and was comprehensive enough to preclude the interaction effects of individual differences in belief in self-efficacy and specific characteristics of self-efficacy enhancement interventions.

Enhancement of belief in self-efficacy following treatment was closely correlated with reduction of fear of flying. Because correlation does not allow causal inferences, it cannot be concluded that enhancing self-efficacy produced the reduction of fear obtained. However, some empirical evidence was found that a greater belief in self-efficacy predicted maintenance or even further improvement of treatment gains. Participants with a greater belief in self-efficacy at post-treatment (T8) showed a somewhat better outcome at the 6 and 12-month follow-ups, even after statistically correcting for differences in fear of flying at post-treatment. These results suggest that belief in self-efficacy cannot be solely regarded as an epiphenomenon of anxiety and leave open the possibility than changes in belief in self-efficacy are of some causal importance in the process of reducing fear.

The design of this study had several methodological shortcomings. It was an uncontrolled open study and components had a fixed order. Because a control condition was not included, it remains undetermined whether the changes during treatment were primarily due to the specific treatment components or could perhaps be better explained by a general nonspecific intervention, maturation, or test-retest effect. The fixed order of treatment components precluded any definite conclusion about the individual efficacy of the various components. Ethical considerations rendered infeasible the methodologically more desirable approach of randomizing the order of treatment components in a counterbalanced design. Consequently, conclusions about the relative effect of the various treatment components have to be interpreted in the context of the particular order of treatment components in the present program. Although the largest increases were observed after relaxation training with breathing exercises and exposure, we do not know whether exposure alone would have resulted in the same gains without other therapeutic preparation. Nor do we know whether relaxation training with breathing exercises would have had the same impact if introduced later in the therapy. However, as stated above, therapeutic and ethical problems would have arisen if participants were exposed to a flight without preliminary treatment preparation, since most participants in this study had flown before and found it very distressing.

Furthermore, self-efficacy expectancy measures have been developed mainly for research; however, their suitability for use in clinical settings is unknown. They have good logical or face validity, are brief and straightforward, highly specific with regard to problem behaviors and problem situations, and lend themselves to use at frequent intervals to efficiently monitor patient progress (Goldfried & Robins, 1982). However, in this study, there was a close correlation between changes in self-efficacy and changes in flight anxiety. Self-efficacy theory is proposed as a model of behavioral change, not of emotional experience. Bandura (1984) stated, "Self-efficacy scales ask people to judge their performance capabilities and not if they can perform non-anxiously." In this study, it was assumed that self-efficacy could also be applied to controlling anxiety-related cognitions and that this cognitive change could help control anxiety states. There are strong correlations of changes in self-efficacy and self-reported flight anxiety, but they are not high enough to suggest that the two concepts are the same. However, whether the construct validity of the self-efficacy questionnaires is sufficient and whether these questionnaires do not also measure anxiety to a substantial extent could be questioned. Consequently, it cannot be concluded that the effect of the treatment is through improvement in self-efficacy.

A related point for discussion is that there are likely also other factors that could explain the change in anxiety apart from self-efficacy. Self-efficacy theory focuses on the more cognitive aspects of mastery and effectiveness expectancies and values than on more affective constructs such as needs, motives, and feelings. It was already known that exposure works as an anxiety reducing therapy (Emmelkamp, 1994), but exposure affects more than just self-efficacy. Change processes could also be taking place in other areas. Perhaps other processes were responsible for changes, which cannot be explained on a cognitive level or by self-efficacy theory. It is most likely that habituation plays a leading role subsequent to an exposure program.

The self-efficacy model and the measurement of self-efficacy could help clinicians target specific competency-related beliefs and situations, predict areas of potential difficulty, and tailor interventions to meet patients' special needs. A self-efficacy scale that provides detailed information about "at risk" situations can also help therapists clarify, anticipate, and prevent problems patients typically encounter when attempting new or anxiety-provoking behaviors. However, the results of this study indicated that self-efficacy measures were no crucial predictors of therapy success, neither at pretreatment or post-treatment, given the close correlations with anxiety measures.

A recommendation for future studies is to investigate whether self-efficacy expectancy measures directly measure psychological adjustment and also whether low self-efficacy expectancies are sufficient for diagnosing psychological dysfunction or whether high self-efficacy expectancies form a guarantee of psychological health. Another recommendation could be to investigate whether randomizing the order of treatment components in a counterbalanced design would give a different result for the individual efficacy of the various components in comparison to this study.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th ed., rev.).* Washington, DC: Author.

Arnarson, E. O. (1987). The prevalence of flight phobia among Icelanders. Paper presented at the *First Nordic Meeting, Aviation and Space Medicine*. Oslo, Norway.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 191-215.

Bandura, A. (1984). Recycling misconception of perceived self-efficacy. *Cognitive Therapy and Research, 8,* 231-255.

Bandura, A. (1986). *Social foundations of thought and action: a social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1995). *Self-Efficacy in changing societies*. Cambridge University Press.

Bandura, A. (1998). Self-efficacy regulation of affective states and biological stress reactions. J.F. Cruz (Ed.), *Stress anxiety and emotional disorders.*

Clark, D. M. (1986). A cognitive approach to panic. *Behaviour Research and Therapy, 24* (4), 461-470.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences.* New York, U.S.A.: Academic Press.

Dean, R. B. & Whitaker, K. M. (1982). Fear of flying: impact on the U.S. air travel industry. *Journal of Travel Research, 21,* 7-17.

Ekeberg, O. (1991). *Flight phobia: prevalence, sympathetic responses and treatment.* Department of Internal Medicine, Ulleval Hospital, Oslo, Norway: University of Oslo.

Emmelkamp, P. M. G. (1994). Behavior therapy with adults. In A.E. Bergin & S.L. Garfield (eds.), *Handbook of psychotherapy and behavioral change* (4th ed., pp 379-427). New York: John Wiley & Sons, Inc.

Frank, J. D. (1961) Emotional and motivational aspects of the disarmament problem. *Journal of Social Issues. 17*, 20-27.

Goldfried, M. R. & Robins, C. (1982). On the facilitation of self-efficacy. *Cognitive Therapy and Research, 6*, 361-379.

Haug, T., Brenne, L., Johnsen, B. H., Berntzen, D., Götestam, K. G., & Hugdahl, K. (1987). A three-system analysis of fear of flying: a comparison of a consonant vs. a non-consonant treatment method. *Behaviour Research and Therapy, 25* (3), 187-194.

Howard, W.A., Murphy, S. M., & Clarke, J. C. (1983). The nature and treatment of fear of flying: a controlled investigation. *Behavior Therapy, 14* (4), 557-567.

Maddux, J. E. (1991). Self-efficacy. In Snyder, C. R. (Ed); Forsyth, Donelson R. (Ed). *Handbook of social and clinical psychology: The health perspective.* Pergamon general psychology series, Vol. 162. (pp. 57-78). Elmsford, NY, US: Pergamon Press, Inc.

Maltby, N., Kirsch, I., Mayers, M., & Allen, G. (2002). Virtual reality exposure therapy for the treatment of fear of flying: a controlled investigation. *Journal of Consulting and Clinical Psychology, 70*, 1112-1118.

Nordlund, C. L. (1983). A questionnaire of Swedes' fear of flying. *Scandinavian Journal Behavior Therapy 12*, 150-168.

Rothbaum, B. O., Hodges, L. F., Smith, S., Lee, J. H., & Price, L. (2000). A controlled study of virtual reality exposure therapy for the treatment of fear of flying. *Journal of Consulting and Clinical Psychology, 68*, 1020-1026.

Rothbaum, B. O., Hodges, L. F., Anderson, P. L., Price, L., & Smith, S. (2002). Twelve-month follow-up of virtual reality and standard exposure therapy for the fear of flying. *Journal of Consulting and Clinical Psychology, 70,* 428-432.

Thompson, S. C., & Wierson, M. (2000). Enhancing perceived control in psychotherapy. C. R. Snyder & R. E. Ingram (Eds.). *Handbook of psychological change. Psychotherapy processes & practices for the 21st century.* (pp. 177-197). New York: John Wiley.

Van de Wal, R. C. (1993). *Nieuwe effectmaten voor de vliegangstbehandelingen: de zelfvertrouwenschaal en de copingschaal [New effect measures for the treatments of flight anxiety: the self-confidence scale and the coping scale].* (Unpublished Thesis). Leiden, University of Leiden.

Van Gerwen, L. J. (1992). *Vliegangst behandeling draaiboek [Fear of flying treatment manual].* Leiden, The Netherlands: The VALK Foundation (unpublished).

Van Gerwen, L. J. & Diekstra, R. F. W. (2000). Fear of flying treatment programs for passengers: An international review. *Aviation, Space, & Environmental Medicine, 71*, 430-437.

Van Gerwen, L. J., Diekstra, R. F. W., Arondeus, J., &. Wolfger, R. (2003). *Fear* *of flying treatment programs for passengers: An international update*. Manuscript submitted for publication.

Van Gerwen, L. J., Spinhoven, Ph., Diekstra, R. F. W., & Van Dyck, R. (1997). People who seek help for fear of flying: Typology of flying phobics. *Behavior Therapy, 28*, 237-25.

Van Gerwen, L. J., Spinhoven, Ph., Diekstra, R. F. W., & Van Dyck, R. (2002). Multi-component standardized intervention for fear of flying: description and effectiveness. *Cognitive & Behavioral Practice, 9*, (2) 138-149.

Van Gerwen, L. J., Spinhoven, Ph., Van Dyck, R., & Diekstra, R. F. W., (1999). Construction and psychometric characteristics of two self-report questionnaires for the assessment of fear of flying. *Psychological Assessment, 11* (2) 146 – 158.

W iederhold, B. K., Gevirtz, R. N., & Spira, J. L. (2001). Virtual reality exposure therapy vs. imagery desensitization therapy in the treatment of flying phobia. In Riva, G. & Galimberti C. (Eds.). *Towards cyber Psychology: Mind, Cognitions and Society in the Internet Age* (pp. 254-272). Amsterdam, IOS Press.

International Journal of Applied Aviation Studies, 3, 2

Training Development Papers

Biometric Technology: A Primer for Aviation Technology Students

Stephen J. Elliott
Purdue University
West Lafayette, IN 47907-1416

Abstract

Post September 11, 2001, there has been an increased focus by the airline industry and governments to evaluate various technologies associated with security and identification. Prior to September 11, Automatic Identification and Data Capture (AIDC) technologies have been used extensively in airports and the aviation industry for a number of years in applications ranging from bar coded baggage tags to magnetic stripes on boarding cards. Although used in limited applications before September 11, there is a renewed focus on another branch of automatic identification technology, namely biometrics. This paper will present a primer on biometric technologies and provide a theoretical framework, an understanding of the role of biometrics, and case studies, specifically within the airline sector.

Introduction

Since the September 11, 2001 terrorist attacks in New York, Washington, and Pennsylvania using commercial airliners, there has been considerable coverage, both politically and within the media, on transportation security, with the majority of the focus being placed upon the aviation industry. Although prior to September 11, there were installations of biometric technologies within airports,

Requests for reprints should be sent to Beverly Laughead, FAA Academy, AMA-502-OU, P.O. Box 25082, Oklahoma City, OK 73125.

these were focused on immigration efforts by the Immigration and Naturalization Service (INS), called the INSPASS, and employee access control. The passage of the Aviation and Transportation Security Act focused on new and emerging technologies that may include biometrics (Lazarick, 2002). In addition, Section 403 of the U.S. Patriot Act mandated "the Attorney General and the Secretary of State work with the National Institute of Standards and Technology to develop a technology standard that can be used to verify the identity of persons applying for a United States visa for the purposes of conducting background checks for a United States visa or such persons seeking to enter the United States pursuant to a visa" (U.S. Patriot Act," 2002). As a result of the work in this standards committee, biometrics is being considered for a number of transportation-related applications. As such, those in academia now may want to include biometrics as part of a course in aviation security or airport management. This paper is written as an introduction to biometric technologies, outlining the general concepts and definitions that students in aviation technology will come across, the specific classifications, and the specific issues within an airport environment that may affect the performance of individual biometric technologies. There will also be a section on biometric technologies that have been successfully deployed within an aviation setting.

Biometric Definitions

There are three common ways of identifying someone's identity: through an individual's knowledge, such as a password; through something an individual has, for example an identification card; and the third method, through something they own. In many airport applications, individuals gain access to specific areas by providing a card and personal identification number (PIN). A combination of these provides a more robust security option.

Biometric identification is defined as the "automatic identification or identity verification of (living) individuals based on behavioral and physiological characteristics" (Wayman & Alyea, 2000, p. 269). Physiological biometrics include facial recognition, finger, face, eye, and hand, whereas behavioral includes speaker, keystroke, and dynamic signature verification (Rejman-Greene, 2001).

Furthermore, a biometric must be measurable, robust, and distinctive (Newton & Woodward, 2001). There are two applications of biometric technologies: positive and negative identification. Positive identification is "to prove you are who you say you are," whereas negative identification is "to prove you are not who you say you are not" (Wayman & Alyea, 2000, p. 269).

All biometric systems follow a generic biometric model - shown below in Figure 1 (Wayman, 2000a). The generic biometric system consists of five different sections - data collection, signal processing, decision, transmission, and storage.

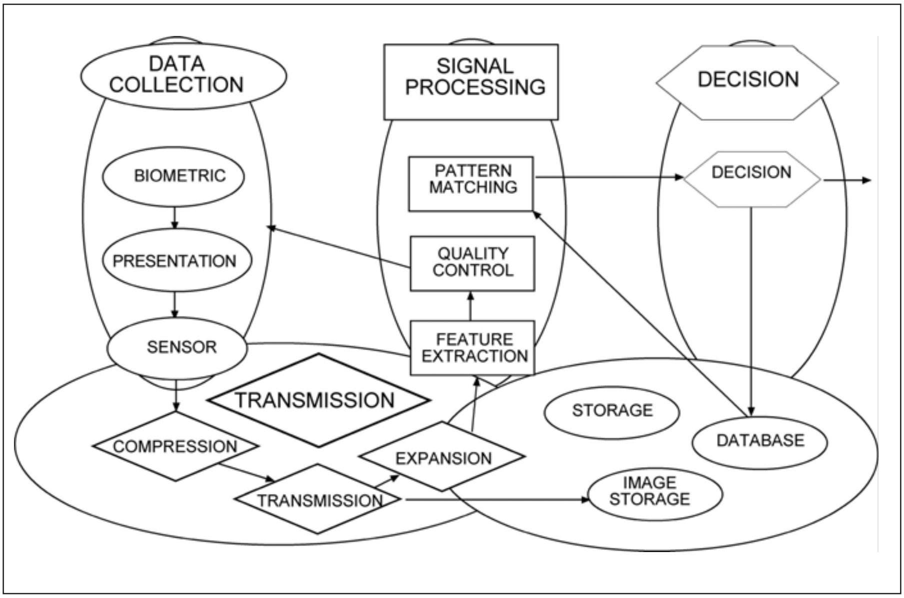

Figure 1. The General Biometric Model (Wayman, 2000a).

Assume the following scenario. An individual needs to gain access to a particular area that is secured by a biometric system. Once their identity has been confirmed by presenting the access issuing authority a set of credentials, they need to be entered into the system. An individual is entered into the system by placing their biometric near or on a sensor.There are several ways of collecting the biometric image. For a fingerprint, the finger is placed on a sensor and an image taken; for face recognition, an individual looks into a camera; and for hand geometry, an individual places their hand under a camera. The biometric image then is transmitted to the pattern matching section, where features are extracted and the features are processed through a quality control. If acceptable, the images are stored in the database. If these images are not acceptable, there is a failure to acquire (FTA), and the individual is requested to present their biometric again. Figure 2 shows an example of a fingerprint collected from a sensor, and Figure 3 shows the extracted features called minutiae points. The sample is taken a number of times. Some face applications can take up to 100 pictures and fingerprint sensors may collect between three and seven images. These initial images then create a template.

Figure 2. Fingerprint

Figure 3. Extracted features from the fingerprint

Several factors may affect the enrollment. The first is whether the individual has measurable characteristics, if not, then an image may not be able to be collected. The second is whether the image is robust – for example, if the biometric sample is damaged or changed overtime. The third factor is distinctive – how it varies amongst the enrolled population.

The images are stored as a template, which is then stored, awaiting for the next visit of the individual to the device. When that occurs, the individual presents the biometric to the sensor.Again, if the sensor cannot pick up a good enough quality image, the individual is asked to present their sample again. This could be due to poor presentation; for example, the individual may not have put his/her entire finger on the device, may have misaligned his/her hand, or it may be due to some other environmental condition. This environmental condition may occur when the enrolled device and the access area are in different locations.

Once the individual presents his/her sample to the biometric device, the algorithm goes through a process of matching the two images. There are four different outcomes associated with this matching computation. If the sensor does not pick up a good enough sample, it is a failure to acquire. If the sensor matches the sample to the wrong template, this is a false match and an impostor gains access to the area. If the sensor does not match the sample to the correct template (called false non match), the correct individual is denied access to the area. A threshold is acceptance or denial of a biometric sample based upon the score falling above or below a threshold. The threshold is variable so that the levels of security can change (AfB/ICSA, 1998). Verification is the process of comparing a submitted biometric sample with a previously stored biometric template in order to determine the identity of the subject (Wayman, 2000b).

The issue of biometric use and its cultural, moral, and social ramifications depends largely on the application. Applications can be classified into seven categories: cooperative versus non-cooperative, covert versus overt, habituated versus non-habituated, attended and non-attended, and whether the system is public or private, open or closed. Cooperative versus non-cooperative refers to the action of the deceptive user or impostor.Applications verifying the positive claim of identity, such as access control occur when the deceptive user is cooperating with the system in order to be recognized as someone he or she is not (Wayman, 2000a). A non-cooperative application is a deceptive user who is not cooperating with the system in order not to be identified. When a claimant is unaware that a biometric identifier is being used, the use of the system is classified as covert. If the user is aware that the biometric identifier is being taken, the system is classified as overt. A habituated application is everyday use of the biometric device; such as entry to a particular room, or log-on to a network computer. However, if the use of the biometric device is infrequent, the system is un-habituated. All systems will be un-habituated at the installation of the system, and may have a mixture of habituated and un-habituated users throughout the operation of the device as new claimants and frequent and infrequent claimants use the device. A biometric device is non-attended if it does not have an operator or someone is guiding users with the device. A standard environment is the environment in which the biometric device operates. The sixth category classifies the users, such as employees of the company (private) or customers (public). If there is a requirement to share with other biometric systems then the system is classified as open; if not, it is closed.

In the aviation industry, there are a number of different application classifiers with which to deal. For the airport worker who needs to gain access to the ramp on a number of occasions during their workday, the application classifier will be cooperative, overt, habituated, non-attended, private, and closed. This is the same application classifier as with the existing badges; however, biometric technologies add an additional level of security to the token and pin. For other security needs, these application classifiers will change. Applications may need to be covert or have non-habituated users. Other applications may have non-cooperative individuals, with the resulting security setup altered to take this into consideration.

Negative and Positive Identification

There are two applications of biometric technologies: positive identification is "to prove you are who you say you are" and negative identification is "to prove you are not who you say you are not." For positive identification, verifying the claim of the individual is through the comparison of the sample to an enrolled template. In the negative identification system, the user makes no claim to identity; therefore, requiring the search of the entire database (UKBWG, 2000). Therefore, when enrolling in a negative identification system, a comparison is made of the enrollment template with all other enrollment templates in a system to make sure that there is not a match (UKBWG, 2000).

Positive identification does not require the use of biometrics; other forms of physical identification such as driver's licenses, passports, and passwords can positively identify the individual. Positive identification applications include the INSPASS system, which enables frequent travelers the opportunity to enroll in a hand geometry database to gain entry into the United States, used by the U.S. Immigration and Naturalization Service (Wayman, 2000b).

Conversely, negative identification, is only achieved using biometrics (Wayman, 2000b). Some applications require the use of negative identification, such as biometrics on commercial driver's licenses. When enrollment occurs in a negative identification system, the system compares the samples with all the templates in the database to ensure that there are no duplicate records. Table 1 below shows the relationship between Positive and Negative Identification as explained in UKBWG (2000).

Table 1.
Positive and Negative Identification Classifications

Positive To prove I am who I say I am	Negative To prove I am not who I say I am not
Comparison of a submitted sample to a single claimed template Can use alternative forms of identification such as drivers licenses and passports	Comparison of a submitted sample to all enrolled templates No alternatives to this form of identification exists
Voluntary	Not voluntary
Biometric linked to external information only through external documents	Linkage to personal information not required

Biometric Technologies Commonly Used in Airports

There are a number of biometric technologies already in use in a number of airports throughout the world. The most commonly deployed biometric technology in airports is hand geometry. Other biometrics described are fingerprint, iris and face recognition.

Hand Geometry

Hand geometry measures hand characteristics such as finger length, width, thickness, and curvatures (Jain, Bolle, & Pankanti, 2001). It does not measure palm prints or take into consideration variables such as fingerprints, scars, or color of the skin. Hand geometry has been implemented at a number of installations for immigration, access control, and time and attendance. One such example is at the San Francisco International Airport, which has installed over 600 machines at doors for access control. The hand geometry readers are hardwired into an electromagnetic door lock. Individuals swipe a magnetic card with the individual's biometric template and then place their hand in the reader. This benefits the airport since they previously had a problem with lost ID cards.

Attaching a biometric requirement to the ID card adds another layer of security to access control. It also ensures that if an airport ID card is lost, the card can be deactivated along with the biometric. Therefore, if someone were to find the ID card, he/she would not be able to simply use the card to gain access, as he/she would also have to present a biometric. Another example of pre-September 11th installation of hand geometry is time and attendance applications. According to Biometrics (1999), the airline has over 2,000 full- and part-time employees using the system. The advantage in using biometric authentication for time and attendance based systems is that it eliminates 'buddy punching' and provides the airline with an accurate time and attendance record. At Ben Gurion International Airport (2000), departing or arriving passengers use hand geometry readers in conjunction with a credit card used for initial verification to gain a receipt, allowing them to proceed through security.

Fingerprints

Fingerprints have a long history in personal identification and authentication, the foundations of which were established by Galton in 1882, in a book titled

Fingerprints, and Henry in 1901, who revised Galton's features, also known as minutiae points. Fingerprints are partitioned using the Henry Classification and include right loop, left loop, whorl, arch, and tented arch. Galton features are used for matching (Halici, Jain, & Erol, 1999). Galton defined four characteristics of a fingerprint: ridges, specifically the beginning and ending points of them; forks; islands; and enclosures. Galton features were expanded to include dots, short ridges, crossover, bridge and spur (Halici, Jain, & Erol, 1999). Within biometric systems, there are two types of finger image typologies – minutia- and pattern-based systems. Minutiae-based systems have traditionally been used within the forensic approach, and are the de facto standard for AFIS (Automatic Fingerprint Identification Systems). In this methodology, the minutiae are extracted and recorded based on their location. For a more detailed discussion on minutiae-based fingerprint feature extraction, see Halici, Jain, & Erol (1999).

The pattern-based approach has been driven by the requirements of smaller sensors that can be deployed on mobile computers and PDA's. According to Soutar (2002), a pattern-based approach consists of two stages – a conversion of the raw fingerprint image to "a cropped and down-sampled finger pattern" followed by "cellular representation of the finger pattern to create the finger pattern interchange data" (p. 7).

Because of the Federal Aviation and Transportation Security Act of 2001, all airports are to conduct criminal history record checks on all employees that have unescorted access to secure areas. Fingerprint devices are being installed in several airport facilities. As a response to the change in the law, many airports are purchasing biometric fingerprint technology. O'Hare International Airport uses finger imaging to secure entry and delivery into the universal air cargo area. Combined with a smart card that holds manifest information, the fingerprint system identifies drivers and their corresponding trucks and cargo.

Iris Recognition

Iris recognition uses camera technology to recognize patterns within an iris at a distance. These patterns include "arching ligaments, furrows, ridges, crypts, rings, corona, freckles and a zigzag collarette" (Daugman, 2001, p. 2). Iris recognition has been deployed in a number of airports for employee identification and identification at immigration halls. The first U.S. installation of iris recognition for airport security was at the Charlotte/Douglas International Airport (North Carolina, USA). The system includes an access control station for U.S. Airways pilots and flight attendants, and a separate station for airport employees and other airline personnel. At Frankfurt Airport, Germany, an iris recognition-based identification system controls airport and airline employee access to restricted areas and enhances security operations. This alternative eases the cumbersome paper- and token-based identification processes and ensures security and convenience for airlines and employees. A five-month trial using iris recognition is underway at Heathrow Airport. The trial is aimed at reducing the time in the arrival hall where immigration documents are processed. Up to 2,000 frequent-flyers on Virgin Atlantic Airlines and British Airways will participate in the program (Eye Ticket, 2002).

Facial Recognition

There are several methodologies for facial identification, as discussed in Howell (1999) and Podio (2001). Facial recognition is being trialed in a number of airports including Dallas-Fort Worth and Palm Beach International, where the technology will be deployed in checkpoint areas (Visionics, 2002). In Canada, Thunder Bay International Airport utilizes facial recognition as an access control solution (Industry Updates, 2002).

Biometrics and the Airport Environment

As discussed earlier, within the airport environment, there will be many different types of users; and many different classifiers may be employed in the same facility, for example, attended enrollment at a security badge office, but unattended at the door or checkpoint. According to the Aviation Security Biometrics Working Group (ASBWG, 2001), there are five focus areas. These are transportation employees, surveillance, passengers, pilots and flight crew, and air traffic controllers.

Each of these particular users will interact with the device in different ways. While the classifiers described above are important in evaluating a biometric before implementation, user psychology is important when assessing the implementation success of the biometric installation. According to Ashborne (2000), biometric literature rarely discusses user psychology. If a user does not want to use the system, he or she may not be consistent in the use of the system (e.g. the presentation to the sensor), and will produce a wide variance in distance measurements resulting in higher average error rate. The additional effect of rejecting the user, even if claiming the correct identity, provides more reason for the individual to be non-cooperative with the system. Conversely, someone who is enthusiastic about the device will produce lower-than-average error rates. For particular airport applications, door security for example, there are tangible benefits for users to accept the system. Check-in validation (due to the tangible benefits) may induce the individual to perform at a lower-average-error rate. However, checkpoint screening at a choke-point within an airport may result in higher errors, as people may or may not cooperate.

Identification brings problems of its own. Researchers have classified the population into four groups: sheep, goats, lambs and wolves (Doddington, Liggett, Martin, Przybocki, & Reynolds, 1998). Sheep are defined as the normal populace that have no problem enrolling and using biometrics, goats as people that for one reason or another simply cannot enroll, lambs as people that their biometric measurements are easily mimicked in some way, and wolves as those that can have measurements taken that will pass for someone else (Woodward, 2002).

This poses challenges to airport authorities when trying to select the best biometric for their specific application. The strengths and weaknesses of these biometric technologies are a function of the application to which they are being introduced. When selecting a specific technology, airport managers need to focus not only within the application classifiers described above, but on a number of areas. User participation is a factor for successful deployment.

If a user does not want to use the system, he or she may not be consistent in the use of the system (e.g., the presentation to the sensor), and will produce a wide variance in distance measurements and a resulting higher average error rate. The additional effect of rejecting the user, even if claiming the correct identity, provides more reason for the individual to be non-cooperative with the system.

However, several factors can improve the successful deployment of a biometric application. The first is to make sure that the biometric installation is not dependent solely on the specific biometric alone. Biometrics is just one part of the security system, and there are other factors that can affect the overall system chances of success or failure. The biggest mistake that individuals make when introducing a biometric system is to assume that they know their population. However, the case is that biometrics may perform poorly when there are outliers in the population. As discussed earlier, user perception is also important. For an employee access control solution at an airport, managers can educate individuals and answer their questions or concerns. For a broader biometric program, such as a frequent traveler program, individuals may not want to use the technology.This may be for a number of reasons such as fear, misconceptions, invasion of privacy or cultural objections. Another question is to establish who the biometric users are – employees, contractors, or passengers. Again, these will have different application classifiers. What biometric is going to be collected? This may be mandated or recommended by government agencies, due to historical reasons. One such example is fingerprints for background checks. However, if the system is going to be closed, as the system in San Francisco, the airport may choose the specific biometric technology. Other biometric technologies might be mandated or recommended by international bodies. Airport managers might also have to take into consideration device influences, such as sensor and hardware issues – cleaning the sensors due to dirt and smears, residual samples (maybe the case in fingerprinting), sensor qualities, as well as human factors. Another factor for airport managers to consider is throughput rates – at the security check-in for example, there will need to be higher throughput than at an access control gate to the ramp.

Biometric Standards and Airport Security

However, some guides will aid airport managers and those making decisions for biometric implementations. The USA Patriot Act (Act of 2001) in Public Law 107-56 Section 403c outlined the need for technology standard to confirm identity.The International Committee of Information Technology Standards (INCITS), through its proposal IT/01-0917, established a new technical committee with the purpose of developing standards for biometric technologies. As such, new documents within the committee include proposals for biometric data conventions, formats for identification documents and card, application profiles for interoperability and data interchange for the biometrics-based verification, identification of transportation workers, and verification and identification for border crossings.

One of the goals for biometric standardization and aviation industry is being pursued by the Transportation Security Administration (TSA), where the establishment of a common credential across the entire transportation system is envisioned. This credential is envisioned to be used across all modes and for access to secure areas of the transportation system. The goal of this standardized credential is the Transportation Workers Identification Card (TWIC). As it involves identification of an individual, biometric technologies are being examined as part of this framework. According to Lazarick (2002), biometrics within the airport environment will be reference and operational. Reference biometrics will be used at the initial application and re-issuance. Lazarick anticipates the use of fingerprint technology, where the template will be stored on the card and also in a central database. This will, in effect, be a negative identification scenario to prevent alias enrollment. The operational biometric will be used for verification at points of access, and it is envisioned that local airports will select the devices based on technology.As discussed in this article, there are many different types of biometric technologies, each suited for a number of different applications. Therefore, local airports may be made up of a number of different devices, capitalizing on the strengths of each biometric technology at a specific application (access control, perimeter control, etc).

Conclusion

Given the current level of attention on aviation security, biometric technologies may provide an additional level of security that may not be available with traditional tokens or password PIN combinations. Biometric identification will play an important role for students in aviation. As discussed in this paper, there are a number of biometric technologies available within the airline industry.A caveat is that proper understanding of the environment and the application will provide airports with the enhanced security and customer service required by today's traveling public.

References

AfB (Association for Biometrics) & International Computer Security Association (ICSA_). (1998). *1998 Glossary of Biometric Terms.* Retrieved January 25, 2000 from the World Wide Web: http://www.afb.org.uk/public/gloosuk1.html

ASBWG. (2001). *Evaluating Biometrics for airport security*. Washington D.C.

Ashborne, J. (2000). *Biometrics: Advanced identity verification.* New York: Springer-Verlag.

Ben Gurion International Airport. (2000). *Recognition Systems.* Retrieved April 14, 2002, from the World Wide Web: http://www.ingersoll-rand.com/biometrics/07.html

Biometrics. (1999). *Biometrics.* Retrieved March, from the World Wide Web: http://www.scmagazine.com/scmagazine/1999_03/cover/cover.html

Daugman, J. (2001). *High confidence recognition of persons by Iris Patterns.* Paper presented at the IEEE 35th International Carnahan Conference on Security Technology.

Doddington, G., Liggett, W., Martin, A., Przybocki, M., & Reynolds, D. (1998). *Sheep, goats, lambs and wolves. An analysis of individual differences in speaker tecognition performance.* Paper presented at the ICSLP, Sydney, Australia.

Eye Ticket. (2002). Press Release. *Eye Ticket.com.* Retrieved April 15, 2002, from the World Wide Web: www.eyeticket.com

Halici, U., Jain, L. C., & Erol, A. (1999). Introduction to fingerprinting. In L. C. Jain, Halici, U. Hayashi, I., Lee, S. B., Tsutsui, S. (Eds.), *Intelligent biometric techniques in fingerprint and face recognition* (pp. 1-35). Boca Raton, FL: CRC Press.

Howell, A. J. (1999). Introduction to face recognition. In L. C. Jain, Halici, U. Hayashi, I., Lee, S. B., Tsutsui, S. (Eds.), *Intelligent biometric techniques in fngerprint and facial recognition* (pp. 219-283). Boca Raton, FL: CRC Press.

Industry Updates. (2002). Retrieved, from the World Wide Web: http://www.canadiansecuritymag.com/newpages/industry1.html

Jain, A., Bolle, R., & Pankanti, S. (2001). *Biometrics: personal identification in networked society.* Norwell, MA: Klewer Academic Publishers Group.

Lazarick, R. (2002). *DOT Transportation Security Administration's initiatives on employee & passenger identification and air terminal surveillance.* Paper presented at the Biometrics Consortium Conference, Washington, DC.

Newton, H. & Woodward, J. D. (2001). *Biometrics: a technical primer*. Rand, Santa Monica, CA

Podio, F. L. (2001). *Personal authentication through biometric technologies.* Paper presented at the IEEE 4th International Workshop on Networked Appliances.

Rejman-Greene, M. (2001). *Biometrics - real identities for a virtual world.* BT Technology Journal, 19, 3, 115-121.

Soutar, C. (2002). *Finger pattern based interchange format (Version 7)*, Retrieved July 1, 2003 from the World Wide Web: http://www.incits.org/tc_home/m1htm/docs/m1030233.pdf

U. S. Patriot Act. (2001). Public Law (Vol. 107).

UKBWG. (2000). *United Kingdom biometric best practice document.* UK Biometric Working Group, CESG. Retrieved November 23, 2000 from the World

World Wide Web: http://www.afb.org.uk

Visionics. (2002). *Visionics FaceIt ARGUS pilots go live at Dallas-Fort Worth and Palm Beach International Airports*. Visionics Inc. Retrieved April 15, 2002, from the World Wide Web: http://ir.shareholder.com/vsnx/ReleaseDetail.cfm?ReleaseID=70091

Wayman, J. & Alyea, L. (2000). *Picking the best biometric for your applications*. In SJSU Biometrics. San Jose, California: San Jose State University. Retrieved April 14, 2000 from the World Wide Web: http://www.engr.sjsu.edu/biometrics/publications_tech.html

Wayman, J. L. (2000a). A definition of biometrics. In *National Biometric Test Center Collected Works* (pp. 21-25). San Jose, CA: National Biometric Test Center.

Wayman, J. L. (2000b). Fundamentals of biometric authentication technologies. In *National Biometric Test Center Collected Works* (1st ed., pp. 1-21). San Jose, CA: National Biometric Test Center.

Woodward, C. (2002, July 16). Security hassles deter many frequent fliers: Struggling airlines study how to speed process, Placate Their Best Customers. *USA Today*. p. B08.

International Journal of Applied Aviation Studies, 3, 2

Passion! Aviation's Best-Kept Leadership Secret: A Brief Synopsis of Qualitative Inquiry into One Emotional Aspect of Aviation Leadership

Mary N. Kutz
6108 Winfield Drive
Oklahoma City, Oklahoma 73162
E-Mail Address: drkutz@cox.net
Phone: 405-720-9091

Abstract

History is replete with examples of powerful organizations led by powerful people who recognized the value of passion or love for the job in accomplishing the extraordinary. Aviation is no exception. In fact, aviators are known for their passion for flying that sometimes borders on arrogance. Successful aviation leaders have recognized the value of that pride and passion in achieving organizational mission success.

Contemporary research of aviation leadership indicates that aviators frequently see themselves as part of a prideful business that is unique in its history and its achievement. The uniqueness of aviation leadership is attributed to: the volatility and futuristic nature of the business; the safety issues and unforgiving nature of the profession that demands the highest standards; the visibility and public impact of aviation activity; the capital intensive requirements of the industry; the technical skills and single focus sometimes required for success; and the camaraderie and esprit' de corps of the aviation community who share a passion for the industry. This article utilizes the findings of a 1998 study of the Characteristics of Successful Aviation Leaders of Oklahoma to launch a review of the literature and explore the power of passion in leading aviation professionals.

Requests for reprints should be sent to Beverly Laughead, FAA Academy, AMA-502-OU, P.O. Box 25082, Oklahoma City, OK 73125.

Introduction

Passion can be described in a variety of ways ranging from words of emotion such as affection, devotion, and even love, to words of conviction such as enthusiasm or fervor for a topic or profession. The literature on leadership in general uses these and many other synonyms to describe the powerful commitment and emotion expressed as a *"fire in the belly"* kind of experience associated with doing something you love. For purposes of this article, that expression will be used to describe passion in leadership and to explore the power of that passion in the aviation environment.

The emotion of passion can be associated with many professions. It is often more readily associated with professions such as firefighting, military service, medicine, and aviation, all of which require self-sacrifice, high risk, or noble action. However, contemporary research has unearthed some surprising findings related to passion in a variety of other professions including management and leadership. The power of leadership passion coupled with a passion for a high-risk profession such as aviation provides a fertile field for future research.

Even before the Wright Brothers and the early attempts at flight, a fascination with and passion for flight has been a part of the history of aviation itself. That passion manifested itself in the earliest years of flight with barnstormers and their recklessness and willingness to risk their lives to engage in their profession. It has not diminished today, yet little has been done to research the concept of passion in aviation leadership. This article will describe some of the literature pertinent to passion in leadership and examine the impact of combining passion for both aviation and leadership in leading aviation professionals.

Contemporary Research

A beginning, albeit a small beginning, in researching the concept of passion in aviation leadership had its origin in the 1998 findings of a qualitative study of the Characteristics of Successful Aviation Leaders in Oklahoma which surfaced some interesting perspectives on the implications of passion. In that study (Kutz, 1998), 18 aviation leaders from a variety of different aviation specialties were interviewed regarding their background, opinions, and recommendations for future aviation leaders. When asked what if any differences they had noticed in achieving success as a leader in aviation versus another environment, 50 percent of the respondents perceived the aviation leadership role as unique and many of them cited passion as a reason for that uniqueness. The leaders cited examples of their observations of passion at work in the aviation industry and their perception of why aviation leaders often demonstrate a significantly higher level of passion. Their perceptions included:

1. the natural high of a high-risk profession;
2. a pride and certain glamour that comes from the ability to conquer the skies especially in the early years of aviation;

3. the greater emphasis on quality and higher standards necessary in aviation because of the unforgiving nature of the environment;
4. the dynamic nature of the profession both technically and in terms of the business risk and capital intensiveness of the business;
5. the more structured and disciplined nature of the environment;
6. the critical importance of being futuristic and visionary in such a volatile industry;
7. the impact of the overall aviation mission on the public;
8. the complexity of the technical skills necessary to survive;
9. the camaraderie or esprit de corps of aviation professionals; and
10. the sometimes single focus on aviation based on the love for the field.

As the research questions probed more deeply into the concept of motivation and leaders were asked what motivates aviation leaders to perform exceptionally, the most frequently identified motivator of successful aviation leaders was a love for the aviation business, a love for flying and a love for their job and what they do (Kutz, 1998). The Kutz findings were consistent with those of Kouzes and Posner (1995), and other more generic leadership research.

Kouzes and Posner (1995) described, "Love—of their products, their services, their constituents, their clients and customers, and their work" as possibly "the best-kept leadership secret of all" (p.14). This observation followed several years of research in which Kouzes and Posner created a research base of 60,000 leaders and constituents in a variety of occupations worldwide.

Review of the Literature

The Power of Leadership Passion

The concept of passion and love as an act of will or conviction and a source of power is supported in the literature in a variety of ways. That power is a derivative of a number of specific qualities noted in effective leaders.

Active, energetic leadership. Clemmer (1995) lamented the misuse of that power when managers create a "passionless culture of budgets, business plans, strategies, and projects and try to energize people by mission statements, visions, values and other *leader speak*" (¶2). Clemmer's article went on to describe effective leadership as action, not a position. Effective leaders generate action that comes from creating energy through excitement or urgency (¶7).

In his study of hundreds of 'peak performers,' Garfield (Clemmer, 1995) alluded to that energy which produces action when he described preference with a passion or commitment to what they do as one of the single most dramatic differences between peak performers and their less productive colleagues (¶5).

Leadership service. In an article entitled "Starting the Young on Passion for Leadership," Jimenez (2003) cited Maxwell's description of a true leader in "The

21 Indispensable Qualities of a Leader" as someone with "a passion to serve and build up others to the point of sacrificing your own personal wants for their sake and benefit." He described that passion as "loving your people more than your position" (¶15).

Primal Leadership, Realizing the Power of Emotional Intelligence, Goleman, Boyatzis, and McKee (2002), a team of renowned researchers, explored the role of emotional intelligence in leadership and described the discovery of "The Ideal Self" as the point where change begins. When leaders engage only at the intellectual level, it is difficult to maintain the energy and commitment that comes with a deep passion about one's profession. It is that connection with one's dreams that releases passion, energy and excitement and arouses enthusiasm in those who lead. They described the process of reaching deep into one's gut level to develop the ideal image and becoming passionate about the possibilities life holds (pp.115-116).

Peters and Austin (1986) described a seminar where a young undergraduate asked what the most important criterion for success might be. Peters wrote on the board, in letters a foot-and-a-half high, the word "Passion," and proceeded to elaborate that "you gotta love what you do. You gotta care" (p. 339). He later quoted Vince Lombardi, coach of the Green Bay Packers, as saying, "I don't necessarily have to like my associates, but as a man I must love them. Love is loyalty. Love is teamwork. Love respects the dignity of the individual. Heartpower is the strength of your corporation" (p. 341).

Neff and Citrin (1999) of the renowned executive search firm, Spencer Stewart, described their efforts to identify the most successful business leaders in America based upon the most exacting standards imaginable. In their book, *Lessons from the Top, The Search for America's Best Business Leader*, they interviewed each of the fifty leaders regarding their accomplishments, beliefs, careers, etc. and identified ten traits each of the leaders seemed to have in common. The first and most noticeable trait, according to Neff and Citrin, was the passion that the leaders shared for their people and their companies. When asked for their advice to young people, they almost unanimously advocated loving what you do (pp. 379-380).

Conviction. In *Eagle's Flight, Learning that Powers Performance*, an article entitled "Conviction Must Include Passion," passion is described as an inner fire or an emotion that "carries conviction…from the heart of one person…to the inner self of others" (2002, ¶4). It further implored leaders not to allow their convictions to motivate only themselves, but "through the magic of passion, your visible excitement, allow them to be passed on to those looking to you for leadership" (¶5).

Aviation Leadership Passion

Lindbergh (1953) described the passion for aviation as "a love for the air and sky and flying, the lure of adventure, the appreciation of beauty. It lay beyond the descriptive words of men—where immortality is touched through danger, where life meets death on an equal plane; where man is more than man" (p.3).

Petzinger (1995), in *Hard Landing, The Epic Contest for Power and Profits That Plunged the Airlines into Chaos*, described flying as "an act of conquest, of defeating the most basic and powerful forces of nature" (p. xvii). It involves the power to wipe out hundreds of lives with one mistake.

The margin of error in the industry itself is thin with the ability to implode and wipe out a long list of companies that once were going strong such as Pan Am, Eastern, Braniff, and many others. Petzinger described it this way. "The union of devilish details and "godlike power," as Lindbergh found in the act of flying, makes commercial aviation compelling for yet another reason: the anthropology of the executive suite.... The men who run the airlines of America are an extreme type; calling them men of ego would be like calling Mount McKinley a rise in the landscape" (p. xix). He further describes these executives as men "of an age and of a type..all of them stricken with the same infatuation with aviation and all of them committed to achieving personal triumph" (Pettzinger, 1995, p. xxi).

Nowhere is the love or passion for aviation and aviation leadership more apparent than at Southwest Airlines which has built its phenomenal success in the industry not only on their passion for aviation and for their company but for their love for their customers, co-workers, the aviation public and people in general. Freiberg and Freiberg (1998), in their book entitled "*Nuts! Southwest Airlines' Crazy Recipe for Business and Personal Success*," devoted an entire chapter to the concept of love at Southwest. They described the LUV ticker symbol at Southwest not just an advertising theme but a symbol of the loving character of the company that permeates all levels of the organization as well as the hiring practice based on people who can externalize or focus on other people. They recognized that even though "macho managers might never use the word 'love' in the workplace, most would admit that, when it comes right down to it, the deepest need in human existence is the need to be loved and accepted and that need does not mysteriously disappear when employees walk through the door at work" (p. 217). Southwest believes love is an act of will, something you do and that love permeates the organization all the way to its leadership and the CEO level where it is reflected in the actions of Herb Kelleher as he helps load baggage on Thanksgiving or refuses to change his schedule when he has a commitment to Southwest employees (p. 221).

Are there circumstances when passion, love, energy, and drive to action can impact aviation leadership in a negative way? Does a passion for aviation necessarily equate to a passion for leadership?

I Just Want to Fly

Aviation students frequently share the passion for flight, which is often expressed in frustration with the need for courses that develop skill in communication and leadership as well as other soft skills. The student perceptions are that they will find a career that lets them operate solely within their passion for aviation. Much to their surprise, they discover that human communications, team leadership, and technical skills are essential even in the cockpit. Because of their technical skills, they more than likely will be thrust into organizational leadership roles and the transition from a passion for aviation to a passion for leadership is sometimes difficult.

Unrealistic Expectations and Impatience with Perceived Underperformance. Passion for aviation can even become a negative in leadership when there is a tendency toward impatience and perceived disrespect for the talents of non-pilots in the organization. In *The Leadership Team: Leading Lines. Coaching to Help You Lead and Succeed*, Geisler, Dunlap, Favre, and Johnson (2001) described feedback where managers were commended for their passion but were also described as brusque, easily frustrated, impatient, all of which were descriptors of poor interpersonal skills. It seemed that the evaluators were "sending a message to their leaders that their passion may explain why you treat us this way, but it doesn't excuse it; we want your passion channeled in more positive ways" (¶3).

Passion for Aviation versus Wise Decision Making. Aviation management textbooks often caution students about the impact of their passion for aviation that gets in the way of wise decision-making in a management role. Decisions driven by a love for aviation rather than a passion for wise management of the company can severely impact the effectiveness of the organization. For example, capital investments in new aircraft may be excessive if a decision to purchase is driven by a fascination with new technology that outweighs wise business practice.

Petzinger (1995) described aviation leaders whose passion for the industry and for the job contributed not only to their personal demise but to that of the company they led.

> "….in the end most of these men were exiled from the executive suite…when that same overwhelming ambition that drives so many executives to the top also assures their failure; that when executives form emotional attachments in business, whether to people, markets or machines, they deprive themselves of their best business judgment; that those who know an industry best

are the most likely to take for granted, and ultimately ignore, its most inviolate principles; that although the rebuke may be slow in coming, greed in the end, is almost always punished; that economics in short overpowers ego. Just as they reflect the excesses of business in so many other respects the airlines bespeak these lessons in spades" (p. xxii).

Conclusions and Recommendations

The concept of passion for aviation coupled with a passion for aviation leadership and the subsequent dynamics of leadership that could harness the combined power of both emotions deserves much more extensive study.The Oklahoma State University study (Kutz, 1998) simply exposed the deficit of understanding and the potential for a line of inquiry into a powerful leadership tool for future generations of aviation leadership students. The power of passion and the potential for subsequent problems associated with that passion are fertile fields of further inquiry. In the meantime, it is important to recognize that emotions such as passion and love are not mutually exclusive to success in leadership. On the contrary, preliminary research and actual organizational examples such as Southwest Airlines demonstrate the importance of these emotions to success in aviation leadership.

References

Clemmer, J. (2000). Leaders arouse passion and create energy. In J. Clemmer, *Pathways to performance; a guide to tranforming yourself, your team, and your organization.* Rocklin, CA: PRIMA Lif4estyles. Retrieved July 2003 from http://www.tcm.com/clemmer/c30leaders.htm.

Eagle's Flight. (2002, July). Conviction must include passion. *Creative Training Excellence Learning That Powers Performance.* Retrieved July 2003 from http://www.eaglesflight.com/leadership/conviction_passion.html

Freiberg, K. & Frieberg, J. (1998). *Nuts, Southwest Airlines crazy recipe for business and personal success.* New York: Broadway Books.

Geisler, J., Dunlap, L., Favre, G. & Johnson, P. (2001, July 9). The leadership team: Leading lines, coaching to help you lead and succeed. *Poynter.org Poynter Seminars & Info.* Retrieved January 2003 from http://www.poynter.org/lead/070901.htm

Goleman, D., Boyatzis, R. & McKee, A. (2002). *Primal Leadership, Realizing the Power of Emotional Intelligence.* Boston, MA: Harvard Business School Press.

Jimenez, B. (2003, January 26). *Starting the Young on Passion for Leadership*. Retrieved January 2003 from http://www.inq7.net/lif/203/jan/26/lif_10-1.htm.

Kouzes, J. M. & Posner, B. Z. (1995). *The Leadership Challenge.* San Francisco, CA: Jossey-Bass Publishers.

Kutz, M. N. (1998). C*haracteristics of successful aviation leaders of Oklahoma*.

Unpublished doctoral dissertation, Oklahoma State University, Stillwater, OK.
Lindberg, C. (1953). *The Spirit of St. Louis*. New York: Scribner Classics.
Neff, T.J. & Citrin, J.M. (1999). *Lessons From the from the Search for America's Best Business Leaders*. New York: Doubleday.
Peters, T. & Austin, N. (1986). *A Passion for Excellence, The Leadership Difference*. New York: Warner Books.
Petzinger, T. (1995). *Hard Landing. The Epic Contest for Power and Profits That Plunged the Airlines Into Chaos*. New York: Times Books, Random House, Inc.

International Journal of Applied Aviation Studies, 3, 2

Book Reviews

Psychological Perspectives on Fear of Flying edited by Robert Bor and Lucas van Gerwen

Ashgate Publishing Limited
Aldershot, Hampshire
England

Reviewed by
Todd P. Hubbard
The University of Oklahoma
Norman, OK

The 251-page textbook is a pleasant departure from the dozens of self-help books that promise relief from fear, but never tell you if their advice really works. Robert Bor and Lucas van Gerwen treat their readers to a buffet of clinical studies, therapy critiques, technological aides, suggestions for air carrier personnel, and therapy success stories. Most of the studies followed a survey research format, rather than an ethnographic, empirical approach, since the aim of the research studies was not to study a population of phobics, but rather to address the root cause of the phobia and apply meaningful therapeutic strategies. Data collected were generally of the ordinal scale variety, analyzed by non-parametric statistics. Therefore, most of the statistics were descriptive in nature and generalizations were more meaningful for the populations sampled rather than for the greater population of those with a fear of flight. Each study concluded with a list of limitations or recommendations. In all cases, the authors were honest about their results and did not overemphasize their meaningfulness.

Sample populations used in these studies tended to be potential passengers on commercial aircraft who suffered from one or more phobias. Fear of flying in pilots or aircrew members, caused by traumatic aircraft accidents, was not addressed.

From the outset, this text puts to rest any notion that the contents are anecdotal accounts or that the authors are largely unknown souls doing poor work.

Requests for reprints should be sent to Beverly Laughead, FAA Academy, AMA-502-OU, P.O. Box 25082, Oklahoma City, OK 73125.

Before the preface, which was written by such notables as Robert Wolfger and Josine Arondeus, the reader is presented with summaries of each author's work and credentials. With the aid of some foreshadowing of text's content in the introduction, the chapters were gift wrapped with thoughtful introductions and informative conclusions. If a format is suggested, it is that the reliability studies for the flight anxiety situations (FAS) questionnaire, the visual analogue flight anxiety scale (VAFAS), and the flight anxiety modality (FAM) questionnaire are presented in the first half of the text, allowing the editors to place companion studies in the second half of the text. Occasionally, the editors interleaved a position paper among the studies, perhaps to break up the steady flow of statistical information. Placing *Flight Crew Involvement* (Chapter 14) and *Putting Fear to Flight* (Chapter 15) at the end of the text was a brilliant move. Chapter 14 gave air carriers something to do and Chapter 15 ended on the high note of success.

Forty-two percent of the contributors had ties to the VALK Foundation, a collaborative venture between the University of Leiden, KLM, Royal Dutch Airlines, and the Amsterdam Airport Schiphol. Out of 15 chapters in the text, six were written by those associated with the foundation. The VALK Foundation (http://www.valk.org) has been in operation since 1989 and has continued to perfect the scientific basis for therapies related to the cessation of fear of flying. Clearly, the VALK Foundation is making a difference.

European authors are in the majority in this text. Exposure to European studies expands the reading list for most American readers. Comparison of fear of flying treatment protocols, among practitioners in the United States, Australia, and Europe, give readers an idea of where the western world is currently focused. Although two contributors were from Australia, the text does not represent the possibility of alternative approaches to treatment for fear of flying from Asia, Africa, South America, or the Near East. Future texts on this subject should include the effect of culture on treatment methods.

Robert Bor and Lucas van Gerwen, in the introduction to their text, characterized a portion of their target audience as psychologists, psychiatrists, pilots, air carrier employees, law enforcement, clergy, medical doctors and nurses, and rescue workers engaged in the many aspects of fear of flying awareness and treatment. However, it appears that most of the chapters were written for psychologists or those directly involved in treatment (doctors, nurses, flight attendants, and clergy). Pilots, air carrier employees, law enforcement, and rescue services personnel might appreciate the topics in Chapters 2, 3, and 14, but might not appreciate or fully understand the implications of the rest of the text without tutorials on statistics, psychological therapies, and jargon used in this field.

Those suffering from fear of flying might also benefit from this text, not because it provides aid to their suffering, but because there is evidence that some treatments have a higher success rate than other treatments. Today's patients like to know that there is a return on investment for treatment. Many of the programs are costly, and as Mandy said in Chapter 15, "I'm not doing anything else because it's cost me a fortune and I can't afford any more" (p. 238).

The text has utility for the classroom, as a companion text. The content might not be lengthy enough for a stand-alone text, but if paired together with other research studies in the same field, it would give the students a foundation from which to launch into discussion and further study. Instructors will find it easy to rearrange the chapter sequence as best fits their classroom needs.

At the close of this review, special attention should be given to the editors. Readers often judge an edited text by the strength of credentials of its editors. They will not be disappointed by what they discover about Robert Bor and Lucas van Gerwen. The text is a masterful work of scientific research on fear of flying, well packaged, and thoughtfully presented. Buy one for your library or use it as a course text. You will not be sorry.

International Journal of Applied Aviation Studies, 3, 2

Passenger Behaviour *edited by Robert Bor*

Ashgate Publishing Limited
Aldershot, Hampshire
England

Reviewed by
Raymond E. King
Civil Aerospace Medical Institute
Mike Monroney Aeronautical Center
Oklahoma City, OK

It is often said that it is a small world. That is certainly the case in the world of aviation psychology. Before being asked to review *Passenger Behaviour*, I had reviewed the proposal for the book by request of the publisher. Despite my initial doubts, Robert Bor, the volume's editor, has pulled off an ambitious project. American readers, however, are cautioned that it contains numerous examples of Winston Churchill's assertion that the English and the Americans are separated by a common language. For example, I had no idea what a "hen weekend" was, so to clue me in, I had to find a bona fide Brit (no easy feat in Oklahoma). Such distractions, however, are a very small price to pay for the wisdom of the chapters and lend to its charm. There are some deeper cultural issues, however, that may prove a little more cumbersome. I happened to sit next to a travel agent from Calgary at the SkyDome in Toronto who explained to me the concept (and popularity) of package tours in Canada and Europe, a marketing technique infrequently used in the United States. Having dispensed with these minor issues, let's now consider specific aspects of *Passenger Behaviour*. Most readers will not read this book from cover to cover the way I did to prepare this review and will instead use it as a reference tool. Nevertheless, each section deserves careful consideration.

The reader is first impressed with the stable of talent that Bor has assembled to produce this book. The contributors are listed alphabetically; Bor is listed first as a function of his last name being first alphabetically. The contributors include a lawyer, travel consultants, medical doctors, and research and clinical psychologists, and many other specialists.

Requests for reprints should be sent to Beverly Laughead, FAA Academy, AMA-502-OU, P.O. Box 25082, Oklahoma City, OK 73125.

Bor sets the stage with his excellent introduction. He informs us that the book is intended primary for flight and cabin crews and others who work in commercial aviation. This reviewer would be remiss not to point out a glaring error in the introduction (page 6): Flight 93 did not crash in a field in New Jersey; it took off from Newark, New Jersey and crashed into a field in the neighboring state of Pennsylvania. Similar to the contributor Simon Calder who "was born two miles from the runway at Gatwick Airport," I was born within two miles of the runways at Newark. To their credit, all of the other contributors get this detail right.

Glenvil Smith provides a brief but comprehensive chapter on the legal aspects of passenger behavior. He successfully avoids the perils of writing about laws that do not pertain to the majority of readers and instead focuses on the issues of interest to an international audience. Iain B. McIntosh convincingly maps the genesis of air rage as a manifestation of flight-related stress and over-indulgence of alcohol. Despite writing a comprehensive and well thought-out chapter, he errs when he suggests that air travel "continues to increase" (page 17) despite the events of September 11. Other contributors are more accurate on this point. Alex Cruz and Linda Papadopoulos provide a chapter on the evolution of the airline industry and the consequences on passenger behavior.The requirements for the early "air hostesses" (page 34) are particularly amusing. While they had to be nurses and weigh less than 115 pounds, their duties included baggage handling and refueling the aircraft! The reader is introduced to a bygone era of passenger aviation that was expensive, exclusive, luxurious, and sometimes unreliable. Nevertheless, passengers were on their best behavior, and usually in their best clothes. What happened to this exclusivity? The authors point to the revolutionary low-cost carrier, "People's Express." As a side note, the airline was actually named *People* Express, another fact I know from my childhood proximity to Newark (now Liberty) International Airport. The authors also delineate a theory of air rage. The root causes include the deterioration of cabin service, cabin crew profile, smoking bans, flight delays, and the profile of the passengers themselves (page 41).

Elaine Iljon Foreman gives us an overview of the fear of flying. She contends that fear of flying is actually a cluster of disorders to include fear of crashing, heights, confinement, instability, panicking, and lack of control. This chapter is of great interest to the more psychologically minded reader as it delves into some pretty heavy theories of cognitive behavioral psychotherapy. I enjoyed it, but I'm a clinical psychologist and I eat this sort of stuff right up. Other readers may wish to skim this chapter. Julia Heller continues this clinical trend by writing about psychological and psychiatric difficulties among airline passengers. She asserts on pages 60 and 61 that "medical consultation was sought on approximately 24 per cent of U.S. flights." It just can't be accurate that medical consultation is needed on a quarter of all U.S. flights. In any case, those who are forced to deal with passengers' anxiety while at 35,000 feet will find this chapter to be of some use.

Olga Levitt and Robert Bor explicitly deal with a topic that the other contributors also touch on when they consider air travel and the implications for relationships. They succinctly state: "Transitions perturb people (page 66)." Their exploration of Bowlby's attachment theory is not for those uninterested in psychology. Susanne Robbins introduces us to homesickness. She points out "passengers on the same flight will typically be flying for very different reasons and will experience the flight in very different ways (page 81)." While members of various cultures may differ in their experience of homesickness, Robbins maintains that couples saying goodbye all bear resemblance to children being separated from their primary caregivers.

Angela Dahlberg, in considering air rage, takes issue with the concept of an expectation that passengers assist when things go awry in the cabin. She maintains that such an expectation is not in line with the promises of safe transportation. She asserts that new selection techniques are needed to staff the cabin to address newly emerged security concerns.

Lauren J. Thomas writes about passenger attention to safety information and concludes, on the basis of published survey data, that many passengers simply to not pay attention to the information that is presented to them. Ignoring this information, which passengers assume is just redundant with information they have heard on other flights, puts themselves and others at risk. She specifically addresses the duties of those of us who tend to populate the emergency row exits and notes the counterintuitive nature and difficulty of operating a Type III exit hatch. This reviewer has had occasion to operate this type of exit hatch, during a training exercise, and could not agree more completely. Ed Galea, in the largest chapter of the book, expands upon the topic of passenger behavior in emergency situations and notes passengers' tendency to exit from the door they boarded. He notes the research facilities at the Civil Aerospace Medical Institute (CAMI) and Cranfield University but notes that real-life evacuations differ from laboratory studies due to social bonding – passengers attempt to reunite with separated traveling companions. The reader is advised that the author uses the acronym CAMI twice before defining it. This chapter makes extensive use of photographs and other illustrations.

Man Cheung Chung notes the devastating impact aircraft disasters have on the communities in which they occur in his interesting chapter on the psychological impact of aircraft disasters. He also notes the trauma that rescue and recovery workers face, particularly when children are among the victims.

Margaret A. Wilson changes the focus of the book a bit by focusing on hostage situations aboard aircraft. Drawing upon a rich history of hijacking events, Wilson delineates likely outcomes of hijacking scenarios suggesting that hijackers and their victims follow preordained scripts. Wilson finally concedes that the events of September 11 may have changed all the rules of behavior in future hijacking events.

The book then takes a medical turn and addresses the issues of the physiology of flying (Richard Dawood), illness among passengers (Jane N. Zuckerman), and travel fatigue and jet lag (Jim Waterhouse, Thomas Reilly, and Ben Edwards). These are useful chapters for passengers and aircrew alike as they may help prepare humans for the unique adaptations required for existing in a small space, at altitude, for extended periods of time and then finding themselves in a time zone not in synchrony with their body clock Smokers and liberal alcohol imbibers, in particular, will find useful information.

Next Peter Jones and Margaret Lumbers consider appetite and in-flight catering. Their chapter provides inside information on the logistics of serving a hot meal in the setting of an aircraft. Their treatment of thirst and its management is equally fascinating. Stephen Clift tackles the topic of sex and international travel. This reviewer thought he would spend more time on the type of behaviors some passengers engage in while flying. Rather, Clift focuses on those individuals (the authors note that this tendency is not limited to male travelers) who fly specifically to reach locations where they may engage in sex more freely than at home ("sex tourism"). Such behavior has public health implications due to the spread of disease. These practices are made even more dangerous and exploitive due to the practice of importing women and girls from third world countries expressly for sexual purposes. Finally, Simon Calder addresses the status of civil aviation. He notes that some airlines have de-evolved to the point where they promise nothing - and then deliver it. Dealing with modern airports is no more relaxing for the traveling public. Calder concludes by stating: "violent or aggressive behavior by passengers can never be justified, but the reasons for it can be understood, and steps taken to reduce the causes of disruption" (to include improved service and a reduction in the amount of alcohol offered to passengers, particularly those in the premium classes).

Bor's book will not be considered light reading by anyone lucky enough to have an opportunity to study the wisdom contained within its covers. As noted, it is also not perfectly consistent from chapter to chapter. For those of us who fly on a regular basis, however, obtaining and studying this book promises to return substantial

International Journal of Applied Aviation Studies, 3, 2

The U.S. Air Service in the Great War, 1917-1919 by James J. Cooke (199f)

Praeger Books, Westport, CT

Reviewed by
Mark Sherman and Deak Arch
Ohio University
Department of Aviation

In general, the role of aviation is only briefly touched upon in historical treat ments of World War I, also referred to as the Great War. Historians focus primarily upon causative events leading up to war, postwar diplomacy including Wilson's fourteen points, and terms Germany was forced to accept under the Versailles Peace Treaty. Because of the role this treaty played in leading to a Second World War, positive contributions of World War One Aviation has not been accorded the study it deserves.

Dr. Cooke successfully used multiple primary, secondary, unit histories, and periodical articles while attempting to detail contributions by the United States Air Service to the Great War. As stated in his introduction, this work started with research into the Rainbow Division, composed of National Guard units from 26 states and the District of Columbia. The four infantry regiments were respectively 165th (formerly New York's 69th), 166th (formerly Ohio's 4th), 167th (formerly Alabama's 4th), and 168th (formerly Iowa's 3rd). It was during his Rainbow Division research Dr. Cooke realized the division usually contained some form of aviation assets: either pursuit aircraft or observation balloons, or both. He also states the official four-volume *U.S. Air Service in World War One* contained a wealth of information. However, this resource failed to appreciate and tie together the larger picture of how operational aviation assets (pursuit, observation, and bombardment squadrons) logistically operated with maintenance, supply, and U.S. flight training occurring in England, France, and the United States.

Requests for reprints should be sent to Beverly Laughead, FAA Academy, AMA-502-OU, P.O. Box 25082, Oklahoma City, OK 73125.

Dr. Cooke attempts to integrate all aspects of aviation together by concentrating only upon American Air Power and its use toward a successful allied prosecution of the war. Before the war, aviation in America was a novelty most individuals regarded as foolish with only a few individuals, primarily the Aero Club of America, calling unsuccessfully for aviation to be adapted for military needs. The work begins by outlining American Congressional and War Department attitudes towards military aviation. After American entrance into the Great War in 1917, military aviation assets existed but not in sufficient numbers to support American Ground Forces against German Airpower. Dr. Cooke takes the reader on a journey from America's early period of aviation unpreparedness to creation of a strong aviation arm within the U.S. Military. Once Congress approved the aviation wing of the U.S. Military, it had to be organized, schools constructed, pilots trained, equipment transported to Europe and further training conducted in the war zone. This effect began with few assets and less experience.

The bias aviation encountered by traditional military combat arms is presented along with the attempt by aviation-minded individuals to work with traditional combat arms. If not for attempts by aviation-minded individuals like Billy Mitchell, Frank Lahm, and individuals who previously had flown with either the Royal Flying Corp or Lafayette Escadrille, the American Air Service efforts, finding itself in a new technological war, would have been greatly hindered. This is evident with the discussion of air operations during the Marne Counteroffensive, St. Mihiel Offensive, and battle for Meuse-Argonnne. The author does not stop here but takes the reader through both demobilization and occupation duty.

This book is significant in the fact that it explores World War One Aviation, not from a pilot or unit perspective along with the glory those publications tend to immortalize, but from a military command and control perspective. Subject matter such as this is seriously needed in Great War Aviation history. Mature scholars with an interest in American Air Power organization during World War I, Aeronautical logistics, or military command and control functions should read this book.

It is readily apparent both students and staff in attendance at the U.S. Air Force Air War College would utilize this work, rich with unit numbers and statistics, which non-scholarly and individuals without previous military experience might find tedious. Having retired from Naval Aviation, I feel this military experience is essential for comprehension of each political decision or rational for each command and control policy adopted during the war.