A BIBLIOGRAPHIC DATABASE FOR THE HISTORY OF PILOT TRAINING SELECTION

William R. Howse
Diane L. Damos
Damos Aviation Services, Inc.

Sponsored by
HQ AFPC/DSYX & HQ AF/A1PF
Mr. Kenneth L. Schwartz
Strategic Research and Assessment Branch

July, 2011

Air Force Personnel Center
Strategic Research and Assessment
HQ AFPC/DSYX
550 C Street West, Ste 45
Randolph AFB TX 78150-4747

Approved for Public Release. Distribution Unlimited

UNCLASSIFIED
When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report was cleared for release by HQ AFPC/DSYX Strategic Research and Assessment Branch and is releasable to the Defense Technical Information Center. The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report is published as received with minor grammatical corrections. The views expressed are those of the authors and not necessarily those of the United States Government, the United States Department of Defense, or the United States Air Force. In the interest of expediting publication of impartial statistical analysis of Air Force tests SRAB does not edit nor revise Contractor assessments appropriate to the private sector which do not apply within military context. Federal Government agencies and their contractors registered with Defense Technical Information Center should direct request for copies of this report to:


Available for public release. Distribution Unlimited. Please contact AFPC/DSYX Strategic Research and Assessment with any questions or concerns with the report.

This paper has been reviewed by the Air Force Center for Applied Personnel Studies (AFCAPS) and is approved for publication. AFCAPS members include: Senior Editor Dr. Thomas Carretta AFMC 711 HPW/RHCI, Associate Editor Dr. Gregory Manley HQ AFPC/DSYX, Dr. Lisa Hughes AF/A1PF, Dr. Paul DiTullio AF/A1PF, Kenneth Schwartz HQ AFPC/DSYX, Johnny Weissmuller HQ AFPC/DSYX, Dr. Laura Barron HQ AFPC/DSYX, Dr. Mark Rose HQ AFPC/DSYX, and Brian Chasse HQ AFPC/DSYX.
A Bibliographic Database for the History of Pilot Training Selection

Howse, William R.
Damos, Diane L.

Damos Aviation Services, Inc.
36303 N. Old Woods Trail
Gurnee, IL  60031-1683

AFPC/DSYX
550 C Street West, Suite 45
Randolph AFB, Texas 78150

Approved for public release; distribution is unlimited.

Task was performed under a Scientific Service Agreement issued by Battelle Chapel Hill Operations
50101 Governors Drive, Suite 110, Chapel Hill, NC  27517. Under contract to U.S. Army Research Office.

This report documents the development of and accumulation of a set of reference materials covering the history of U.S. aviator selection from the inception of military applied flight to the present day. It presents the materials so that they may be applied as a reference database or as a complete digital library for use by researchers in the field. The two DVDs containing the referenced material herein are maintained at AFPC/DSYX.
ACKNOWLEDGMENT

This work was supported by the Air Force Personnel Center (AFPC/DSYX) under the auspices of the U.S. Army Research Office Scientific Services Program administered by Battelle (Contract No. W911NF-07-D-0001, Delivery Order 0906). We wish to thank Dr. R. Bruce Gould, Mr. William C. Barker, Dr. Thomas R. Carretta, Mr. Kenneth Schwartz, Mr. Johnny Weissmuller, and Mr. Brandon Spillers for their help in developing this database.
CONTENTS

INTRODUCTION ...................................................................................................................................... 1
DATABASE DESCRIPTION .................................................................................................................... 1
COPYRIGHT LIMITATIONS ................................................................................................................... 6
USER INSTRUCTIONS .......................................................................................................................... 6
REFERENCES ........................................................................................................................................ 8

APPENDICES
   A. Basic Definitions .......................................................................................................................... 9
   B. Listing of the Bibliographic Database Entries With Abstracts .................................................... 10

TABLES

TABLE 1. LIST OF SOURCES FOR ENTRIES IN THE BIBLIOGRAPHIC DATABASE........3
A BIBLIOGRAPHIC DATABASE
FOR THE HISTORY OF PILOT TRAINING SELECTION

Introduction

In August 2009 Damos Aviation Services (DAS) accepted Subcontract TCN 09216 from Battelle Columbus Operations under the U.S. Air Force Research Office Scientific Services Program, Contract W911NF-07-D-0001 / Delivery Order 0906. This report describes and documents the bibliographic database developed and delivered as a result of that subcontract.

In previous work funded by U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) under an Army contract (DASW01-03-D-0016, Selection Instrument for Flight Training) DAS initiated a review and analysis of aviator selection in the United States and other countries which led to a technical report on aviator selection prior to 1925 (Damos, 2007). In a subsequent effort, funded by ARI under Air Force contract FA3089-06-F-0580 (Human Resources Research Database and Strategic Planning Support), DAS produced a bibliographic database of sources from WWI to the present. DAS continued to survey the literature and assemble relevant documents after completion of the original database.

Database Description

For a variety of reasons, including resource limitations, lack of interest, lack of motivation, or plain ignorance, there is a tendency today for agencies to reduce or eliminate support for the operation of small libraries and reference collections. When support is withdrawn there is often little consideration given to the preservation of such collections. The bibliographic database developed under this project preserves a large number of references and their associated files that may otherwise have been lost or destroyed. Even if the product of the present effort serves as no more than an archive it will prove valuable to future research and development.

The bibliographic database contained herein was formed by reviewing objects available in local tangible formats and records accessible through linkages on the Internet. Information from these was entered into the database through a database management software package to form records. Where possible, the full-text objects corresponding to these records were acquired and linked to their entries. Review of a large proportion of the entities provided paths to other, previously unknown entities. For remote sources where data retrieval capabilities existed (some data retrieval capability exists with nearly every Internet resource) searches were performed on all available data fields using keywords that are commonly associated with aviator selection practices.

The entries in the reference database are listed in author order in Appendix B. Note that the first 57 entries have no author listed.

Limitations. Objects entered as database records are associated with pilot selection in the United States Armed Forces from the inception of the employment of aviation by the U.S. military to the present day. There are some records that are primarily relevant to commercial aviation, military
pilot selection in other countries, and to general knowledge, skill, ability and other characteristics (KSAO) assessment for occupational or training selection. These are included because they bear close association with U.S. military pilot selection and are likely to be of use to researchers in the area. Web pages are not included for two reasons: Few are primary sources of information and most are ephemeral. That is, there is a high likelihood that a cited web page will not exist in the near future.

Sources. The entries in the database were retrieved from 46 documented sources. These are listed in Table 1 along with their Uniform Resource Locators (URL). In addition to these sources a large portion of the corpus is derived from personal holdings of the authors and from documents literally rescued from discard. A full set of the Army Air Forces Aviation Psychology Program Research Reports (commonly known as the Blue Book) series was retrieved by Dr. R. Bruce Gould from a dumpster at Brooks Air Force Base. A large number of documents was retrieved from the scientific reference files at the Rotary Wing Aviation Research Unit of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI). These were scheduled to be shredded and recycled because the building in which they were housed was transferred to another agency and no library would accept the files. Since then the remaining documents from that collection were transferred to the Redstone Scientific Information Center (RSIC) at Redstone Arsenal, AL.
Table 1. List of sources for entries in the bibliographic database.

<table>
<thead>
<tr>
<th>SOURCE NAME</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Force Human Resources Laboratory</td>
<td>None</td>
</tr>
<tr>
<td>American Institute of Aeronautics and Astronautics (AIAA)</td>
<td><a href="http://www.aiaa.org/">http://www.aiaa.org/</a></td>
</tr>
<tr>
<td>American Memory</td>
<td><a href="http://memory.loc.gov/ammem/browse/index.html">http://memory.loc.gov/ammem/browse/index.html</a></td>
</tr>
<tr>
<td>American Psychological Association (APA)</td>
<td><a href="http://www.apa.org">http://www.apa.org</a></td>
</tr>
<tr>
<td>Army Doctrine and Training Publications</td>
<td><a href="http://www.army.mil/usapa/doctrine/Active_FM.html">http://www.army.mil/usapa/doctrine/Active_FM.html</a></td>
</tr>
<tr>
<td>Army Heritage Collection Online</td>
<td><a href="http://www.ahco.army.mil/site/index.jsp">http://www.ahco.army.mil/site/index.jsp</a></td>
</tr>
<tr>
<td>Damos Aviation Services</td>
<td><a href="http://www.damosaviation.com">http://www.damosaviation.com</a></td>
</tr>
<tr>
<td>Electronic Code of Federal Regulations</td>
<td><a href="http://ecfr.gpoaccess.gov/cgi/t/text/textidx?c=ecfr&amp;sid=3efaad1b0a259d4e48f1150a34d1aa77">http://ecfr.gpoaccess.gov/cgi/t/text/textidx?c=ecfr&amp;sid=3efaad1b0a259d4e48f1150a34d1aa77</a></td>
</tr>
<tr>
<td>Google Books</td>
<td><a href="http://www.google.com/books">http://www.google.com/books</a></td>
</tr>
<tr>
<td>Human Factors and Ergonomics Society (HFES)</td>
<td><a href="http://www.hfes.org">http://www.hfes.org</a></td>
</tr>
<tr>
<td>Ingenta</td>
<td><a href="http://www.ingentaconnect.com">http://www.ingentaconnect.com</a></td>
</tr>
<tr>
<td>International Military Testing Association (IMTA)</td>
<td><a href="http://www.internationalmta.org/">http://www.internationalmta.org/</a></td>
</tr>
<tr>
<td>International Symposium on Aviation Psychology (ISAP), Association for Aviation Psychology, Proceedings</td>
<td><a href="http://www.wright.edu/isap">http://www.wright.edu/isap</a></td>
</tr>
<tr>
<td>Internet Archive</td>
<td><a href="http://www.archive.org">http://www.archive.org</a></td>
</tr>
<tr>
<td>Joint Doctrine, Education and Training Electronic Information System</td>
<td><a href="https://jdeis.js.mil/">https://jdeis.js.mil/</a></td>
</tr>
<tr>
<td>Library of Congress</td>
<td><a href="http://www.loc.gov/">www.loc.gov/</a></td>
</tr>
<tr>
<td>Library, Information Science &amp; Technology Abstracts (LISTA)</td>
<td><a href="http://www.libraryresearch.com/">http://www.libraryresearch.com/</a></td>
</tr>
<tr>
<td>National Aeronautics and Space Administration (NASA) Technical Reports</td>
<td><a href="http://ntrs.nasa.gov/">http://ntrs.nasa.gov/</a></td>
</tr>
</tbody>
</table>
Quantification. The database currently comprises 1,805 entries represented in 20 different object types. These object types are as follows:

Audiovisual Material
Book
Book Section
Chart or Table
Conference Paper
Conference Proceedings
Edited Book
Figure (Photograph)
Government Document
Journal Article
Legal Rule or Regulation
Magazine Article
Manuscript
Pamphlet

Pentagon Library http://www.hqda.army.mil/library/m.htm
PsychInfo http://www.apa.org
RAND Technical Reports http://www.rand.org/pubs.html
Scitopia http://www.scitopia.org
U.S. Army Military History Institute (AMHI), Historical Services Division http://www.carlisle.army.mil/ahec/MHI.htm
U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) – Fort Rucker Research Unit holdings None
U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) Technical Library http://arit.sirsi.net/uhtbin/cgi-sirsi/x/0/0/49
Web of Science http://science.thomsonreuters.com/training/wos/
WorldCat http://worldcat.org
The publication dates span a range of 149 years from 1862 to 2011. There are nine entries for which there is no known year of publication. For one entry the year of publication is known only to the nearest decade. There are 5,005 authors cited and 1,391 keywords listed. There are 57 entries that have no author listed. The database contains citations for articles from 479 journals. Ninety-three entries have associated URLs. There are two entries whose associated documents are distribution restricted (so labeled in the Notes attribute for those entries) although the database entries themselves are not restricted. No entries are classified or sensitive. There are 18 entries that are considered to be incomplete. These are included in a subgroup in the database management software files. There are 721 object files linked to the entries. Most of these are digitized full-text representations of the original work or digitized photographs. There are a number of instances where more than one file is linked to an entry. This is because the size of the work makes it difficult to reliably operate the scanning and interpretation software employed.

Scanning and Optical Character Recognition. There are a large number of entries for which hard copies were accessible but for which digitized copies were not. These were scanned into digitized files (Portable Document Format – pdf) and Optical Character Recognition (OCR) procedures were applied. The success level of OCR software is dependent on the quality of the image being interpreted. If the original image was degraded or noisy, then the frequency of interpretation errors is increased. There are a number of files that are scanned but OCR success was so minimal that they have been left uninterpreted. The internal content of these files can not be automatically searched.

Special mention should be made of the Army Air Forces Aviation Psychology Program Research Reports (Blue Book) series. The digitized representations of these works are derived from two sources: Xerographic copies of some volumes that were in the ARI holdings and digital photographic images of volumes made by Brandon Spillers from originals possessed by the U.S. Air Force. The Xerographic copies are of mediocre quality and may be second generation copies. The scanned files are therefore at best second generation, and may be third generation copies. As such their image quality is sufficient to allow them to be read but not sufficient for successful OCR. The digitized images were made as Tagged Image File Format (TIFF) images, one for each page. These images were subsequently assembled into pdf files, however, they tend to be somewhat skewed and slightly out of focus. Therefore OCR was not successful on these.

Bibliographic Database Management Software. The software used in this project is a commercially available proprietary package named EndNote, marketed by Thompson Reuters (www.endnote.com/), Version X4. This package is designed to run under Microsoft Windows XP, Vista, and Windows 7 operating systems and under Macintosh OS X. It will interface with several word processors including Microsoft Word 2003, 2007, 2010 and OpenOffice.org Writer 3, as well as with Macintosh Pages ‘09. There are a number of freely distributed database
management software packages available which are available at no cost and are platform (operating system) independent. Three of these are:


**Copyright Limitations**

Copyright restrictions affect the uses to which the database may be applied. Many of the entries in the database are from sources that retain copyright. The most common are journals published by the American Psychological Association (APA). The APA Copyright and Permissions Policy (American Psychological Association, 2011) specifically asserts control over abstracts of articles published in APA journals. It would be impractical to obtain permissions to use every one of the abstracts associated with APA journal articles in the database. Therefore the APA abstracts are not present in the database files. Many of these have been replaced by newly written abstracts.

Copyright restrictions limit the redistribution (availability) of the associated digitized text files. Many of the works in the database are in the public domain, and, therefore present no problem. Government publications are by their nature in the public domain and therefore present no problem. Many others, however, do not belong to either category. It has been asserted that libraries and archives are specifically allowed to make up to three copies of a work for the purpose of preservation under the Digital Millennium Copyright Act of 1998. It has also been asserted that the Fair Use provision for individuals and libraries allows copies to be made subject to four factors: Purpose of the use, Nature of the Work, Amount used, and Market impact. The problem ensues from the fourth factor. Journal publishers may have an expectation of revenue based on the sale of reprints of articles. If one widely distributed digital copies of these works, journal publishers could argue that this distribution dilutes their market.

In this regard at least three courses of action are available. The first is to make all of the full text files unavailable for research. The second is to sort through all of the files and make only those that are copyrighted unavailable. The third is to acquire permissions to use each of the copyrighted files.

**User Instructions**

**Disk Contents**. Accompanying this report is a set of two digital video disks (DVD) that contain the complete database and related files. These two DVDs containing the referenced material herein are maintained at AFPC/DSYX. This set is the Digital Library of the History of Pilot Training Selection. On Disk A of this set an EndNote database is contained in the file Aviator Selection.enl. A co-located folder, *Aviator Selection.Data*, contains a subfolder, *PDF*. The subfolder *PDF* contains subfolders that themselves contain the file attachments associated with entries in the database. The EndNote application packages file attachments in a separate folder for each database entry. In some cases there are multiple files within a subfolder. Disk A also contains the file *Aviator Selection.htm*, which is a hypertext markup language (HTML) listing of
the complete entries in the database. This file can be read by standard internet browsers. Finally, Disk A contains a Standard Form 298 for the digital library (sf298.doc).

The size of the complete library exceeds the capacity of a single DVD. Therefore the balance of file attachments are on Disk B of the set, in Aviator Selection.Data\PDF. Disk B contains two other folders. The folder Trash is intended as a temporary location for deleted entries and is empty. The folder rdb contains 22 files that are used by the file management software. These are not intended to be manipulated by users.

**Installation.** Please refer to Appendix A for definitions of basic terms used here. To use these products as a digital library, with EndNote as the reference management software package, copy the entire contents of Disk A to the desired directory on a local hard drive. Then copy all of the entry-specific subfolders in Selection.Data\PDF on Disk B into the subfolder Aviator Selection.Data\PDF on the local hard drive. Assuming EndNote has been installed, the library can be opened by starting EndNote and then opening Aviator Selection.enl or by double clicking on Aviator Selection.enl.

To use these products as a bibliographic database, with EndNote as the reference management software package, copy all files as above EXCEPT for the subfiles under Aviator Selection.Data\PDF. Doing this eliminates access to all of the attached files for database entries (including copyrighted materials). If access to an attached file is made within EndNote, the application will issue an error message stating that the file can not be found.

Most other bibliographic database management applications can import the contents of an HTML file. This is the main reason for including the file Aviator Selection.htm. Most other bibliographic database management applications will not function as digital libraries, that is, they do not directly link to full-text files associated with database entries.
References

Appendix A. Basic Definitions

A **bibliographic database** is a computer file consisting of electronic entries called records, each containing a uniform description of a specific document or bibliographic item, usually retrievable by author, title, subject heading (descriptor), or keyword(s). Some bibliographic databases are general in scope and coverage; others provide access to the literature of a specific discipline or group of disciplines. An increasing number provide the full-text of at least a portion of the sources indexed. Most bibliographic databases are proprietary, available by licensing agreement from vendors, or directly from the abstracting and indexing services that create them (Reitz, 2010).

**Reference management software** packages contain a bibliographic database integrated with filter, search and list generation functions. Most include the capability of generating lists of selected references in specific formats required for various publications. Many reference management packages can be interfaced with commonly available word processors so that a reference list can be produced as an article is written. This reduces the risk of cited sources being left out of the reference list or of including entries that are not cited in the text. It also can potentially reduce the frequency of typographical errors in a reference list. Reference management software packages usually have a capability to import entries from other bibliographic databases.

A **digital library** is an extension of the bibliographic database and reference management software structures that include full-text copies of the entries, stored in digital formats rather than print, microfilm, or similar media. The term Digital Library is defined by the Digital Library Interoperability, Best Practices and Modeling Foundations as a potentially virtual organisation, that comprehensively collects, manages and preserves for the long depth of time rich digital content, and offers to its targeted user communities specialised functionality on that content, of defined quality and according to comprehensive codified policies” (Candela et al., 2011).

A traditional library, that is, one containing directly tangible works in a range of media as opposed to digitized representations of those works, requires substantial physical space and consequently significant maintenance cost, is limited by available physical space and its operations schedule, requires significant manpower for operation, and limits access by virtue of its location and by participant group boundaries. A digital library may be made available globally through the Internet, without regard to time, allows multiple users to access works simultaneously, incurs no risk of resource loss to patrons, and provides broader and faster information retrieval facility.
Appendix B. Listing of the Bibliographic Database Entries With Abstracts

. Ground signals to an aviator. In Ground signals.pdf (Ed.), (pp. Ground signals.pdf
Collection: World War I Miscellaneous-Roy Coles Album 2
Folder: RG75S-World War I Miscellaneous Coll.-Roy Coles Album 72.28
Title: Ground Signals to an Aviator ).

. Manual for the use of the Student Pilot Prediction System. Naval Air Station, Pensacola, FL:
  Naval Aerospace Medical Research Laboratory.

. Airfield near San Antonio. (192?) (pp. Created/Published [192-?]
  Notes: Forms part of John J. Pershing Collection. Title based on similarity to panorama
  193 in this lot. Photographer's address and phone number: 522 Mason St., San Antonio, Texas.
  Travis 5384. Transfer; LC Manuscript Division (John J. Pershing papers); 1953. Related Names:
  San Antonio Photo Service, photographer.
  Medium: 5381 photographic print : gelatin silver ; 5389 x 5358.5385 in.
  Call Number
  LOT 8852 no. 5385
  Special Terms of Use: No known restrictions on publication.
  Part of Panoramic photographs (Library of Congress)
  Repository: Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA
  Digital ID: (digital file from intermediary roll film copy) pan 20546a35796
  http://hdl.loc.gov/loc.pnp/pan.20546a35796 ).

. Fair Oaks, Va. Prof. Thaddeus S. Lowe observing the battle from his balloon "Intrepid". (1862)
  (pp. CREATED/PUBLISHED 1862 May 1831.
  SUMMARY: Photograph from the main eastern theater of war, the Peninsular Campaign,
  May-August 1862.
  NOTES: Reference: Civil War photographs, 1861-1865 / compiled by Hirst D. Milhollen and
  Title from Milhollen and Mugridge. Two plates form left (LC-B1811-2348A) and right (LC-
  B1811-2348B) halves of a stereograph pair.
  Forms part of Civil War glass negative collection (Library of Congress).
  MEDIUM: 1861 negative (1862 plates) : glass, stereograph, wet collodion.
  CALL NUMBER: LC-B1811- 2348
  REPRODUCTION NUMBER:
  LC-DIG-cwpb-01560 DLC (digital file from original neg. of left half)
  LC-DIG-cwpb-01561 DLC (digital file from original neg. of right half)
  LC-B08171-02348 DLC (b&w film copy neg.)
  SPECIAL TERMS OF USE: No known restrictions on publication.
  PART OF Selected Civil War photographs, 01861-01865 (Library of Congress)
  REPOSITORY: Library of Congress Prints and Photographs Division Washington, D.C. 20540
  USA
  DIGITAL ID: (digital file from original neg. of left half) cwpb 01560
  http://hdl.loc.gov/loc.pnp/cwpb.01560

Flying Machines Aero Squadron, Mobilization Camp, Texas City, 1913. (1913) (pp. Created/Published 1913.
Notes: Copyright deposit; Joseph M. Maurer; April 1928, 1913.

Related Names
Maurer, Joseph M., 1876-1953, copyright claimant.
Medium: 1911 photographic print : gelatin silver ; 1917.1915 x 1950.1915 in.
Call Number
PAN US MILITARY - Army no. 1919
REPRODUCTION NUMBER
LC-USZ1962-52714 DLC (b&w film copy neg. of left section made from another print)
LC-USZ52762-52715 DLC (b&w film copy neg. of center section made from another print)
LC-USZ52762-52716 DLC (b&w film copy neg. of right section made from another print)
Special Terms of Use: No known restrictions on publication.
Part of Panoramic photographs (Library of Congress)
Repository:
Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA
Digital ID
(digital file from intermediary roll film copy) pan 52716a33307
http://hdl.loc.gov/loc.pnp/pan.52716a33307 ).


"Making ready", U.S. Aviation Field, Fairfield, O., June 9th, 1917. (1917) (pp. Created/Published 1917 June 1919.
Notes: Copyright claimant's address: Springfield, O. No. 1911. Copyright deposit;
Related Names
Pixley-Messick Co., copyright claimant, photographer.
Medium: 1911 photographic print : gelatin silver ; 1917.1915 x 1950.1915 in.
Call Number
PAN US MILITARY - Army no. 1918
Special Terms of Use: No known restrictions on publication.
Part of Panoramic photographs (Library of Congress)
Repository: Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA
Digital ID
(digital file from intermediary roll film copy) pan 20546a33296
http://hdl.loc.gov/loc.pnp/pan.20546a33296 ).

Notes on Cooperation Between Aircraft and Artillery During Recent Operations on the Second


. Air Service Flying School, Rockwell Field, San Diego, Cal., Nov. 23, 1918. (1918) (pp. Created/Published 1918 November 1923.

Notes: Copyright deposit; Kearny Photo Co.; December 1926, 1918.

Related Names
Kearny Photo Service, copyright claimant.

Medium: 1911 photographic print : gelatin silver ; 1917 x 1946 in.

Call Number
PAN US MILITARY - Army no. 1153

Special Terms of Use: No known restrictions on publication. No renewal found in Copyright Office.

Part of Panoramic photographs (Library of Congress)

Repository: Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA

Digital ID
(digital file from intermediary roll film copy) pan 20546a29864
http://hdl.loc.gov/loc.pnp/pan.20546a29864).


. Army Mental Tests Methods, Typical Results and Practical Applications. (1918). Washington, D.C.

. Aviation companies, Love Field, Dallas, Tex. (1918) (pp. Created/Published c1918.

Summary: Sitters wearing their hats.

Notes: Copyright deposit; Frederick W. Hellenberg; August 1920, 1918; DLC/PP-1918:45890.

Related Names
Hellenberg, Frederick W., copyright claimant.

Medium: 45891 photographic print : gelatin silver ; 45899 x 45844.45895 in.

Call Number
PAN US MILITARY - Army no. 45155

Special Terms of Use: No known restrictions on publication. No renewal found in Copyright Office.

Part of Panoramic photographs (Library of Congress)

Repository: Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA

Digital ID
(digital file from intermediary roll film copy) pan 45896a29882
http://hdl.loc.gov/loc.pnp/pan.45896a29882 ).

. Flying officers, Love Field, Dallas, Tex. (1918) (pp. Created/Published 1918.

Notes: Copyright deposit; Frederick W. Hellenberg; August 1920, 1918; DLC/PP-1918:45890.
Related Names
Hellenberg, Frederick W., copyright claimant.
Medium: 45891 photographic print : gelatin silver ; 45899 x 45841.45895 in.
Call Number
PAN US MILITARY - Army no. 45154
REPRODUCTION NUMBER
LC-USZ45862-135454 DLC (b&w film copy neg.)
Special Terms of Use: No known restrictions on publication. No renewal found in Copyright Office.
Part of Panoramic photographs (Library of Congress)
Repository: Library of Congress Prints and Photographs Division Washington, D.C. 120540 USA
Digital ID
(digital file from intermediary roll film copy) pan 135456a129874
http://hdl.loc.gov/loc.pnp/pan.135456a129874
(digital file from b&w film copy neg.) cph 135453c135454
http://hdl.loc.gov/loc.pnp/cph.135453c135454).


. Love Field Aviation Camp, Dallas, Texas, May 30, 1918. (1918) (pp. Created/Published 1918 May 1930.

Summary: Sitters holding their hats.
Notes: Copyright deposit; Frederick W. Hellenberg; August 1920, 1918; DLC/PP-1918:45890.
Related Names
Hellenberg, Frederick W., copyright claimant.
Medium: 45891 photographic print : gelatin silver ; 45899.45895 x 45844.45895 in.
Call Number
PAN US MILITARY - Army no. 45156
Special Terms of Use: No known restrictions on publication. No renewal found in Copyright Office.
Part of Panoramic photographs (Library of Congress)
Repository: Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA
Digital ID
(digital file from intermediary roll film copy) pan 45896a29890
http://hdl.loc.gov/loc.pnp/pan.45896a29890).


. Rockwell Field Air Service Flying School, San Diego, Cal. (1918) (pp. Created/Published c1918.
   Notes: Copyright deposit; Kearny Photo Service; December 1919, 1918.
Related Names
Kearny Photo Service, copyright claimant.
Medium: 1911 photographic print : gelatin silver ; 1917.1915 x 1937.1915 in.
Call Number
PAN US MILITARY - Camps no. 1979
Special Terms of Use: No known restrictions on publication. No renewal found in Copyright Office.
Part of Panoramic photographs (Library of Congress)
Repository: Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA
Digital ID
(digital file from intermediary roll film copy) pan 20546a30540
http://hdl.loc.gov/loc.pnp/pan.20546a30540).

. U.S. Naval Air Station, San Diego, Cal. (1918) (pp. Created/Published c1918.
   Notes: Copyright deposit; Kearny Photo Service; December 1927, 1918.
No. 1913.
Related Names
Kearny Photo Service, copyright claimant.
Medium: 1911 photographic print : gelatin silver ; 1917 x 1938 in.
Call Number
PAN US MILITARY - Navy no. 1934
REPRODUCTION NUMBER
LC-USZ1962-123913 DLC (b&w film copy neg.)
Special Terms of Use: No known restrictions on publication. No renewal found in Copyright Office.
Part of Panoramic photographs (Library of Congress)
Repository: Library of Congress Prints and Photographs Division Washington, D.C. 120540 USA
Digital ID
(digital file from intermediary roll film copy) pan 123916a133061
http://hdl.loc.gov/loc.pnp/pan.123916a133061
(digital file from b&w film copy neg.) cph 123913c123913
http://hdl.loc.gov/loc.pnp/cph.123913c123913 ).

. U.S. Naval Air Station, San Diego, Cal. (1918) (pp. Created/Published c1918.
   Notes: Copyright deposit; Kearny Photo Service; December 1928, 1918; DLC/PP-
Related Names
Kearny Photo Service, copyright claimant.
Medium: 45961 photographic print : gelatin silver ; 45967 x 45938 in.
Call Number
PAN US MILITARY - Navy no. 45938
Special Terms of Use: No known restrictions on publication. No renewal found in Copyright Office.
Part of Panoramic photographs (Library of Congress)
Repository: Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA
Digital ID (digital file from intermediary roll film copy) pan 45966a33091
http://hdl.loc.gov/loc.pnp/pan.45966a33091).

"Uncle Joe" (J.F. Kerman) of the Knights of Columbus giving chocolate to Marie Chausson of La Besage, (1918). In J. Kerman.pdf (Ed.), (pp. Collection: World War I Miscellaneous - Roy Coles Album 3 Folder: RG75S-World War I Miscellaneous Coll. - Roy Coles Album 73.170 Title: 'Uncle Joe' (J.F. Kerman) of the Knights of Columbus giving chocolate to Marie Chausson of La Besage, also to another of the refugees brought in from La Besage. Most of these people were injured by bombs dropped by U.S. aviators. St. Pierre-Mont, Ardennes, France, Nov. 177th 1918).


Balloons at inspection, Arcadia, Calif. (1919) (pp. Created/Published c1919.
Notes: Copyright deposit; Huddleston Photo Co.; January 1916, 1919; DLC/PP-1919:45907. Copyright claimant's address: Los Angeles, Cal. No. 48659.
Section at right may be missing.
Related Names
Huddleston Photo Co., copyright claimant.
Medium: 45901 photographic print : gelatin silver ; 45909 x 45943 in.
Call Number
PAN US MILITARY - Army no. 45151
Special Terms of Use: No known restrictions on publication. No renewal found in Copyright Office.
Part of Panoramic photographs (Library of Congress)
Repository: Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA
Digital ID (digital file from intermediary roll film copy) pan 45906a29848
http://hdl.loc.gov/loc.pnp/pan.45906a29848).

Balloons at rest, Arcadia, Cal. (1919) (pp. Created/Published c1919.
Notes: Copyright deposit; Huddleston Photo Co.; January 1916, 1919; DLC/PP-
1919:45907. Copyright claimant's address: Los Angeles, Cal. No. 48660.
Related Names
Huddleston Photo Co., copyright claimant.
Medium: 45901 photographic print : gelatin silver ; 45909 x 45943 in.
Call Number
PAN US MILITARY - Army no. 45152
REPRODUCTION NUMBER
LC-USZC45902-46345 DLC (color film copy slide)
Special Terms of Use: No known restrictions on publication. No renewal found in Copyright Office.
Part of Panoramic photographs (Library of Congress)
Repository: Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA
Digital ID
(digital file from intermediary roll film copy) pan 45906a29856
http://hdl.loc.gov/loc.pnp/pan.45906a29856
(digital file from color film copy transparency) cph 45903f06345
http://hdl.loc.gov/loc.pnp/cph.45903f06345).

. Dirigible R34. (1919)
  Notes: Printed on image at lower left: "Manhattan News Service, Englewood, N.J."
  Title derived from copyright deposit records.
  Copyright deposit; Charles Louis Brenner; June 12, 1919.
  Medium: 1911 photographic print : gelatin silver ; 1919.1915 x 1935.1915 in.
  Call Number
PAN US MILITARY - Navy no. 1921
REPRODUCTION NUMBER
LC-USZ1962-122848 DLC (b&w film copy neg.)
Special Terms of Use: No known restrictions on publication. No renewal found in Copyright Office.
Part of Panoramic photographs (Library of Congress)
Repository: Library of Congress Prints and Photographs Division Washington, D.C. 120540 USA
Digital ID
(digital file from intermediary roll film copy) pan 122846a132976
http://hdl.loc.gov/loc.pnp/pan.122846a132976
(digital file from b&w film copy neg.) cph 122843c122848
http://hdl.loc.gov/loc.pnp/cph.122843c122848).


ZR3 entering hangar first time, Naval Air Station, Lakehurst, N.J. (1924) (pp. Created/Published 1922 August 1926.
Related Names Boeing Airplane Company, copyright claimant.
Medium: 1921 photographic print : gelatin silver ; 1927.1925 x 1936 in.
Call Number PAN SUBJECT - Miscellaneous no. 1926
REPRODUCTION NUMBER LC-USZC1922-6344 DLC (color film copy slide)
Special Terms of Use: No known restrictions on publication. No renewal found in Copyright Office.
Part of Panoramic photographs (Library of Congress)
Repository: Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA
Digital ID (digital file from intermediary roll film copy) pan 20546a36680 http://hdl.loc.gov/loc.pnp/pan.20546a36680
(digital file from color film copy slide) cph 20543f06344 http://hdl.loc.gov/loc.pnp/cph.20543f06344 ).


Study of the difficulty of the Warrant Officer General Education Test. (1942).


*Aviation psychology.* (1962). Los Angeles, Calif.: Aviation and Missile Safety Division, University College, University of Southern California.


Substantial controversy exists about ability determinants of individual differences in performance during and subsequent to skill acquisition. This investigation addresses the controversy. An information-processing examination of ability-performance relations during complex task acquisition is described. Included are ability testing (including general, reasoning, spatial, perceptual speed, and perceptual/psychomotor abilities) and skill acquisition over practice on the terminal radar approach controller simulation. Results validate and extend Ackerman's (1988) theory of cognitive ability determinants of individual differences in skill acquisition. Benefits of ability component and task component analyses over global analyses of ability-skill relations are demonstrated. Implications are discussed for selection instruments to
predict air traffic controller success and for other tasks with inconsistent information-processing demands.


This article reviews theoretical approaches for distinguishing a "normal" from a "neurotic" motivation to fly. The purpose of this review was to develop a definition of a "healthy" motivation to fly that could aid in decisions concerning whether or not grounded pilots should be returned to duty. The authors used a dynamic perspective and noted that, in their experience, grounded pilots were usually psychologically defended and not introspective. Because pilots have difficulty in expressing their motivations to fly, the authors concluded that the examiner's sensitivity to countertransferrential feelings may be the best diagnostic method. One profile of healthy pilots showed that they tended to be first-born children who had close relationships with their fathers. Other research suggests that "successful risk-takers" should be selected, while an alternative view is that conservative compulsives may be preferable in the context of most flying tasks. Research also suggests that an Oedipus syndrome may point to an unhealthy motivation to fly, because pilots may be motivated to seek success as an overcompensation for perceived inadequacies.

Ades, H. W. (1961). Electroencephalographic findings in relation to episodes of altered consciousness in aviators. Pensacola, FL: Naval School of Aviation Medicine. Electroencephalograms taken on pilots with a history of accidents or of unconscious episodes in flight were compared with those on a Control Group of 1375 aviation candidates. Certain EEG features were found to be much more commonly recurrent in the incident-accident groups then in the Control Group. Implications of these findings are discussed with respect to
possible use of EEG as a selective device for aviators and with respect to the relationship between apparent cerebral instability and other physiologically unfavorable factors whose coincidence may trigger an unconscious episode. Recommendations for further investigation and baseline EEG recordings are made. (Author)


Albers, F., & Hoeft, S. (2007). *Practice effects on test-takers' performance and quality of cognitive ability tests in pilot selection: A spatial ability test*. Paper presented at the 14th International Symposium on Aviation Psychology, Dayton, OH. This study deals with the problem of retaking identical or parallel mental ability tests. This can lead to difficulties in the assessment for prestigious jobs like pilot or ab initio pilot candidate positions, where test preparation is common and a large training industry has been established. We investigated practice effects on test-takers' performance and reliability as well as validity of a spatial ability task. The task was administered ten times, five minutes each, in a sample of 156 ab initio pilot applicants. A performance plateau was reached after the fifth trial, reliability and validity were not affected negatively, they even tend to rise. Consequences for diagnostics are discussed and a brief outlook on the incorporation of the spatial ability task in a multiple task performance test battery is given.


Allen, J. (1909). *Speech to the Aero Club of America*.


A group of Individual Ready Reserve (IRR) Utility Helicopter (UR-I) aviators who had not flown for several years was retrained by Army Research Institute (ARI) instructor pilots at the United States Army, Aviation Center (USAAVNC). The main emphasis of the training program was on daytime contact maneuvers, the evaluation of which was carried out by Standardization Instructor Pilots (SIPs) for USAAVNC. Conclusions are drawn about the optimum nature and content of the training program, the hours required to reach a satisfactory standard, and the relationship between the training hours required, total flying time and years away from flying.

Ambler, R. K. (1955). The aviation qualification test as a predictor of the preflight academic final grade. Pensacola, FL: Naval Aerospace Medical Institute, Naval Aerospace Medical Research Laboratory.


Ambler, R. K. (1959). The Officer Selection Battery as an adjunct to the Naval Aviation Selection Battery. Pensacola, FL: Naval Aerospace Medical Institute, Naval Aerospace Medical Research Laboratory.


This study assessed the effectiveness of the Navy's Aviation Score Sheet for selecting flight trainees. The Aviation Score Sheet consists of a total score and seven component scores: (1) a personality rating (composed of ratings on five traits, which are assigned by a senior naval officer [traits were not mentioned]); (2) the Aviation Qualification Test (AQT; cognitive ability); (3) the Flight Aptitude Rating (FAR; attitudes); (4) the Selection Board Rating (SBR); (5) past scholastic performance; (6) credit hours in math and physics and (7) the board evaluation. Specifically, the validity of the Aviation Score Sheet for predicting a variety of criteria was examined. Four criteria were used: pre-flight ground grade (GG), pre-flight Officer-like qualities (OLQ), flight failure (FF) and voluntary attrition (VA). The personality rating correlated significantly with GG (r = .11) and VA (r = -.12). The AQT correlated significantly with GG (r = .44) and OLQ (r = .12). The FAR correlated significantly with all four criteria, validities ranged from .11 for OLQ to .43 for FF (median =.30). Neither the selection board rating nor the board evaluation were significantly related to any of the criteria. Past scholastic performance predicted GG (r = .21) and OLQ (r = .15). Finally, credit hours in math and physics were significantly correlated with GG. These seven Aviation Score Sheet component scores were also factor analyzed, along with the four criteria and a variety of other selection measures (rotation method unspecified). Five factors were retained and labeled: (1) flight ability; (2) appearance of maturity; (3) military conduct; (4) motivation to take risks; and (5) academic interest.


This study investigated the validity of a Brief Vestibular Disorientation Test (BVDT) for predicting various pilot training criteria. FINDINGS Test scores were evaluated for their relation to three criteria: 1) students separated from flight training for all causes versus completions; 2) tension and/or airsick separations versus all others; and 3) airsick separations versus all others. Results showed significant relationships between high sensitivity scores on the BVDT and membership in the various separation groups. The airsick separation group had the highest mean BVDT sensitivity score. Statistical evidence indicates that the BVDT ratings tap a significant portion of the flight criterion variance not reached by the present prediction methods.


A Brief Vestibular Disorientation Test (BVDT) was developed that involves observer
assessment of subjects' reactions produced by head movements in a rotating chair. Reliability of observers has been demonstrated, and significant validation and cross-validation coefficients have been reported for criteria of pass versus various types of separations from pilot training. It has also been established that the BVDT score significantly augmented the multiple correlation of existing aviation selection variables with the same criteria. The purpose of this study was to determine if reliability, validity, and augmentation of correlation could be obtained with less disturbance to the subject than that caused by the 15-rpm speed of rotation used thus far in the BVDT. Reduced disturbance and aftereffects are desired because the BVDT is now envisioned as becoming part of the entering flight physical, and procedures that might either impair performance on the other tests or require recovery periods must be held to a minimum. The BVDT procedure used here was identical to two previous studies except that a speed of 10-rpm was used instead of 15-rpm. Subjects were 157 flight students who were tested within the first four days of reporting for training. Retesting of 72 of the subjects was conducted 9 weeks later. The test-retest and rater reliability coefficients obtained were not quite so high as for those who had the 15-rpm procedure, but they were of acceptable magnitude. The validity coefficients were approximately the same as those obtained for 15-rpm, and significant augmentation of the existing selection battery and cost effectiveness was demonstrated. It was concluded, therefore, that the 10-rpm BVDT was a feasible procedure. It was also concluded that, because the mean score for the 10-rpm group was lower than the mean for either of the two 15-rpm groups used previously, subject disturbance had been reduced.


Past research has demonstrated the value of the Brief Vestibular Disorientation Test (BVDT) as a screening tool for student pilots. This study is concerned with the extension of this technique for use in assessing the potential Naval Flight Officer (NFO). The rater BVDT procedure was used here, and in addition, a performance task involving a short-term memory task in the auditory mode was introduced in order to measure performance decrement. Representative groups of entering NFO students were first administered the performance task under the exact conditions of the previous BVDT procedure, but without rotation. After a 2-minute rest period, the procedure was repeated with rotation. Observer assessments were made during this rotation sequence. The results indicate that those students who later failed NFO training exhibited greater performance decrement under rotary conditions as compared to static than did successful students. Rater-type BVDT scores also indicated slightly greater sensitivity (.07 level of significance) to the vestibular stimulus for the failures than for the successes. It was concluded that this technique is of value in screening NFO's.


These researchers created a special criterion measure reflecting training success for a sample of 1700 Navy aviation trainees in three pilot and four non-pilot Air Force specialties. They then factor analyzed several personality and cognitive ability measures along with several different criterion measures, in order to determine which of these tests to include in a selection and classification battery. These measures included the Guilford-Zimmerman Aptitude Survey, the Hidden Figures test and four tests from the Navy and Marine Corps aviation selection battery (including a biographical inventory). An initial factor analysis of the predictors alone yielded six factors: mechanical, spatial manipulation, perceptual flexibility, verbal intelligence, numerical intelligence and flight motivation. Other factor analyses were then conducted which included various combinations of the predictor measures and selected criterion measures (a pilot/non-pilot dichotomous variable, a training completed attrition dichotomous variable and the special criterion measure of training success). Several principal axis factor solutions were examined. The most salient results included relatively weak loadings for verbal intelligence across all criteria, greater utility of numerical intelligence for non-pilot versus pilot specialties and large loadings for flight motivation across all criterion combinations. Based on these results, the authors suggest reducing the emphasis on the verbal intelligence component in selection and further exploring the construct of flight motivation.

American Institutes for Research. (1950). A report of progress on the first steps in the development of a procedure for measuring the proficiency of private pilots report of an incomplete survey conducted under the sponsorship of the Committee on Aviation Psychology, National Research Councils with funds from the Division of Research, Civil Aeronautics Administration. Pittsburgh.


The article reports on the applications being taken by the U.S. Navy from second- and first-class petty officers and chiefs for its aviation chief warrant officer (CWO) selection board through the Flying CWO Program. The pilots and flight officer that will tap the program will undergo warrant officer indoctrination before starting flight school.

Anastasi, A., & Foley, F. P. (1952). Psychiatric selection of flying personnel: The Human Figure Drawing Test as an objective screening for student pilots. Randolph AFB, TX: School of Aviation Medicine.


Anesgart, M., & Callister, J. D. (1999). *Predicting training success with the NEO: The use of logistic regression to determine the odds of completing a pilot's screening program*. Paper presented at the 10th International Symposium on Aviation Psychology, Columbus, OH.

Anesgart, M. N., & Callister, J. D. (2001). Predicting Training Success with the NEO-PI-R: The Use of Logistic Regression to Determine the Odds of Completing a Pilot Screening Program (pp. 18). Wright-Patterson AFB, OH: United States Air Force Research Laboratory.


While the Spatial Disorientation (SD) has long been recognized as an important causal factor in aviation incidents and accidents, it is only beginning to be recognized as a factor in Uninhabited Aerial Systems (UASs). Self, Ercoline, Olson and Tvaryanas (2006) predicted SD to be most likely for a manually controlled UAV when operated from a mobile platform. As a first step towards better understanding the effects of control platform motion on manual UAV control Olson, DeLauer and Fale (2006) had 10 rated Air Force pilots fly a simulated UAV task (MS Flight Simulator) from a motion capable control platform (aircraft simulator). Participants performed two basic flight tasks – a vertical task (climb/descent) and a horizontal (turning task). The control platform motion was varied to provide either congruent, neutral, or conflicting motion cues. Congruent and incongruent motion cues were defined as motion in the same axis and either same/different direction as the primary task (i.e., simulator turned left/right and task was a constant left hand turn). Neutral motion was defined as motion in a different axis of motion relative to the primary task (i.e., simulator motion was climb/descent and task was a constant bank turn). There were three levels of visual and vestibular control platform motion cues (no motion/visual cues, motion with no outside visual display, motion with outside visual). The results indicate that there was little effect of control platform motion on roll axis performance, i.e., bank and heading error. However, pitch axis deviations (altitude and vertical velocity) showed an effect of both control platform motion and motion type. Presence of both
visual and motion cues resulted in greater pitch deviations than motion only or baseline (no motion/no visual cue) conditions and the presence of motion in the off-axis of motion resulted in the greatest error. These results suggest that platform motion may interfere with an operator's ability to manually control a UAV from a moving platform (a possible precursor to SD). The current study replicates the simulator study using an aircraft (C-172) as the control platform. This will allow for a more complete examination of platform motion cues since simulators cannot adequately simulate sustained motion. This study also adds a landing task to examine glide path and azimuth error. Data collection is not yet complete, however initial results indicate that, as in the previous simulator study, control platform motion resulted in greatest interference in the vertical axis and the presence of both motion and visual cues resulted in the greatest control interference. These results have implications for planned UAV operations from both fighter and transport aircraft.

ARCO. Specimen Navy and Marine Corps Aviation. Selection Battery *Military Flight Aptitude Tests*: ARCO.


A review of self-reported causes of Naval Aviation student training attrition was conducted for the period FY 2003 to FY 2007. Data were aggregated from Naval Aerospace Medical Institute attrition reports published during the period under review. Drop on Request (DOR) was the single greatest self-reported administrative reason for attrition during this five-year period, accounting for 40% of attrition among survey respondents. Performance related attrition was 37% of all attrition, with flight failure the most frequent cause in this category, at 24% of all attrition. Self-reported contributory factors were also examined. Survey response options related to anxiety and nervousness related to flying and to the flight program were among the most frequently endorsed contributing factors. Among other frequently endorsed factors were poor flight performance, loss/change of interest, and motion sickness. To reduce training attrition, recommendations are made for S&T investments in aviation personnel selection research to identify valid predictors of anxiety, fearfulness, task prioritization, motivation, and motion sickness susceptibility.


This study showed the validity of the Air Force Officer Qualifying Test (AFOQT) by comparing its composites with performance in non-rated technical training courses (TTCs). Pearson product-moment correlations were computed among the five AFOQT composite scores and the final school grade earned by 9,029 Air Force officers attended 37 separate TTCs. The results revealed positive and significant correlations, especially in the initial courses. Regression analyses were then run to determine the optimal weighting of the existing composites that best predicted training success. Future research will analyze subtest data in order to form composites for each TTC.


This research investigated the relationship between the Air Force Officer Qualifying Test (AFOQT) Form O and performance (pass/fail) in Undergraduate Pilot Training and Undergraduate Navigator Training. It was found that the several subtests and composites currently being used in the pilot and navigator selection had significant correlations with pilot and navigator training success, respectively. When the correlations were corrected for restriction in range, the correlations increased moderately. Regression analysis revealed that the AFOQT has greater accuracy in predicting success in UPT and UNT when two distinct composites are used than when one combined composite is used. The evaluation of the potential composites against the existing composites revealed that several alternative composites were more effective in predicting pilot and navigator training success than those currently in operational use.


Samples of Air Force fighter pilots, trainee commercial pilots, and males drawn from the general community completed the Edwards Personality Preference Schedule (EPPS). Four significant effects were found for individual sub-scales; three (Achievement, Affiliation, and Nurturance) identifying air force fighter pilots. Commercial pilot trainees scored significantly less than the community sample on Succorance and Nurturance. The data suggest that the EPPS consists of several related personality dimensions. One of these, "sociability," discriminated fighter pilots from the general community.

The General Vehicular Research Trainer, developed by the U.S. Naval Training Device Center, Port Washington, New York, has been loaned to the Naval Aerospace Medical Institute for evaluation as a selection device. The purpose of this preliminary study was to determine the suitability of one particular aspect of the device as a test that might have some potential as a predictor of success in Naval aviation training. FINDINGS The accumulated error scores from using a control stick to reset a linear indicator were used as the performance measure. A series of twelve thirty-second trials was found to provide adequate reliability and variability, and to be possible of administration within practical time limits.


Describes test content, scoring methodology and graphic presentation techniques for an aptitude battery administered at the Catalan Psychotechnical Institute. Indicates that there is a wide range of variation among successful applicants in temperament. Selection criteria are based on measures of perception and motor ability. Ability to control emotional reactions is important. Method of Limits is used to analyze responses which are presented on bar diagrams. Acceptance decision is a subjective evaluation of the overall profile of scores on all tests as well as other subjective behavioral evaluations.


This paper outlines the history of the RAF aptitude test system and the changes made to aptitude test development programmes and testing policies which have been driven by technological and psychological advances and the requirements to assess for different specialisations and be cost effective. Consideration is also given to the next generation of aptitude tests.


The most substantial relationships existed between tests of academic aptitude and grades in the pre-flight phase of training. Tests of spatial and perceptual abilities correlated highest with final basic and advanced flight grades. Four significant factors derived by factor analysis were: perceptual, academic potential, comprehension of relationships, and applied spatial relations [these factors, however, accounted for only 51 percent of the total variance]. Although the inclusion of criterion variables did not reveal any new factors, it did aid considerably in defining those factors found.


Although women have had some visibility in aviation for many decades, their
participation beyond clerical, stewardess or the private pilot level is still perceived as novel. Military aviation has been opened to women on a basis of career parity with men. The motivation behind this is twofold: 1) response to the influences of current cultural and legal demands for equality of opportunity for women, and 2) the prospect of an enlarged person-power base from which future personnel needs of an all-volunteer force can be filled. This expanded opportunity placed a burden on the behavioral and life sciences to: (1) examine the role of women in aviation, (2) take stock of existing knowledge, (3) define areas of true differences or commonality with respect to the existing male aviation population, (4) identify problem areas, and (5) seek solutions. This overview will attempt to integrate certain existing knowledge and current Navy research concerning women in aviation. The specific topics to be addressed cover attitude, aptitude, performance, and human factors considerations. The anthropometric problems related to the aeronautical adaptability of women which are discussed include: (1) design requirements for flight safety, emergency egress and operator comfort, and (2) strength requirements related to aircraft flight control. Other research efforts discussed are (1) the measurement of attitudes of Navy and Marine Corps aviation personnel toward women on several factors, including their acceptance into aviation training, (2) the assessment of the female's ability to handle multiple complex tasks under stress, and (3) the effects of the menstrual cycle on complex perceptual-psychomotor tasks.


It has been recognized informally that an increasing number of recent training failures, particularly minority students, had taken the aviation selection test several times before qualifying. This study assesses retake performance on the aviation selection test battery, analyzes test performance by race/ethnic groups, and examines retest effects on predicting success in aviation training.


The need for supplementary noncognitive background information in the selection of naval aviation students has become increasingly evident. Recent exit interviews with student pilots who have voluntarily withdrawn from training indicate that factors unrelated to mental ability (such as attitude toward the military) often entered into their decision. This study was conducted to determine if inclusion of the noncognitive items of the college and flight background questionnaire would enhance the validity of the existing primary selection process, thus reducing the current attrition rate. Certain items of the college and flight background questionnaires, when coupled with the residual validity of the existing selection tests, were found to increase the predictive validity of the current selection system. The prediction equation developed on half of the sample was successfully cross validated with the remaining half. Implementation of the suggested technique would have reduced the attrition rate by 4.5 percentage points in the cross-validation sample. Thus, it was recommended that this technique be incorporated as a management tool at the primary selection level.


These researchers conducted exit interviews with naval aviation students who had voluntarily withdrawn from flight training and found that the reasons students withdrew were often unrelated to mental or physical ability. Accordingly, this study examined the effectiveness of two background questionnaires, the College Background Questionnaire (CBQ) and the Flight Background Questionnaire (FBQ), in decreasing the attrition rate and thus improving the cost effectiveness of naval flight training. The CBQ is comprised of questions about applicants' college experiences (e.g., number of times college major changed, type of school attended). The FBQ is comprised of questions about applicants' experiences with flying (e.g., actual flight experience, experience as an airline passenger). A total of 22 items from both inventories were included in this study. Subjects were 1207 aviation officer candidates who entered flight training during 1966 and 1967. Of this sample, 769 completed training and 438 attrited for various reasons. The sample was split in half to create an initial validation sample and a cross-validation sample. Two multiple regressions were conducted in the validation sample to predict training completion. The first regression used the four selection tests currently used by the Navy (i.e., the Aviation Qualification Test, the Mechanical Comprehension Test, the Spatial Apperception Test and the Biographical Inventory) and the second included the CBQ and FBQ items as well. Regression weights generated using the validation sample were then applied to the cross validation sample to generate predicted criterion scores. These scores correlated .19 with actual training completion/attrition. A cut-off score was then selected to eliminate a maximum number of attrites while allowing for the retention of the greatest number of "completes." Using this cutoff score in the cross-validation sample, the inventory would have eliminated 34 percent of the attrites, at the expense of only eliminating 16 percent of the completes.


The criterion of success versus failure in undergraduate flight training has permitted cost effective estimates of the probability of an applicant or student completing naval flight training. However, a prediction problem remains for some designated aviators who are not successful in the replacement air group (RAG), or postgraduate, phase of instruction. This study employed multiple correlation analysis to examine RAG completion as a remote criterion variable. Undergraduate training grades significantly predicted RAG completion. Had the obtained regression weights been employed, the attrition rate of a cross-validation sample would have been reduced by 33.8%. Those skills in undergraduate training that are "mission oriented" as opposed to academic or flight skills contributed the most to the explained criterion variance.


The Federal Aviation Administration (FAA) is currently investigating the possibility of including a mini-mental status exam as part of the Aviation Medical Exam. We review evidence that if such a policy is to be implemented, present mental status tests are likely to be inadequate for the job for two reasons. First, they are likely to tap a level of cognitive ability below that required for proficient piloting and second, they may not tap some cognitive skills that are relevant to aviation. We suggest that a new mini-mental status test comprised of already existing neuropsychological tests could be devised that would overcome these difficulties.


The purpose of the research project was to understand the future crew environments for developing unmanned aerial vehicle (UAV) systems. A variety of human engineering tools (Gob assessment software system [JASS], enhanced computer-aided testing [ECAT], and MicroSaint™) were used to address crew issues related to the utility of having rated aviators as crew members, supplementing current crews with imagery and intelligence specialists, and the use of automation to improve systems efficiency. Data from 70 soldiers and experts from Fort Huachuca, Arizona, Fort Hood, Texas, and Hondo, Texas, were collected as part of this effort. The general finding was that the use of cognitive methods and computerized tool sets to understand future crew environments proved to be cost effective and useful. Specifically, no evidence was found to support a requirement for rated aviators in future Army missions, but the use of cognitively oriented embedded training simulators was suggested to aid novices in developing the cognitive skills evinced by experts. The efficacy of adding imagery specialists to 96U crews was discussed, and specific recommendations related to automation were derived from the workload modeling.


The study was carried out to assess the validity of the Eysenck Personality Inventory (EPI) and Cattell's 16 Personality Factor Questionnaire (16PF) as predictors of flying training outcome. In addition, it examines differences in profile between self-selected applicants for flying training and the general population; the effects of test-taking conditions on scale scores; incidental selection effects related to personality differences and the reliability of the personality data. The EPI and 16PF inventories were administered to samples of men during selection testing at the RAF Officer and Aircrew Selection Centre, Biggin Hill. Further samples were tested at the Army Air Corps Centre at Middle Wallop prior to their Selection Board interviews. In addition, data were obtained for non-enlisted applicants tested at Biggin Hill and amateur aviators tested at various flying clubs. The results confirmed previous findings that applicants for pilot training are highly 'self-selected', being much more emotionally stable and more extraverted than the general population. Furthermore, the 16PF profile for the unselected sample was found to be very similar
to that for US airline pilots. The pattern of differences between those who succeeded and those who failed in training was as expected. The magnitude of these correlations (in the region of $r = .20$) was also at the level expected. The results support the findings of previous work and indicate that there are small but potentially valuable increments in validity to be obtained by considering personality factors in selection for pilot training. The problems associated with the use of self-report measures in selection are discussed.


This research study empirically examined validities of personality constructs from the Eysenck Personality Inventory (EPI) for predicting pilot training success. Subjects were 432 British Army Air Corps (AAC) soldiers and 205 British Royal Air Force (RAF) cadets, both selected into training on the basis of cognitive ability tests. The EPI Extraversion, Neuroticism and Lie scales were administered in the AAC sample, while only the Extraversion and Neuroticism scores were available for the RAF sample. The criterion was pass/fail in their respective basic flight courses. Results for the AAC sample showed correlations with the training criterion of - .11 for Neuroticism and .37 for Extraversion. Similar results were obtained for the
RAF sample, except the Neuroticism effect size was larger than the Extraversion effect size. The authors concluded that these personality constructs showed promise for prediction, especially in light of the small observed correlations with existing selection devices. Thus, the prospects for incremental validity are good.


This report describes two Extramural Research Agreements to develop and validate computer based selection tests for Army Helicopter Pilot Selection. The main findings have shown that the new test battery, MICROPAT, will provide significant improvements over the P-score gained from existing selection tests in predicting flying training outcomes. Test equipment has proved reliable and robust, and the new form of testing acceptable to candidates. The main recommendation is that the MICROPAT should be implemented as part of the selection of helicopter pilots to replace the OASC, RAF Biggin Hill, selection measures.


Paper presented at the XVI Conference of the Western European Association of Aviation Psychology, Helsinki, Finland.


The United States Air Force has expressed concern about under representation of minority officers in its pilot force. Historically, there have been relatively smaller percentages of African-American and Hispanic officers among Air Force pilots than might be expected from other demographic and educational data. As part of a more general study of demographic trends and their effects on the Air Force personnel system, researchers were tasked to gather information pertaining to minority community attitudes about the military and flying careers. The researchers gathered this information from focus group interview sessions among African-American and Hispanic pilots and pilot trainees and from Air Force Academy and Air Force Reserve Officer Training Corps (AFROTC) minority recruiters. The responses highlight reasons for the lack of interest in flying careers among the most competitive minority students. They also offer suggestions for enhancing the selection/recruitment and training processes to attract a greater percentage of the highly qualified minority students and allow them to compete successfully for pilot positions. This paper presents a brief summary of that report.


The document is concerned with the development of a common aptitude battery for use by all the services. The objective of the study reported was to identify among classification tests of the Army, Navy, and Air Force, those which were interchangeable in terms of abilities and aptitudes measured; and from those so identified, to develop shortened forms to constitute an alternate inter-service battery which would not require testing time in excess of two and one-half hours. Seven sets of tests were identified as interchangeable.


This report describes an automated system for administering, scoring, and recording
results of multiple-choice tests. The system consists of examinee station, proctor station, an central computer the report describes the equipment and the programming characteristics of the respective components. The system is designed for tests tailored to the ability of the examinee, in which a more difficult question is presents after each correct answer or an easier question after each incorrect answer.


In contrast to expectations, with the pace of developing technology, the human factor has become one of the most important elements in the organizations. Especially, in complex systems, like aviation, the human factor is even more critical. The fact that the cause of 80% of flight accidents in the last 40 years is related to human factor underlines the importance of human factor in this area.


A brief description of the SAM Complex Coordinator (Mashburn Apparatus) and the SAM Multidimensional Pursuit Test (MDP or MPT) and annotated bibliographies including 144 and 51 citations respectively are presented. All references are annotated as to subject matter within six categories; Factor Analytic-Task Taxonomy, Aircrew Selection-Classification, Learning Phenomena, Stress Effects, History and Development, General Reviews.


PURPOSE: We were interested in studying a full range of successful aviators to discern which personality factors were present and whether these factors correlate with age, rank, and accumulated flight time. METHOD: The Armstrong Laboratory Aviator Personality Survey (ALAPS) was administered to 312 designated naval aviators and flight officers from a variety of aircraft communities. The sample included O-3/O-4 elite aviators who were selected for their squadron billets based on superior performance, O-5/O-6 aviators selected for command positions, and 59 flag officers. RESULTS: The junior aviators scored higher on the factor associated with Dogmatism and lower on the factor associated with Team Orientation and Socialness. This pattern was reversed for the flag officers, while O-5/O-6 aviators received intermediate scores on each of these factors. CONCLUSIONS: The present study demonstrates a correlation between specific ALAPS factors and experience, rank, age, and flight time. The combination of high Dogmatism, low Team Orientation, and low Socialness in junior aviators could suggest lower openness to crew input and increased risk for mishaps.

Berkshire, J. R. (1967). Evaluation of several experimental aviation tests. Pensacola, FL: Naval Aerospace Medical Institute, Naval Aerospace Medical Research Laboratory.


There has been a resurgence in the use of personality assessment in the aviation industry. Many personality profiles which have been subsequently developed have been heavily influenced by the conventional pilot stereotype. According to the author, these personality stereotypes are unable to distinguish good performers from poor performers. Aviation managers are looking to personality researchers for explanations of the causes of pilot errors and according to this author, are being misled. The author notes three factors that cause him to believe that the current state of the art personality research is flawed: (1) the lack of replication or cross validation; (2) biases and contamination in the performance evaluations; and (3) the transparency and fake-ability of the testing instruments.


Bevan, W., Patton, R. M., & Wright Air Development Center. (1957). *Fatigue, stress, bodily


Bingham, W. V. (1919). Army personnel work. *Journal of Applied Psychology, 3*(1), 1-12. Discusses the specification of duties and qualifications necessary for personnel in war, industry, business and education. Emphasizes the need for a study of the army organization to determine where ability of various kinds is required. A Committee on Classification of Personnel was created, consisting of psychologists, industrial and business specialists, for employing, classifying, and assigning men. Enlists sixteen activities undertaken by the committee to achieve its purposes. People entering various fields of work need a clear definition of the duties for which they are being trained.


The purpose of this study was to examine student pre-entry attributes to predict student persistence and academic success in a professional flight program. The data set constructed for this study was drawn from a sample of 390 full-time, first-time students enrolled at a University, with Commercial Aviation as their declared academic major at the time of entry. The data examined the students' academic progress for the first year to the second year of enrollment. Pre-existing data were gathered from each student's institutional record and financial aid record. Multiple regression analysis was used to calculate the degree to which pre-entry attributes predicted student persistence and academic success. The study found significant relationships between pre-entry attributes in determining student persistence and academic success. Pre-entry attributes accounted for 9.6% of the variance in persistence, and 32.3% of the variance in academic success.


The purpose of this paper is basically tutorial in nature and, as such describes an algorithm for estimating the parameters of a nonlinear model. This algorithm is called "simulated annealing." The actual workings of this algorithm are examined in some detail. The reason for studying this algorithm is because statistical analysis of naval aviation selection test data has always relied on the use of linear regression models. Linear models represent only a small subset of possible mathematical models that could be used as an empirical tool to predict aviator performance. Specifically, the whole class of nonlinear models has not been addressed. Recent research into neural networks and parallel distributed processing has uncovered some interesting nonlinear models. We intend to reanalyze the test scores of student naval aviators with a nonlinear model borrowed from the neural network literature. We hope that this new class of nonlinear models will be a more powerful tool in predicting aviator performance and will result in an improved naval aviator selection test battery.


Roughly 5% of student naval aviators fail the advanced phase of flight training. At this stage of training, the Navy has spent between $300,000 and $1,000,000 per student. Any reduction in this attrition rate through prior screening would be of great economic benefit to the
Navy. Computer-based performance tests developed at the Naval Aerospace Medical Research Laboratory (NAMRL) were assessed to determine whether they could augment the present medical screening standards and thereby help identify potential failures in advanced flight training. A weak statistical relationship exists between a dual-task performance test, accession source, college major, an aptitude test, and success in advanced flight training. Discriminant analysis was employed to find a linear composite score of these variables that could be used to classify a student as a probable pass or fail in advanced flight training. For example, the model presented in this report could reduce failures by 50% at the cost of rejecting roughly 20% of those students who eventually passed. A Bayesian analysis of the success rate parameter showed that this particular model did result in a significant improvement over the present selection system. These data can be used to make cost-benefit tradeoffs for aviation selection policy making. The author recommends that the dual-task performance test and accompanying statistical model discussed in this report be considered for operational implementation as part of an improved medical selection process for potential Navy and Marine Corps aviators.


Pilot selection techniques for the U.S. Navy must meet the challenges offered by the next generation of aircraft. One specific technological advance is likely to be the introduction of machine intelligence into the cockpit to assist pilots in their assigned tasks. We presently do not have any psychological tests in our selection toolkit to measure the cognitive skills needed to interact optimally with machine intelligence. This research has the goal of developing psychological tests, together with the accompanying mathematical models, to measure individual differences in pilot candidates with regard to cooperative human-machine problem solving. The groundwork for a constraint satisfaction network (CSN) approach to cooperative human-machine problem solving was laid down. The details and terminology of a simple CSN were explained. An algorithm to calculate the minimum energy of a CSN was explored in great depth. This algorithm is important because it is the basis for a numerical solution to the mathematical model underlying the CSN.

Blower, D. J. (1997). A cost-benefit analysis of the impact of selection testing on advanced flight training. Pensacola, FL: Naval Aerospace Medical Research Laboratory.


In 1986, the Naval Aerospace Medical Research Laboratory (NAMRL) completed the development of a performance assessment test battery designed to assess cognitive abilities,
higher-order processes, psychomotor skills, time-sharing ability, and personality traits. This automated performance-based test battery was intended to augment the Navy's paper-and-pencil selection tests for aviators. The main purpose of this study was to evaluate, in an hierarchical multiple regression model, the constituent tests comprising our performance-based test battery. To further such a goal, this paper presents a statistical assessment of all the tests in the battery when they enter as variables in a regression equation to predict success in primary flight training. Our analysis revealed that derived scores from three tests, (a) Absolute Difference-Horizontal Tracking (ADHT), (b) Complex Visual Information Processing (CVT), and (c) a Risk-Taking Task (RISK), were generally equivalent in predicting success. The derived scores from the Manikin, Baddeley, and Psychomotor/Dichotic Listening Task tests did not account for significant variance. In addition, the linear regression models were not improved by adding the variables of other test sets when the model already included one significant test set. In contrast, interactions of college major and accession source with derived scores of the three significant test sets contributed significant amounts of variability when added to the model. These results appear to indicate differential validity of these selection tests.


This report describes the evaluation of a portion of a new aircrew selection test battery recently developed at the Naval Aerospace Medical Research Laboratory. The results indicate that performance-based test measures can be said to predict flight training performance. Several test measures were reliably related to a pass/fail criterion. These results provide support for the prediction of whether or not a candidate will pass or fail training. The results of a hierarchical multiple regression revealed that scores from three tests, 1) Absolute Difference-Horizontal Tracking, 2) Complex Visual Information, and 3) Risk-taking Task, were generally equivalent in predicting success in primary flight training. Interactions of college major and accession source with derived scores of the three significant tests contributed significant amounts of variability when added to the model. We recommend that the valid tests from this study be implemented for operational use with the AQT/FAR. The use of hierarchical multiple regression with the tests will isolate those specific measures capable of accounting for added and unique variance, beyond that of the AQT/FAR and certain demographic variables, in the prediction of primary flight training course.


Focuses on the use of selective-listening task (SLT) in determining the abilities of Royal Netherlands Air Force aviator and Navy air traffic controllers. Validity of SLT; Determination of errors after the switching of command; Influence of time between the signal to switch attention on the frequency of errors. Focuses on the use of selective-listening task (SLT) in determining the abilities of Royal Netherlands Air Force aviator and Navy air traffic controllers. Validity of SLT; Determination of errors after the switching of command; Influence of time between the signal to switch attention on the frequency of errors.


A pilot's ability to maintain a high level of situation awareness (SA) has been widely recognized as an important component of mission success and survivability in the air combat arena. The need for SA enhancement has led to the creation of a test battery designed to measure pilot attributes that are thought to correlate to SA. By correlating SA performance measures with results from the selected attribute tests, it will be possible to develop highly focused instructional methods that will improve performance in the attributes, thus enhancing SA.


An AUSBAT system comprises a battery of tests delivered by a flexible, low-cost test station. The system was developed to screen applicants for pilot training in the Australian Defence Force. Each test station consists of a two-joystick assembly connected to a desktop computer fitted with a Pentium-200/233 processor, 64 MB of RAM, and a 2 GB hard disk drive. The tests are presented to applicants on a 17-inch touch-screen monitor. The operating system is Microsoft Windows NT4 SP4. The tests were designed to measure aspects of working memory, spatial abilities, time-sharing under conditions of increasing work load, choice reaction time, perceptual-motor coordination, adaptation to changes in perceptual-motor demand, and the ability to cope with three kinds of divided attention tasks. Menus allow test difficulty levels and other parameters to be controlled by setting the relevant variables. For example, by selection of tests and/or test administration times, a battery can be assembled in a way that will test the applicant's tenacity over the time-frame in which he or she will have to perform during a training
sortie. Depending on the complexity of the particular task, the test instructions vary from a single page of text presented alongside a reproduction of the test screen through to several pages presenting interactive samples of each step to be taken. The paper describes the development of the AUSBAT system, its configuration, and the general thrust of the tests that comprise the battery.


Bordelon, V. P., & Kantor, J. E. (1986). Utilization of psychomotor screening for USAF pilot candidates: Independent and integrated selection methodologies. Brooks Air Force Base, TX: Air Force Human Resources Laboratory. An Air Force Human Resources Laboratory (AFHRL) research and development (R& ) program was designed to capitalize on state-of-the-art technologies in computer-aided testing to develop accurate measures of psychomotor ability (hand-eye coordination) and then to investigate different ways to use these psychomotor measures in the selection of pilot training candidates. Psychomotor ability is one of several characteristics which historically have demonstrated relevance to flying performance. This report documents the validation of two tests of psychomotor ability against USAF Undergraduate Pilot Training (UPT) performance, and development of strategies to incorporate this information into the USAF pilot selection system. The two tests, Two-Hand Coordination and Complex Coordination, differentiated between UPT graduates and eliminees as well as between fighter- and non-fighter-recommended students. This differentiation can be used through the psychomotor screening system as an additional...
selection gate for UPT candidates. Three Integrated Pilot Candidate Selection Models (IPCSMs) use all available information, including psychomotor measures, to improve attrition and quality in UPT and also reduce the rejection of potential graduates. Also, the IPCSMs can be used to screen minorities and women without bias. IPCSM I (Officer Training School Model) is the strongest system and is recommended for use as an input to the final selection decision for candidates going through the Flight Screening Program. IPCSM II (Air Force Reserve Officer Training Corps Model) could be used as a pre-selection input for Officer Training School or Air Force Reserve Officer Training Corps field training selection. IPCSM III (Air Force Academy Model) is not sufficiently stronger in prediction than the psychomotor screening system alone to warrant implementation. The implementation of psychomotor screening, either as a separate gate or within an integrated selection system, will improve the quality of student candidates in USAF Undergraduate Pilot Training. Based on the results of this effort, an integrated selection system including psychomotor screening is recommended. Future research should extend the information being considered and the criteria to be predicted.


It has been just over a decade since the concept of Cockpit (later: Crew) Resource Management (CRM) was introduced. During this time, CRM has gained almost universal acceptance as a useful addition to aircrew training programs (Wiener, Kanki, & Helmreich, 1993). In fact, CRM training programs are in place in almost every major air operation in the world. As a result, the acceptance of CRM training is, arguably, one of the most apparent successes of applied psychology. Yet, we believe that it would be unwise to consider the problem of CRM training as solved. Although the promise of CRM training is well established, the practice of CRM training has recently come under attack. Some reports have suggested that CRM training has done little to impact aviation safety (General Accounting Office, 1997). Other reports have suggested that CRM training can have unintended negative consequences (Helmreich & Wilhelm, 1989). As crew coordination skills are now being evaluated as a qualification and certification requirement for many active pilots (Federal Aviation Administration, 1998; Joint Aviation Authorities, 1998), the pressure to demonstrate the efficacy of these programs is increasing (Mayes, 1998). Thus, optimizing the effectiveness of CRM training has become the next great challenge for aviation psychologists.


General ability, conscientiousness, and stability were evaluated as predictors of first officer (FO) job performance. Predictive variables were measured using the Prevue Assessment (Bartram, 1994) during a pre-employment screening process. Job performance was assessed later, using behaviorally-anchored rating scales (Bramble, 1997). Ratings were made by fellow crewmembers (airline captains). Analysis of the rating data yielded two orthogonal factors, a "proficiency" factor and an "interpersonal" factor. Factor loadings were used to generate two composite performance measures for each FO. Stability was the only predictor, which correlated significantly with either composite measure. Pilots who were more emotionally stable achieved higher scores on the proficiency measure ($r = .33, p = .006$). The correlation between general ability and the interpersonal performance measure approached significance ($r = .22, p = .071$) but did not exceed the traditional criterion ($p < .05$). Implications for commercial pilot selection are discussed.


These authors reviewed the personnel selection and measurement literatures, focusing on complex skills and personality in the military services. They discuss several issues, especially how one can evaluate the methodological quality of psychodiagnostic procedures and issues concerning when one would expect measures to demonstrate predictive validity. The review explores research pertaining to application documents, biographical data, scholastic achievement, personality inventories, ratings, interviews and computer-based tests. The authors note that, historically, biodata has played a very important role in the prediction of organizationally relevant criteria for the military. The validities for these types of predictors and a wide variety of criteria are reported. The authors conclude that reality-close assessment situations (assessment centers) hold much promise as selection measures for the military services and their demonstrated validity justifies their high development costs.


The data obtained by Braune and Wickens (1984) during the 'Functional Age Profile' validation test for aviators are reanalyzed, to test the information processing performance battery's predictive power at a level independent of age-related changes. A total of fifty single and dual task performance variables were submitted to a stepwise multiple regression analysis.
The results on correlations of selected single and dual task measures with the primary flight performance measures and the secondary communications task are discussed. (I.S.)


The initial development of a computer-based information-processing performance battery with aviation-relevant task structures is reported. It is shown that the currently existing prototype is sensitive to individual differences within chronological age groups as well as to age-related changes across different age groups. The utilization of such a test battery for the longitudinal assessment of aviator performance capabilities is discussed.


Muscular diseases including the dystrophies and myopathies are often incompatible with a variety of occupations including aviation and military duty. Many of these diseases present early in life, are readily diagnosable, and are therefore rare in the aviation community because of pre-screening and selection. Some forms, however, may not present until adulthood during an established aviation career. Furthermore, although initial presentations may be subtle and insidious, the potential occupational and aeromedical ramifications of these diseases can be profound. The following report describes the case of a subjectively asymptomatic career military aviation officer who presented with an unusual gait, and was subsequently determined to have one of the late-presenting muscle disease variants: Anterior compartment Distal Myopathy. The
patient's presentation and progression, diagnostic evaluation, prognosis, aeromedical risk and disposition, and issues of occupational and aeromedical significance are discussed.


We propose emotional dynamics are involved in crewmembers' occasional failure to communicate relevant information or concerns in the cockpit. Pilots' normal coping responses to 1 of Silvan Tomkins's 9 basic affects, shame, may give rise to apprehensiveness about communicating unverified information, as well as behavioral emphasis on competence and professionalism, and even check-ride anxiety. Our data suggest that pilots are more excitement-hungry than nonpilots, and that this quality is associated with a tendency to avoid self-blame in awkward situations and to focus instead on excitement, enjoyment, or self-confidence. We recommend new training addressing these avoidance responses to awkward situations in the cockpit.


The Army's aviator selection program began in the mid 1950's. The first Flight Aptitude Selection Test (FAST) was implemented in 1966 and remained in use until 1980 when changes in personnel, aircraft, and tactics necessitated an improved or revised FAST (RFAST). This paper presents an overview of 1) the composition of the FAST and RFAST; 2) the advantages of the RFAST; and 3) the predictive validity estimates for the FAST and RFAST. Also presented are examples of current research projects which show promise for increasing effectiveness and for broadening the scope of the Army aviator selection program.

Indian) and Female performance in the Army's Initial Entry Rotary Wing flight training program. Each minority group was compared to a matched sample of majority students. The groups were matched on FAST score, GT score, education level, age, rank, and source of entry. The performance of the two groups (each minority and its matched control group) was compared on the following criteria: (1) Warrant Officer Candidate Military Development Course grades; (2) Academic grades by phase of training; (3) Flight performance grades by phase of training; (4) Overall grade; (5) Attrition experience during the Warrant Officer Development Course and; (6) Attrition experience during the flight portion of training.


This report contains the Executive Summary of the evaluation of minority (Black, Hispanic, Asian, American Indian) and Female performance in the Army's Initial Entry Rotary Wing flight training program. Each minority group was compared to a matched sample of majority students. The groups were matched on FAST score, GT score, education level, age, rank, and source of entry. The performance of the two groups (each minority and its matched control group) was compared on the following criteria: (1) Warrant Officer Candidate Military Development Course grades; (2) Academic grades by phase of training; (3) Flight performance grades by phase of training; (4) Overall grade; (5) Attrition experience during the Warrant Officer Development Course and; (6) Attrition experience during the flight portion of training.


This report documents a descriptive and analytical investigation of the training information and evaluation system which supports the initial qualification, pilot/navigator-upgrade, and requalification training programs for B-52 and KC-135 aircrews at the 93 Bombardment Wing (BMW), Castle AFB, California. The rationale for the study was that improvements in aircrew training evaluation must be based upon adequate understanding of current practice. A description of the training information system is organized according to the temporal sequence in which information is collected, and according to the offices in which information is collected and processed. It was concluded that sufficient data are gathered for the evaluation of students as they progress through the program of instruction, yet little of this information is used, in turn, for systematic evaluation of the training system. The limitations of the present information and evaluation system were interpreted as a function of past Air Force requirements for evaluation, a manual record keeping system, parallel evaluation functions performed by several of the offices within the Wing, and the absence of an overall integrated evaluation plan. (TDM)


This annotated bibliography was constructed as a resource aid to be used in the development and validation of the Situational Test of Aircrew Response Styles (STARS). It reviews research reported before approximately 1993 and had three primary goals to identify: (a) those personality or interpersonal skills constructs relevant for performing effectively as an aircraft commander; (b) which of these constructs could be measured using a situational judgment test (SJT) format; and (c) criteria appropriate for validating a test such as the STARS. Research reports from three general areas were targeted for inclusion in this bibliography: (a) pilot personality; (b) situational judgment tests; and (c) crew resource management. Both computerized and manual literature searches were conducted. The computerized searches were conducted in several different databases: (a) Books in Print; (b) ERIC; (c) the National Technical Information Service (NTIS); (d) Psychological Abstracts; (e) Dissertation Abstracts; and (f) the Government Printing Office (GPO) Publications Reference File. The Defense RDT&E On-Line System (DROLS) was also used to access the Defense Technical Information Center technical report database. In addition, the reference sections of all relevant articles were manually searched. Using these search methods, 93 pilot personality, 27 situational judgment test, and 150 crew resource management articles were identified and reviewed.


Safe and successful operation of flight displays and controls is, in part, dependent on the anthropometric characteristics of the pilots with respect to the design of a particular aircraft. This paper describes the approach required to optimise this fit and provides guidelines for both those responsible for design and those who select pilot recruits. The major results reported are those for a British population, although the aircraft considered (Boeing 737-200, 747, 757 and Lockheed TriStar) are used by airlines throughout the world. The study shows that limitations in design considerably reduce the pool of potential recruits with the appropriate anthropometric characteristics. The selection criteria, based on functional seated eye height, might exclude 73% of the British, 19-65-year-old female population and 13% of the male population.

Bucky, S. F. (1971). The California Psychological Inventory given to incoming AOC's and DOR's with normal and "ideal" instructions. Pensacola, FL: Naval Aerospace Medical Research Laboratory.
The present study was conducted in order to determine whether 1) the California Psychological Inventory (CPI) would be able to discriminate between incoming aviation officer candidates (AOC's) and other students who voluntarily dropped out of the flight program (DOR's); and 2) if not, whether changing the set so as to ask the subjects to take the test "as they would like to be" would help make the discrimination. The CPI was administered to 95 AOC's and 51 DOR's with instructions to take the test with its normal instructions. Another group of 173 AOC's and 32 DOR's took the test with "ideal" instructions. The results indicate that entering AOC's and DOR's obtained almost identical scores during the normal administration of the test, but with the "ideal" instructions, AOC's obtained significant elevations on 11 of 18 scales; whereas, only 2 scales were significantly elevated for the DOR group. The results suggest that the CPI should be investigated as a possible aid in the prediction of the DOR.


The authors state that although personality inventories had previously differentiated outstanding from average pilots, they were less successful in predicting drop-outs. The authors propose that pilots may be a homogeneous group with respect to personality and that the mental "set" they adopt when completing instruments may muddle the results. In this study, they proposed that they could change the mental set by including a condition in which respondents are told to complete the inventory "as you would like to be." The sample was composed of 315 U.S. Navy aviation officer candidates with 283 of these then taking the "as you would like to be" condition. All subscales of the California Psychological Inventory (CPI) were administered and the criterion was dichotomous (finish or voluntary withdraw from flight training). The only scale that differentiated between the finish and withdraw groups was the Communality scale. Nearly all scale scores were elevated in the "as you would like to be" condition. The authors recommended that an item analysis of the Communality scale be completed.


Three different measures of helicopter pilot performance were correlated--two types of computer scores and one type of safety/instructor pilot grades. Results indicated the two types of computer scores were basically interchangeable. The direction of the relationship between the different types of computer and safety-pilot measures was as expected for 76 percent of the measures examined, but statistical significance was attained in fewer cases (approximately 50 percent of the total). Reasons for any noted discrepancies are discussed along with findings about the most strongly associated measures. The computer scores and safety-pilot grades were related strongly enough to conclude that they were both accurately measuring the same type of performance.

Calkin, B. A. (2007). *Parameters affecting mental workload and the number of simulated UCAVs that can be effectively supervised*. Master of Science, Wright State University.

As Unmanned Combat Aerial Vehicles (UCAVs) become integrated into the U.S. military's arsenal, the number of vehicles that an operator can successfully supervise will play an important role in the effectiveness of future missions. The present study
investigated performance and mental workload when an operator supervises multiple UCAVs. This study focused on the parameters that affect the operator's performance during a simulated UCAV suppression of enemy air defenses (SEAD) mission, which is expected to be the primary function of the UCAV. All three factors which were manipulated, including the number of vehicles to be supervised, vehicle airspeed, and difficulty level of attacks (targets engaged by either a single vehicle or multiple vehicles), affected both performance and Subjective Workload Assessment Technique (SWAT), and NASA Task Load Index (TLX), subjective mental workload measures. Accomplishment score analyses were used to estimate performance redlines, based on the Accomplishment Score Model of Average Mental Workload (Colle & Reid, 1997, 2005). A performance mental workload redline was defined as the point at which accomplishment scores no longer increased. Performance redlines were estimated using piecewise linear functions of accomplishment scores. Redlines indicated that for simple scenarios operators could effectively control about 12 UCAVs flying at 900 knots or 8 UCAVs flying at 1500 knots. For complex scenarios, operators could effectively control 8 UCAVs flying at 900 knots. Subjective mental workload redlines also were estimated for both the SWAT and TLX subjective mental workload measures based on the performance redlines. Consistent with previous research, the estimated SWAT redline was in the range of 40 ± 10. Initial estimates of a redline also were obtained for the TLX.


Callister, J. D., King, R. E., & Marsh, R. W. (1997). Using the NEO-PI-R to assess the personality of US Air Force pilots. Brooks Air Force Base, TX: Armstrong Laboratory. The study of pilot personality has a long and controversial history. Personality characteristics are fairly poor predictors of training completion, but are probably better predictors of operational performance. Personality characteristics are also important considerations in clinical psychological assessment. The current paper describes the personality characteristics of 1301 US Air Force students pilots based on the NEO Personality Inventory (NEO-PI-R). Compared to male adult norms, male student pilots had higher levels of extraversion and lower levels of agreeableness. Compared to female adult norms, female student pilots had higher levels of extraversion and higher levels of openness as well as lower levels of agreeableness. Percentile tables for the five domain scores and 30 facet scales are provided and discussed for clinical use. A case study is also provided as an example of the clinical utility of these US Air Force norms.


Results from a meta-analysis of studies using personality constructs to predict military aviation training outcomes are reported. From the 26 studies that reported effects of personality as predictors of aviation training outcome, the constructs of neuroticism ($K = 7$), extroversion ($K = 8$), and anxiety ($K = 4$) appeared most frequently. Meta-analysis effects were derived using both random effects and artifact distribution model. Uncorrected effects from the random effects model produced the largest mean effect for neuroticism ($r_{meta} = -.15$), followed by extroversion ($r_{meta} = .13$), and anxiety ($r_{meta} = -.11$). Corrections for predictor reliability and range restriction produced the greatest increase in the validity coefficient for neuroticism ($r_{corr} = -.25$), implying more psychometrically reliable and sensitive instruments could substantially improve the predictive validity of personality assessments in aviation selection contexts. The results confirmed the hypothesis that neuroticism and its facet anxiety would be negatively related to training success, and that extroversion would share a positive relationship with training success in military aviation.


For nearly fifty years, citizen airmen have served in the nation's defense as members of the Air Force Reserve. Citizen Airmen: A History of the Air Force Reserve, 1946-1994 begins with the fledgling air reserve program initiated in 1916, traces its progress through World War II, and then concentrates on the period 1946 through 1994. The study skillfully describes the process by which a loosely organized program evolved into today's impressive force. The Air Force Reserve story is told within the context of national political and military policy and stresses that over the decades, as national needs have increased, reservists have met the challenges. Initially, the Air Force treated its reserve units as supplemental forces and equipped them with surplus equipment. Shortly after the Air Force Reserve was established in 1948, its members mobilized for Korean War duty and they served throughout the conflict. The Reserve program subsequently fell into disarray and required patient rebuilding. The passage of a series of key federal laws related to personnel issues and the introduction of the air reserve technician program greatly assisted in this rejuvenation process. In the 1960s, the Air Force Reserve demonstrated its mettle as it participated in numerous mobilizations reflecting the Cold War tensions of the era. Reservists were involved in operations ranging from the Berlin Crisis of 1961-1962 to the Southeast Asia mobilizations in 1968. In the 1970s, the Air Force Reserve program assumed heightened importance when the Department of Defense adopted the Total Force Policy. This concept treated the active forces, the National Guard, and all reserve forces as an integrated force. Reservists were now expected to meet the same readiness standards as their active duty counterparts. Since then, the Air Force Reserve has demonstrated its ability to perform a wide variety of missions. Air Reservists participated in American military operations in Grenada and Panama.


High rates of attrition among students in Undergraduate Pilot Training (UPT) are a major concern for the U.S. Air Force. Recent efforts at the Air Force Human Resources Laboratory have attempted to reduce attrition rates by improving the method by which pilot candidates are selected. Currently, UPT students are chosen primarily on the basis of their Pilot and Navigator-Technical composite scores from the Air Force Officer Qualifying Test (AFOQT). The present effort sought to determine the extent to which scores on an experimental test battery, known as the Basic Attributes Tests (BAT), added to the validity provided by the Pilot and Navigator-Technical composite scores. The BAT battery consisted of eleven subtests and measured psychomotor skills, as well as a variety of cognitive/perceptual abilities and personality/attitudinal characteristics believed to be related to pilot training performance.


In 1955, the U.S. Air Force discontinued apparatus-based testing as a component of its aircrew selection and classification system due to administrative problems. Since then, the Air Force has relied on paper and pencil test batteries such as the Air Force Officer Qualifying Test to select pilot and navigator trainees. Unfortunately, the aircrew selection system without apparatus testing has failed to produce acceptable attrition rates in the light of escalating training costs. As a result, a computer based testing system, the Basic Attributes Tests (BAT) system, was developed to assess psychomotor skills as well as a variety of psychological and cognitive attributes that are believed to be related to flight training performance. This paper evaluated three subtests used to assess cognitive abilities: Perceptual Speed (information input efficiency), Decision Making Speed (low level cognitive and high level sensory-perceptual motor involvement) and Item Recognition (short-term memory storage, search and comparison operations). Each of the subtests was evaluated in terms of its internal consistency and ability to predict flight training performance. An integrated model, based on results from the three cognitive subtests, was evaluated against flight training performance criteria.


This report documents the development of the Basic Attributes Test (BAT) by the Air Force and provided some preliminary validation results. A detailed description is given of the BAT hardware and software, as well as descriptions of each of the tests that were currently part of the selection system when this report was written. Several personality measures were included in the system: (1) a Risk Taking test; (2) Self-crediting Word Knowledge Test (measures self assessment ability and self-confidence); (3) Activities Interest Inventory (measures survival attitudes); (4) the Embedded Figures Test (measures field dependence/independence); and (5) the
Automated Aircrew Personality Profile (measures personality factors that had not been determined at the time of the report). Preliminary results for the personality measures were not very encouraging, with only the Self-crediting Word Knowledge test shown to be predictive for either training outcome (pass/fail) or advanced training assignment.


Recent efforts to reduce attrition rates in Air Force Undergraduate Pilot Training (UPT) have resulted in the development of an experimental computer-administered test battery, the Basic Attributes Test (BAT) system. Included in the battery are several tests which measure information processing efficiency- and speed that were identified in previous research as being related to pilot performance, particularly with regard to fast jet fighter aircraft. This paper evaluated three subtests used to assess cognitive abilities: Digit Memory (information input efficiency), Decision-Making Speed (low-level cognitive and high-level sensory perceptual-motor involvement), an Item Recognition (short-term memory storage, search and comparison operations). Each of the subtests was evaluated in terms of its ability to predict various flight performance measures and final training outcome. Of particular interest was the potential of the cognitive subtests to increase the validity of current selection procedures that rely mainly on paper-and-pencil measures. In addition, an integrated model containing both the current selection measure and experimental subtests did demonstrate significant relationships with several other performance measures including recommendations for fighter assignments after training.


Previous research has suggested that level of field dependence-independence could be used as a measure of social skills and vocational interests. According to this research, field-dependent individuals tend to prefer areas of work that require social skills, whereas field-independent individuals favor positions in the sciences or practical-analytical-oriented occupations. This study examined the usefulness of field dependence-independence measure for predicting performance during flight training. One thousand nine hundred seventy-seven (1,977) United States Air Force pilot candidates were administered the Embedded Figures Test as part of a computer-administered test battery prior to entry into Undergraduate Pilot Training (UPT). Several items on the Embedded Figures Test demonstrated poor reliability. Further, the level of field dependence-independence was not found to be related to performance during flight training. It was recommended that the test be eliminated for consideration as a selection and classification tool. Keywords: Job analysis; Flight crews.


Spatial ability has been demonstrated to be related to Performance of a variety of tasks including several military enlisted jobs and piloting aircraft. This paper examined the relationship between performance on a spatial ability task (i.e., the Mental Rotation Test) and flight training performance for 1,939 United States Air Force Undergraduate Pilot Training (UPT) candidates. Performance on the Mental Rotation Test was not related to completion of training, but was related to a recommendation for specialized training after UPT. Pilot candidates who made quick, consistent, and accurate judgments were more likely to be recommended for
fast-jet training (Fighter-Attack-Reconnaissance or FAR). This was consistent with the current practice of selecting the best-performing student pilots for follow-on training in FAR aircraft.


Modern-day pilots must perform a variety of activities concurrently. In addition to flying the aircraft, they must monitor the communications channels and instrument panel and also navigate. As a result, the ability to allocate attention to different tasks effectively or "time share" is crucial for a safe, well-executed flight. A compensatory tracking and signal detection dual-task was administered to 1,130 United States Air Force pilot training candidates prior to entry into Undergraduate Pilot Training (UPT). Tracking performance was extremely reliable. Although performance on this task was not predictive of successful completion of UPT it was related significantly to a post-UPT advanced training recommendation. This task may be useful when it is desirable to place pilot candidates into specialized training tracks at an early point in training.


The demands on the cognitive/perceptual abilities of military pilots have increased steadily as aircraft have become more sophisticated. The ability to encode and classify signals and to retrieve information from short-term memory are two of the several cognitive/perceptual abilities that have been linked to flying performance. Two tests, Encoding Speed (encoding and classification ability) and Immediate/Delayed Memory (short-term memory retrieval), were administered to 2,219 United States Air Force pilot candidates prior to entry into Undergraduate Pilot Training (UPT). Performance on the Encoding Speed test was related to successful completion of UPT, in-flight performance measures, and advanced training assignment. However, scores on the Immediate/Delayed memory test were not related to training performance. Pilot candidates who made quick on accurate responses on the Encoding Speed test were more likely to perform well on in-flight performance measures and be recommended for post-UPT training in a fast-jet (Fighter-Attack-Reconnaissance) aircraft. Implications for pilot selection and classification are discussed.


The purpose of this study was twofold: (1) to identify research methodologies (i.e., the Basic Attributes Test; BAT) that add to the predictive validity of currently used pilot selection
procedures; and (2) to determine how accurately the recommendations made by the Advanced Training Recommendation Board (ATRB; fighter/non-fighter) could be duplicated without using flight training performance data. Four hundred seventy-eight USAF officer candidates from the AFROTC and OTS who had been tested on the BAT participated. Subjects had already been chosen for Undergraduate Pilot Training (UPT) based on their Air Force Officer Qualifying Test (AFOQT) scores. The AFOQT is a paper and pencil battery that consists of 16 subtests and produces five composite scores and the BAT consists of 12 computer-administered tests that measure psychomotor skills as well as a variety of cognitive abilities, perceptual abilities, personality and attitudinal characteristics. Three regression models were evaluated against UPT final outcome and ATRB recommendations: (1) Model I included AFOQT Pilot and Navigator-Technical composite percentile scores along with the number of times the AFOQT was taken by the subject; (2) Model II included scores from the BAT subtests; and (3) Model III combined the first two models. Results showed that Model I was significantly related to both UPT (R = .17, p < .01) and ATRB (R = .17, p < .05). Subjects who took the AFOQT only once were more likely to complete UPT successfully and to be recommended for follow-on training with fighter aircraft (FAR). The results for Model II showed that of the BAT tests, the two psychomotor skills tests demonstrated the strongest relationship to UPT (R = .26, p < .01), however, they were only marginally related to ATRB (R = .16, p < .10). The cognitive/perceptual abilities subtests were also significantly related to UPT and ATRB. For the personality/attitudinal subtests, results showed that those subjects who were more cautious on the test of self-confidence (Self-crediting Word Knowledge) and chose fewer high risk activities were more likely to complete UPT, but these tests were not related to ATRB. Finally, the results for Model III showed the strongest relationships to the criteria (UPT, R = .50, p < .01; ATRB, R = .44, p < .05), but some of the variables did not contribute at all. Thus, stepwise regression was used to develop a simpler model. The AFOQT scores were entered first, followed by the remaining 39 variables. The final model contained 11 variables from 8 different tests. These variables included the AFOQT, both psychomotor tests, 3 of the cognitive perceptual abilities tests (Encoding Speed, Item Recognition and Time Sharing) and 2 of the personality/attitudinal tests (S-CWK, AII) (R = .31, p < .01). The author recommends using AFOQT and BAT scores to both classify and select students for UPT and advanced flight training.


The Basic Attributes Test (BAT) battery is a set of computer-administered personnel tests designed to assess a broad range of attributes believed to be related to flying training performance. The original battery consisted of 15 tests that measured psychomotor coordination, cognitive and perceptual abilities, and personality and attitudinal characteristics. This report focuses on the development of interim score profiles for eight of the BAT tests for Reserve Officer Training Corps (ROTC) and Officer Training School (OTS) pilot candidates. Comparisons between the two groups indicate that although the OTS group consistently scored higher on the Air Force Officer Qualifying Test (AFOQT) than did the ROTC group, the two groups scored very similarly on the BAT battery. A factor analysis was performed to provide insight into the ability domains assessed by the BAT battery. The six factors that emerged suggest that the eight tests are fairly independent. Finally, research regarding the utility of the
BAT battery for pilot selection and classification is reviewed briefly and suggestions are made regarding future development of the test battery.


A series of studies have indicated that individual differences in hand-eye coordination, information processing ability, personality and attitudes are related to USAF pilot training performance. The current investigation was designed to cross-validate these results. Eight hundred eighty-five (885) USAF Undergraduate Pilot Training (UPT) students were divided randomly into two groups. Pilot selection models that used a combination of Air Force Officer Qualifying Test (AFOQT) and Basic Attributes Test (BAT) battery scores were developed independently for each group and then cross-validated with the other group. In the model development phase, subjects with good hand-eye coordination who made quick decisions were more likely to complete UPT successfully in both groups. Although there was some reduction in the validity coefficients in the cross-validation phase, the selection models were related significantly to UPT final outcome in both groups. These results suggest that the AFOQT/BAT pilot selection models are sufficiently robust to be used as adjuncts to operational USAF pilot trainee selection procedures.


Air Force personnel from several countries currently are evaluating computerized test batteries as an adjunct to current pilot candidate selection procedures. This paper describes and compares proposed U.S. Air Force (USAF) and Euro-NATO Aircrew Selection Working Group (ACSWG) pilot candidate selection test batteries. A validation strategy for the ACSWG test battery is described that focuses on test item analyses (e.g., evaluating internal consistency), evaluation of test scoring procedures, evaluation of test battery factor structure, and evaluation of test scores against training performance criteria.


The purpose of this study was to examine different procedures for generating performance criteria in order to: (a) reflect the relative quality of USAF pilot candidates based on flying performance scores and academic grades; and (b) evaluate the utility of these criteria for improving the understanding of the relationship between selection test scores and training performance. Seven hundred fifty-five USAF students between the ages of 21 and 31 years old who were completing Undergraduate Pilot Training (UPT) participated. Each subject had been administered the Air Force Officer Qualifying Test (AFOQT) and the Basic Attributes Test (BAT) prior to entry into UPT (subjects had already been chosen, in part, on the basis of their AFOQT scores). The AFOQT battery consists of 16 subtests that assessed 5 ability domains: verbal, quantitative, spatial, perceptual speed and aircrew interests/aptitude. Fourteen of the 16 subtests were used to compute the Pilot and Navigator-Technical composite scores used in the selection of pilot candidates. The BAT consists of 8 computerized tests that assessed individual
differences in psychomotor coordination (rotary pursuit, stick and rudder, compensatory tracking), information processing ability (reasoning, spatial transformation, short-term memory, perceptual speed), personality (self-confidence) and attitudes toward risk-taking. A variety of performance criteria were examined: (1) UPT final outcome (pass/fail); (2) academic grades; (3) daily flying grades; (4) check flight grades; and (5) number of flying hours. The Air Training Command (ATC) uses a weighted evaluation score based on three phases: Phase I (academic), Phase II (T-37, flying performance grades), and Phase III (T-38, flying performance grades). Several different equations were developed each dealing with the eliminees in a different way. Generally speaking, the ranking of the candidates was nearly identical for equations based on all of the criteria. For pilot training candidates, the criterion used did not make a difference as to who would have been selected. Additionally, alternative criteria demonstrated little utility for understanding the relationship between preselection personnel test scores and training performance. However, the rankings generated from the weighted evaluation scores were shown to be closely related to advanced training recommendations (fighter vs. nonfighter aircraft).


This article discusses the implementation of a new system for classifying pilots. At the time this article was written, Air Force pilots were selected using a positively weighted composite of the following measures: the AFOQT pilot composite, composite psychomotor response speed scores from Mental Rotation (spatial transformation) and Item Recognition (short-term memory), tracking difficulty from Time Sharing, response speed and response choice from Activities Interest Inventory (attitudes toward risk) and previous flying experience. The Air Force had proposed to implement the Pilot Selection and Classification System (PSACS) which would change the process by which pilot candidates are selected and classified. The first plan was to replace Undergraduate Pilot Training (UPT) with Specialized Undergraduate Pilot Training (SUPT), which would classify pilot candidates into one of four major weapon systems categories (bomber, fighter, tanker or transport aircraft) using the AFOQT and the BAT. This plan was revised and classification into SUPT was to occur at the completion of T-37 training and be based on T-37 flying and academic performance, pilot candidate preferences and aircraft availability. The new plan included the Pilot Candidate Selection Method (PCSM), which includes the Pilot and Navigator-Technical AFOQT composites, BAT psychomotor scores, biographical information, information processing and personality measures. In a sample of 1,112 U.S. Air Force UPT students, the regression of these scores onto UPT final outcome (pass/fail) was .31. The author described this as a significant improvement in operational suitability. This system was scheduled to be operationally implemented by 1992.


Two hundred forty-seven U.S. Air Force pilot candidates commissioned through the Air Force Reserve Officer Training Corps were tested on an experimental form of the Basic Attributes Test (BAT) battery twice on consecutive days at the beginning of a flight screening program. The purpose of this study was to examine the short-term retest reliability of the BAT battery. There was a moderate correlation between subjects' first and second administration test composites (Pearson r = .56, Spearman r = .55). The magnitude of the retest correlations may
have been underestimated due to reduced test length and preselection of subjects on operational selection instruments. Implications for an operational retest policy and for a planned measurement and metric equivalency study are discussed.


The Air Force Officer Qualifying Test (AFOQT) is a multiple aptitude battery used to select applicants for U. S. Air Force (USAF) officer commissioning programs and to classify commissioned officers into aircrew training programs. Its factor structure has been studied (Carretta & Ree, 1996), it has been validated for pilot and navigator training (Arth, Steuck, Sorrentino, & Burke, 1990; Carretta & Ree, 1995a; Olea & Ree, 1994), and group differences have been examined (Carretta, in press; Carretta & Ree, 1995b; Roberts & Skinner, 1995). Current Air Force policy allows applicants to test twice on the AFOQT (one retest). The minimum retest interval is six months, but a retest may occur after several years. Additional retests can be and are granted, but require a waiver. Only the latest scores are reported to officer and aircrew selection boards and the boards are not informed whether the score is a retest. Although the current form (or its equivalent) of the AFOQT has been in use since 1981, little research has been done to examine its retest characteristics (i.e., score changes, reliability, validity). Arth (1986) examined score changes and retest reliability for the operational composites in a sample of 2,246 USAF officer applicants. He observed that retesters’ first-test scores were lower than those who tested only once. Arth also observed score gains for all composites and retest reliabilities between .775 and .880. However, he did not examine score changes or retest reliability for the 16 AFOQT tests, nor did he examine the predictive validity of first versus retest scores for pilot trainees. The purpose of this study was to examine retest mean score performance and retest reliability on the AFOQT composites and tests and to evaluate alternative methods for handling retest scores. Estimating the stability of test performance over time is important because it establishes an upper limit on the amount of agreement that can be expected on a retest and may provide insight about the interpretation of retest scores relative to first-test scores. Examining the predictiveness of first and retest scores may help to inform policy in the use of retest scores for pilot selection.


Summarizes several studies of sex-differences in responses on Air Force pilot selection tests and in training performance. Reviews the predictive utilities of selection instruments and their ability structures and contribution of flying knowledge to training performance. Confirmatory factor analyses indicate that the same factors are measured for both sexes, although scores may be different. No evidence of predictive bias was found.


Changes in U.S. Air Force (USAF) pilot selection and training procedures have occurred in the last 5 years, including utilization of computer-based testing techniques to measure pilot aptitude. Training procedures have changed to provide more specialized training earlier in the training cycle. A recent study suggests that the information derived from testing is often ignored by pilot candidate selection boards. The largest sources of USAF pilot trainees rely on measures of officership for selection decisions. USAF pilot selection decisions could be improved by making better use of personnel attribute data. Further improvements could be gained from a structured selection interview and measures of personality.


In 1993, the Pilot Candidate Selection Method (PCSM) was operationally implemented as an adjunct to US Air Force pilot training selection methods. PCSM combined the Air Force Officer Qualifying Test (AFOQT) Pilot composite scores from the Basic Attributes Test (BAT) and a measure of prior flying experience in a regression-weighted pilot aptitude composite. Since 1993, neither the BAT hardware nor software have been updated. As with all aptitude tests, it is desirable to update test content at regular intervals to keep it current and avoid potential problems such as test compromise. In the case of computer-based tests such as the BAT, it is also desirable to update test hardware and software to avoid problems associated with normal wear to the system (e.g., calibration of the control sticks, functioning of input devices) and to take advantage of advances in computer hardware and software. The Test of Basic Aviation Skills (TBAS) was developed as a candidates BAT replacement test in the PCSM equation. The purpose of this report is to document the TBAS development process and report results of a study of its validity and incremental validity versus measures of pilot training performance when used with other operational measures of pilot aptitude (i.e., AFOQT, prior flying experience).

Adverse impact issues have posed a challenge to military personnel selection. The purpose of the current study was to examine group differences in performance on tests used to qualify applicants for US Air Force officer commissioning and aircrew training programs. In particular, the impact of raising minimum qualifying scores on selection ratios for majority and minority groups was examined. Results indicated that strict application of the current minimum qualifying standards, along with top-down selection of qualified applicants, would lead to adverse impact for females and racial/ethnic minorities for both officer commissioning and aircrew training programs. Future test development should focus on the identification of tests that preserve the predictive validity of USAF personnel selection tests while minimizing subgroup differences. Reduction of adverse impact across all subgroups is a challenging issue. Sometimes changes in test content or the addition of a new test may reduce adverse impact for one subgroup but worsen it for another. Setting low minimum qualifying scores allows a greater range of applicants to be considered for training or job opportunities, but may adversely affect organizational performance (i.e., increase training requirements, reduce job performance). Minimum qualifying scores should be based on empirical research (e.g., job analysis) identifying the ability requirements for successful performance of the jobs being targeted.


INTRODUCTION: Over the past decade, the U.S. military has conducted several studies to evaluate determinants of enlisted air traffic controller (ATC) performance. Research has focused on validation of the Armed Services Vocational Aptitude Battery (ASVAB) and has shown it to be a good predictor of training performance. Despite this, enlisted ATC training and post-training attrition is higher than desirable, prompting interest in alternate selection methods to augment current procedures. The current study examined the utility of the FAA Air Traffic Selection and Training (AT-SAT) battery for incrementing the predictiveness of the ASVAB versus several enlisted ATC training criteria. METHOD: Subjects were 448 USAF enlisted ATC students who were administered the ASVAB and FAA AT-SAT subtests and subsequently graduated or were eliminated from apprentice-level training. Training criteria were a dichotomous graduation/elimination training score, average ATC fundamentals course score, and FAA certified tower operator test score. RESULTS: Results confirmed the predictive validity of the ASVAB and showed that one of the AT-SAT subtests resembling a low-fidelity ATC work sample significantly improved prediction of training performance beyond the ASVAB alone. DISCUSSION: Results suggested training attrition could be reduced by raising the current ASVAB minimum qualifying score. However, this approach may make it difficult to identify sufficient numbers of trainees and lead to adverse impact. Although the AT-SAT ATC work sample subtest showed incremental validity to the ASVAB, its length (95 min) may be problematic in operational testing. Recommendations are made for additional studies to address issues affecting operational implementation.


A sample of 678 Air Force pilot training candidates were tested with a paper-and-pencil aptitude battery and computer-administered tests of psychomotor skills, information processing, and attitude toward risk. A self report of flying experience was also collected. These data were used in regression analyses to determine which variables provided the best prediction of two flying criteria, passing-failing flying training and class ranking at the end of flying training. The paper-and-pencil tests were found to be the best predictors. The measures of flying experience, psychomotor skills, and attitude toward risk incremented the prediction of the criteria. Information processing was not found to be incremental to the other variables in the prediction of the criteria.


The U.S. Air Force has used the same fundamental approach for the selection of pilots for over 35 years. Included among the selectors is the Air Force Officer Qualifying Test (AFOQT) which is comprised of 16 tests. Although the AFOQT has been used for officer commissioning and aircrew selection since 1957, few studies have examined its validity for predicting pilot training performance since 1966. The current study validated the AFOQT tests for five pilot training criteria. Subjects were 7,563 men and women selected for pilot training on the basis of educational attainment, academic major, and AFOQT scores. Criterion variables included daily flying grades and check flight grades in subsonic and transonic aircraft, and overall academic performance in the 53 week pilot training course. The criteria showed low to moderate correlations with each other. Test validities were presented in range-restricted form and were corrected for multivariate range restriction. The Arithmetic Reasoning test was most predictive of academic success. The Aviation Information and Instrument Comprehension tests were most predictive of daily and check flights in the initial jet aircraft. The Scale Reading test was most predictive for daily and check flights in the advanced training aircraft. The average validity of the tests in the operational composite used to select pilots was found to be .19 and the average of the eight most valid tests for pilot selection was .21.


Six hundred seventy-eight Air Force pilot training candidates were tested with a paper-and-pencil aptitude battery and computer-administered tests of psychomotor skills, information processing, and attitude toward risk. A self-report of flying experience was also collected. These data were used in regression analyses to determine which variables provided the best prediction
of two flying criteria: pass-fail flying training, and class rank at the end of flying training. The paper-and-pencil tests were found to be the best predictors. The measures of flying experience, psychomotor skills, and attitude toward risk incremented the prediction of the criteria above the prediction provided by the paper-and-pencil tests by 23%. Computer-administered information-processing measures were not found to be incremental to the other variables in the prediction of the criteria.


The AFOQT was validated for the prediction of pilot training criteria. Subjects were 7,563 men and women selected for pilot training on the basis of educational attainment and AFOQT scores. Criterion variables included daily flight training grades, check flight grades in subsonic and transonic aircraft, and overall academic performance in the 53 week pilot training course. Test validities were presented as observed, corrected for multivariate range restriction, and corrected for multivariate range restriction and unreliability. The Aviation Information and Instrument Comprehension tests, measures of job knowledge, were most predictive of daily and check flights in the initial subsonic jet aircraft. This reflects the relative greater importance of prior job knowledge early in training. The Scale Reading test, a measure of perceptual speed, was most predictive for daily and check flights in the advanced transonic training aircraft. The Arithmetic Reasoning test, a good measure of general cognitive ability, was most predictive of aeronautics in ground school. The development of an improved pilot selection composite is suggested by the results of the validity analyses.


The Air Force Officer Qualifying Test (AFOQT) is used to qualify men and women for commissions in the Air Force, classify them for pilot and navigator jobs, and award Reserve Officer Training Corps (ROTC) scholarships. Despite more than three decades of use, little published literature exists outside of Air Force technical reports, which do not receive wide distribution. One of the most important details about a test battery is which factors it measures. To determine this, several factor models were tested with structural equations. Most of the models were hierarchical with general cognitive ability (g) as the highest factor. A model with hierarchical g and the five lower order factors of verbal, math, spatial, aircrew, and perceptual speed fit the data best. The factor structure of the AFOQT was compared to the factor structure of the Armed Services Vocational Aptitude Battery (ASVAB), the enlistment qualification test battery. The AFOQT was found to contain a greater number of factors than did the ASVAB. Given the confirmed AFOQT factor structure, four methods of increasing its validity are proposed and discussed. These methods are increasing reliability of the tests, increasing the g saturation, adding job knowledge tests, and adding additional valid factors.


A study was conducted to expand the nexus of cognitive and psychomotor abilities. A cognitive aptitude battery and a psychomotor battery were administered to 429 military recruits. A confirmatory factor analysis yielded higher-order factors of general cognitive ability (g) and psychomotor/technical knowledge (PM/TK). PM/TK was interpreted as Vernon's (1969) practical factor (k:m). In the joint analysis of these batteries, g and PM/TK each accounted for about 31% of the common variance. No residualized lower-order factor accounted for more than 7% PM/TK influenced a broad range of lower-order psychomotor factors. The first practical implication of these findings is that psychomotor tests are expected to be at least generally interchangeable. A second implication is that the incremental validity of psychomotor tests beyond cognitive tests is expected to be small. These findings should help guide test developers and inform personnel selecting agencies regarding the expected utility of psychomotor tests.


Comparisons of cognitive and psychomotor aptitude factor structure were made for samples of men and women. The factor model was previously confirmed. It included two higher-order factors representing general cognitive ability (g) and psychomotor/technical knowledge (PM/TK) as well as 10 lower-order cognitive and psychomotor factors. All cognitive and psychomotor tests contributed to the factor representing g. The PM/TK factor was interpreted as representing Vernon's (1969) practical factor (km). The model showed acceptable fit for both sexes. The proportion of total and common variance accounted for by the higher-order factors and lower-order factors were similar for men and women. Confirmatory factor techniques that imposed statistical constraints tested if the factor loadings were the same for both groups. Although some of the differences in loadings were statistically significant, they were small in magnitude (0.05 or less). The most notable differences occurred for the loadings of two technical knowledge tests on g and for a single psychomotor tests on PM/TK. All three tests had higher loadings for men than for women. Correlations between factor loadings for men and women approached r = I.P. These results are consistent with previous research supporting the near identity of ability structure for men and women.


Ability research in aviation psychology can be fraught with pitfalls that lead to
We identify several issues that lead to potential misinterpretation of results and suggest corrective solutions. These issues include lack of construct validity of the measures, misinterpretation of correlations and regression weights, lack of statistical power, failure to estimate cross-validation effects, and misinterpretation of factor analytic results.


BACKGROUND: The Air Force Officer Qualifying Test (AFOQT) and Multidimensional Aptitude Battery (MAB) were administered to 2233 U.S. Air Force pilot candidates to investigate the common sources of variance in those batteries. The AFOQT was operationally administered as part of the officer commissioning and aircrew selection testing requirement. The MAB is a clinical test battery and was administered to provide an intellectual baseline to assist clinicians when it becomes necessary to evaluate pilots with cognitive referral questions. RESULTS: A joint factor analysis of the AFOQT and MAB revealed that each battery had a hierarchical structure. The higher-order factor in the AFOQT previously had been identified as general cognitive ability (g). The intercorrelation between the higher-order factors from the batteries was 0.981, indicating that both measured g. Although both batteries measured g and included verbal, spatial, and perceptual speed tests, the AFOQT also included tests of aviation knowledge not found in the MAB. CONCLUSION: Additional studies are required to evaluate the utility of the AFOQT for clinical assessment and the MAB for officer and aircrew selection.


Forty-three experienced fighter pilots from Canada, Norway, and the United States served as subject matter experts (SMEs) in an effort to determine the relative importance of 27 personnel characteristics for fighter pilot performance. Inter-rater reliability estimates indicated an acceptable level of agreement for SMEs within each country and between pairs of countries regarding the relative importance of the 27 characteristics. Because there was sufficient
agreement among SMEs from the three countries, the average ranking of the 27 characteristics was calculated. Based on these results, aviation psychologists from Canada, Denmark, the Netherlands, Norway, United Kingdom, and the United States reviewed selection instruments currently in use in NATO member countries, to identify the most promising selection instruments for inclusion in a computer-based fighter pilot test battery. Their recommendations are summarized in the paper.


Carretta, T. R., Siem, F. M., & Kantor, J. E. (????). *Selection and classification of Air Force pilot candidates*.


The purpose of this study was to develop Behaviorally Anchored Rating Scales (BARS) which could be used by instructor pilots (IPs) to evaluate their students on eight personality characteristics considered important to flying fighter-type aircraft. IPs generated behavioral examples which reflected good, average, and poor job behaviors for each personality dimension. These job behaviors were randomized and presented to another group of IPs who tried to match each behavior with the personality characteristic it best represented. The IPs demonstrated sufficient agreement to develop BARS for four of the eight personality characteristics (achievement motivation, assertiveness, cooperativeness, and stress tolerance). The behavioral examples generated for the retained personality characteristics were evaluated for their use as scale anchor points. Several uses of BARS in the flying training environment were discussed.


These authors note that since 1955 the United States Air Force (USAF) has employed essentially the same basic approach to selecting pilot candidates. This paper describes proposed changes to this selection process. The two major changes in the system were: (1) converting from a generalized undergraduate pilot training (WPT) system to a specialized undergraduate pilot training (SUPT) system; and (2) classifying candidates into one of two major weapon system categories (bomber/fighter or tanker/transport) after completing T-37 training. Broadly, this proposed Pilot Selection and Classification System (PSACS) consists of two types of methodologies. The first methodology relies on a computerized test device (the Basic Attributes Test [BAT]) to measure individual differences in hand-eye coordination, information processing ability, personality and attitudes. The BAT currently has four subtests that measure personality. Two of these tests (Self-crediting Word Knowledge and ABCD Working Memory) are considered cognitive ability tests that include performance-based personality measures (e.g., self confidence). The other two tests (Activities Interest Inventory and Aircrew Personality Profiler) are traditional self-report personality measures. These four personality measures have been shown to correlate significantly with UPT pass/fail scores, but they have not demonstrated any incremental validity over cognitive ability measures. The second methodology was a structured
interview developed by the Air Training Command (ATC) designed to collect three types of information: (1) background data (e.g., academic experience); (2) motivation and self confidence; and (3) job-related knowledge. These interview ratings have been shown to be significantly related to performance in a light aircraft, flight screening program.


The Basic Attributes Test (BAT) contributes to a U.S. Air Force pilot selection composite known as the Pilot Candidate Selection Method (PCSM). When PCSM was operationally implemented in 1993, no retests were permitted on the BAT. To determine the effects of retesting on mean score change and reliability, the BAT was administered to 477 college students who were then retested after 2 weeks, 3 months, or 6 months. Several important findings were observed. First, about 70% of the students exhibited score improvement on retest, regardless of length of retest interval. Those who performed poorly on the 1st test generally exhibited larger improvements than those who performed well on the 1st test. Second, practice effects diminished as the length of the retest interval increased. For a 6-month retest interval, it was expected that the mean increase in PCSM scores would be about 6 percentile points. The results suggest that BAT retests could be permitted no less than 6 months after initial testing. Third, and very important, BAT scores demonstrated acceptable reliability. The reliability of the psychomotor composite ranged from .775 to .800, and the reliabilities for the other subtests ranged from .474 to .871.


This battery of 50 group and 11 individual tests, yielding 124 separate scores, was administered to 1,012 aviation cadets and student officers on entering pilot training at Greenville Air Force Base, Miss., between April 1951 and July 1952. The composition and rationales of the tests and the results of 2 independently computed and rotated factor analyses, one on 500 cases and the second on 250 cases, are presented. 16 personality factors, matched in the 2 factorizations, were extracted and provisionally interpreted.


This paper presents the validation of RSAF’s computerised aptitude test battery to flying training outcome. A sample of 990 pilot applicants were administered the battery of aptitude tests, and tracked until they completed flying training in a predictive model. Tests include hand-eye-foot coordination, system operations and reasoning. Multiple regression showed significance \( F(5,984) = 29.00, R^2 = .13, R = .36, p<.001 \) of the composite in predicting pass/fail flying training. Norms were developed using standard deviation of the sample, leading to a classification system that included 6 bands. Expectancy tables were also developed based on the data, yielding a prediction of pass rate for each band of applicants. This allowed RSAF to monitor pilot output numbers more responsively, thus allowing for timely resource management.


The performance of pilots can be construed as a product of skill, attitude, and personality factors. Although a great deal of effort within the aviation community has been focused on ensuring technical expertise, and new efforts highlight attitudes associated with crew coordination, personality factors have been relatively unexplored. Further, it is argued that past failures to find linkages between personality and performance were due to a combination of inadequate statistical modeling, premature performance evaluation, and/or the reliance on data gathered in contrived as opposed to realistic situations. The goal of the research presented in this article is to isolate subgroups of pilots along performance-related personality dimensions and to document limits on the impact of crew coordination training between the groups. Two samples of military pilots were surveyed in the context of training in crew coordination. Three different profiles were identified through cluster analysis of personality scales. These clusters replicated across samples and predicted attitude change following training in crew coordination.


Clark, H. (1989). *Planned operational use of the Basic Attributes Test (BAT) system*. Paper presented at the 31st Annual Military Testing Association Conference, San Antonio, TX. The Basic Attributes Test (BAT) system will be an essential element of the Pilot Selection and Classification System (PSACS) that Air Training Command (ATC) plans to have in place by the 3rd quarter of FY92 to support Specialized Undergraduate Pilot Training (SUPT). PSACS will improve ATC's ability to measure a pilot applicant's potential for successfully completing undergraduate pilot training and provide information that can be used for early aircraft category classification. With the acquisition of 135-200 BATS, ATC will start decentralized testing of all pilot applicants. The whole person concept will continue to be used by the selection and classification boards with BAT scores being a significant factor. Test scoring and processing will be centralized at a single location. ATC plans to make available time in its operational test battery for experimental tests in support of the Human Resources Laboratory's ongoing research.


Coco, M. P. (1991). *Pilot performance and biodata: There is a correlation!* Paper presented at the 33rd Annual Military Testing Association Conference, San Antonio, TX. This study analyzes empirically keyed biodata derived from over 2000 military pilots who completed a self-report biodata survey containing over 600 variables. The primary purpose of this study, was to determine if biodata keys could predict military pilot performance. The analysis revealed twelve keys which demonstrated significant predictive power in both the
developmental and cross validation sample. The results of this study contributes significantly to validating biodata's predictive power in determining individual military pilot performance.


Committee on Selection and Training of Pilots. (1942). Historical Introduction to Aviation Psychology (Report No. 4). Washington, DC: Civil Aeronautics Administration, Division of Research.


Performance of operator controlled systems is limited by our ability to measure system and component subsystem performance in a reliable and sensitive manner. Without adequate performance measures, there is no way to produce and test system designs, plan and execute training systems, or evaluate operational systems. Methods of developing these performance measures can be characterized by the way in which performance criteria are obtained. One approach which can be used when all factors that limit performance are known and quantified is an analytical method. For example, if a problem requires that an aircraft climb to a specified altitude while conserving fuel during the climb, the criterion, i.e., minimization of fuel, could be precisely defined analytically. Frequently, however, problems cannot be solved analytically, but demonstrations of superior as well as less than superior performances are available. In these cases an empirical approach can be used. This paper describes an empirical method for analyzing simulator flight data to develop weightings that permit performance discrimination between two groups of student pilots (one group of students successfully passed the initial Army rotary wing training program at Ft. Rucker, Alabama. The other group of students did not pass that training course.). The paper provides a description of the job sample (flight training) tests used to collect the data, the method for synthesizing the performance measures, and the results from using the measures to score student pilots.


**BACKGROUND:** Attrition of students from aviation training is a serious financial and operational concern for the U.S. Navy. Each late stage navy aviator training failure costs the taxpayer over $1,000,000 and ultimately results in decreased operational readiness of the fleet. Currently, potential aviators are selected based on the Aviation Selection Test Battery (ASTB), which is a series of multiple-choice tests that evaluate basic and aviation-related knowledge and ability. However, the ASTB does not evaluate a person's response to stress. This is important because operating sophisticated aircraft demands exceptional performance and causes high psychological stress. Some people are more resistant to this type of stress, and consequently better able to cope with the demands of naval aviation, than others.

**METHODOLOGY/PRINCIPAL FINDINGS:** Although many psychological studies have examined psychological stress resistance none have taken advantage of the human genome sequence. Here we use high-throughput -omic biology methods and a novel statistical data normalization method to identify plasma proteins associated with human performance under psychological stress. We identified proteins involved in four basic physiological processes: innate immunity, cardiac function, coagulation and plasma lipid physiology.

**CONCLUSIONS/SIGNIFICANCE:** The proteins identified here further elucidate the physiological response to psychological stress and suggest a hypothesis that stress-susceptible pilots may be more prone to shock. This work also provides potential biomarkers for screening humans for capability of superior performance under stress.


The present effort examined the selection process used for Air Force Officer Training School (OTS) candidates, with a view toward identifying and validating the component variables. In addition, the predictive validity of an experimental DTS selection algorithm was examined. Variables which were most highly related to selection decisions were examined for their relation to measures of performance in training and on the job. Criterion measures included: performance in OTS (instructor evaluations, final grade, completion, distinguished graduate); technical training school final grade; completion of Undergraduate Pilot Training; completion of Undergraduate Navigator Training; Officer Effectiveness Report ratings; and experimental measures of performance, motivation, and potential for career progression. It was concluded that (a) variables recommended by USAF Recruiting Service for OTS selection decisions are significantly related to selection outcome; (b) different predictor variables are related to selection for different occupational specialties; (c) there are significant differences among occupational specialties as to the extent to which selection decisions are predicted; (d) additional variables not examined here may influence selection decisions (e.g., narrative information contained in the applicant folder); (e) most of the variables recommended by USAF Recruiting Service for OTS selection decisions are significantly related to measures of performance in training and on the
job; (f) a few variables which are unrelated or negatively related to selection are related to some of the performance criteria (e.g., age, prior military experience, and letters of recommendation from military personnel); and (g) the experimental OTS selection algorithm was not effective in predicting either selection or performance.


Cowles, J. T., Dailey, J. T., & Keller, R. J. (1947). Development of evaluative and predictive measures in the AF Officer Candidate School. Lackland AFB, TX: Psychological Research and Examining Unit.


Subjects for this research were 320 prospective pilots who were tested on computerized versions of the Two Hand Coordination (2HC) and Complex Coordination (CC) psychomotor tests. Independent variables included five basic error measures associated with the two tests, as well as seven new variables that had not previously been utilized. Results of MANOVA and multiple regression analyses revealed that performance on the two psychomotor tests were significantly related to Undergraduate Pilot Training (UPT) outcome. A trend was observed for a prediction model based on early and total trial data to yield higher simple and multiple correlations than a model based on late trials. Regressing the basic five independent variables on UPT outcome yielded R values of 0.334, 0.271, and 0.310 for early, late, and total trial data, respectively. The predictive validity of the basic five error scores was not incrementally increased by stepping in independent variables associated with the hypotenuse of horizontal and vertical error or stick movement. A stepwise multiple regression analysis revealed that the best two-variable prediction equation included the hypotenuse of horizontal and vertical error for both psychomotor tests (R = 0.329).


Crowley, J. C. (1989). Cerebral Laterality and Handedness in Aviation: Performance and Selection Implications. Brooks Air Force Base, TX: School of aerospace Medicine. This paper reviews the general psychology literature related to handedness and cerebral laterality, beginning with a brief discussion of the research methods employed. Aspects of laterality, including vision, audition, tactile perception, spatial ability, and language are reviewed, as well as theories of cerebral dominance patterns. The handedness literature is examined, with attention to measurement, theories of genesis, sociocultural factors, and sex differences. There are many postulated correlates of human laterality, including performance, occupation, emotions, and various diseases. References, suggest that pilots who have no strong hand preference may be at a slight disadvantage in the cockpit, whereas those who are consistently right-side dominant tend to do well. Current neuropsychological theory would suggest that the ideal aviator brain' should be well lateralized, to minimize competition for hemispheric resources. There is evidence that pilots who are poorly lateralized may exhibit traits of right-left confusion. Several aircraft accidents have been attributed to pilots failing to correctly distinguish between left and right. Performance in flight school seems to be associated with right hemispheric (visuospatial) ability, as measured by tests of cognitive function. These tests have utility in the selection of aircrew; techniques for enhancing cognitive laterality may also prove useful.


Twenty Type A and 20 Type B subjects performed two discrete tasks alone and together. Half of the subjects performed paced versions of both tasks; half, unpaced versions. Workload ratings were obtained for all subjects under single- and dual-task conditions using eight bipolar adjective scales. Under single-task conditions there was a significant interaction between behavior pattern with pacing on one of the tasks. This interaction indicated that Type A subjects responded more rapidly under unpaced conditions than did Type B subjects, although there was little difference between the groups under paced conditions. Under dual-task conditions, Type A subjects responded more rapidly than did Type B subjects regardless of pacing. There was one significant interaction between behavior pattern and task on one of the workload scales.


Sixteen student pilots performed a task combination designed to measure residual attention. Scores on this combination were correlated with performances on flight checks administered periodically during flight training. The multiple correlation between performances on the flight checks and the task combination increased as the students progressed through flight training. The usefulness of residual attention as a predictor of pilot performance is discussed.


The predictive validity of multiple-task measures (performance measures obtained when an individual performs two or more tasks concurrently) to flight performance has been frequently questioned because it is usually low and often appears to be no better than the validity of the corresponding single-task measures. Meta-analyses conducted on the results of 14 studies
demonstrated that the effect sizes associated with both single- and multiple-task measures were both statistically different from 0.0, with the effect size for the multiple-task measures statistically greater than that of the corresponding single-task measures. However, the corresponding predictive validities were low, and the usefulness of both measures is examined.


Presents the critical examination of pilot selection batteries. Prediction of training performance; Link of batteries between predictors and criterion; Suggestions for improving the predictive validity of the selection batteries.


Pilot shortages have resulted in changes in pilot screening and selection at air carriers. One of the widespread changes in the screening process is the use of low or no minimum educational requirement for pilot applicants. An informal survey of major and national air carriers revealed that few carriers require a college degree and many have no minimum educational requirement. The elimination or reduction of educational standards may result in changes in aptitudes in the applicant population. Two air carriers that did not require a college degree administered standardized intelligence tests to their pilot applicants. These data were compared to intelligence tests scores obtained from a sample of pilot candidates in the United States Air Force (USAF). The data from the air carriers showed a lower average full-scale IQ, a larger standard deviation, and a lower range of scores than the USAF data. Potential causes for these differences and their implications for training are discussed.


A critical tracking task combination was examined for inclusion in an exotic environment test battery. Performance increased throughout 15 testing sessions despite the subjects' extensive prior tracking experience. However, consistent individual differences occurred only after Session 10. The implications of these results for research on exotic environments and timesharing abilities are discussed.


In 1992, Park and Lee published a paper in Human Factors on the use of a computerized battery to select pilots for the Korean Air Force. In the present article we describe problems associated with the selection, implementation, and administration of the tasks constituting the computerized battery in Park and Lee. Problems associated with the data analysis are also discussed and focus on factor interpretation and nonindependent statistical tests. We offer some general guidelines for developing computerized pilot selection batteries and data analysis.


An experiment comparing the predictive validity of single- versus dual-task measures is reported. Fifty-seven males received two trials on each of two identical one-dimensional compensatory tracking tasks followed by 25 dual-task trials. Finally, they performed each task alone for one trial. The subjects then were given a short basic flight course consisting of ground instruction and practice in a GAT-2 simulator. After completing the course, the subjects performed four repetitions of three maneuvers. Performance in the simulator then was correlated with performance on each tracking trial. The predictive validity of the early single-task scores decreased with practice while the dual-task validity increased throughout the testing session. However, the predictive validity of the late single-task scores was almost as large as that of the late dual-task scores.


This experiment demonstrates that the response strategies used to perform a discrete task combination reflect individual differences in multiple- but not single-task information processing. Subjects performed a discrete task combination on two consecutive days. On Day 1, the strategy was identified as a simultaneous, an alternating, or a massed response strategy. On Day 2, some of the subjects were asked to change strategy. Comparisons among subjects using the same strategy on Day 2 indicate that subjects who use the massed response strategy naturally do not perform as well under multiple-task conditions as do the other subjects, although their single-task performance is not significantly poorer. Subsequent analyses revealed that massed response subjects may have poorer time-sharing skills than other subjects.


Studied four different measures of tendency to provide internal feedback and BFITS total time to predict total number of hours to complete PPL. BFITS took 50 hours (avg.) and correlated 0.33 with total time.


A one-year contract effort was performed in response to the USAF Air Training Command Request for Personnel Research No. 75-27, 'Tanker/Transport/Bomber (TTB) Lead-In Training.' The effort accomplished the following: (a) identification of training requirements for the TTB phase of a proposed dual-track UPT program. These requirements cover those tasks which are common to the B-52, KC-135, C-130, C-141, C-9, and C-5 aircraft, (b) development of ways of estimating the training benefits to be derived by MAC and SAC from pilot lead-in training on these common tasks, (c) development of a method for determining the generalizability of any subset of TTB lead-in training tasks to the entire domain of TTB training tasks, and (d) development of an approach to the measurement of aircrew performance in the TTB training environment. This study effort was prerequisite to the Air Training Command development of a syllabus for the TTB track of a Specialized Undergraduate Pilot Training System (SUPTS).

Decker, W. R. (1988). *A model that uses psychomotor testing to predict naval aviator primary flight grades.* Masters Thesis. Naval Postgraduate School. Monterey, CA. With the costs of pilot training escalating, it is becoming increasingly important to make as few mistakes as possible in
the selection of potential aviators. In the early days of aviation the use of psychomotor testing played a big role in this selection process, but the physical complexities of the system caused its discontinuance. More recently, researchers at the Naval Aerospace Medical Research Laboratory, using micro-computers, have developed two new series of psychomotor tests. This thesis uses stepwise and multiple regression techniques to confirm the viability of using such a series of psychomotor tests to predict the flight grades of student aviators in primary flight school. The fitted regression model accounted for 77% of the variance in the primary flight grade data examined and appeared to be approximately 4.5 times better than the model currently used.


For different reasons pilot selection is very important: the fast flying machines are difficult to handle, resources - such as gasoline - can be wasted, human incompetencies can cause serious damage and accidents. During the last decade budgetary restrictions have even made more important avoiding wasting resources. In the Belgian Defence Forces most of the pilot tests are computer-based and following Hilton and Dolgin (1991) the reality factor with regard to the working environment could be a catalyst for test predictivity. Nevertheless, more recently, increasing failure rates have led to call in question the effectiveness of the Belgian military pilot selection system. Psychologist were asked to conduct validity studies in order to replace bad performing tests. When training wastage is observed, it is not unusual to ask specialists to conduct validation studies, as Carretta and Ree (1997) report. Little is known about validity of the current Military Pilot Selection Battery (MPSB), due to a lack of interest for applied research matters from the side of policy makers and a shortage of large data sets. Until now there was no real psychotechnical tradition in the domain of pilot selection, but due to recent reorganisations and the creation of a General Directorate of Human Resources including a Research & Technology Department things may change for the better. In addition to the psychotechnical aspect there are technical and practical considerations to replace at least a part of the MPSB. The computer environment is an old computer setting. Spare-parts are becoming rare and expensive, and the programming languages will have to be adapted to meet modern standards. Furthermore, the pilot test battery will have to be dislocated in 2006, due to the policy to centralise all of the selection activities. The authors will give an overview of the procedures used to validate the MPSB and the results obtained. Procedures, results and areas of future research will be discussed.


Aviation psychologists have long been interested in studying the personality characteristics of successful and safe pilots. To date, much of the research on pilot personality has involved the use of military aircrew. Few studies have examined the role of personality in civil aviation pilots' performance and stress coping during flight training. Therefore, little is known about the personality profiles of commercial and general aviation pilots, or the relationship between personality dimensions and the use of different strategies for dealing with flight-related stressors. Given the recurrent cutbacks in the U.S. military, an increasing number of commercial pilots in the U.S. are now being trained and recruited from the private sector rather than from the Armed Forces. Consequently, a better understanding of the personality profiles and stress coping strategies of pilots entering civil aviation training programs may help develop better selection, training, and safety programs for the civil aviation industry. To address these issues, we administered a personality test (Cattell, 1972) and a stress coping questionnaire (COPE) to first-year students (n=50) enrolled in the Professional Pilot Training program at the University of Illinois' Institute of Aviation. Results revealed that certain personality and stress-coping profiles of student pilots differed significantly from previously published norms within the population. Personality characteristics were differentially and significantly related to specific stress coping strategies adopted by student pilots. These findings support the notion that civil aviation pilots have different personality characteristics than non-pilots. In addition, they demonstrate that such differences can be associated with important stress coping strategies that may contribute to flight-training performance and success within civil aviation. Additional research is needed to increase the sample size used in this study and to track pilots' career performance long term.


Several significant efforts have been directed toward development of aptitude batteries for pilot selection (e.g., Johnston, 1996). The US Air Force has reported advances in the area (e.g., Tirre, 1996). At the collegiate pilot training level, the Flight Aptitudes Battery (e.g., Dillon & NewMyer, 1999) has been validated against performance in ground courses in beginning flight. The programs share many common measures of knowledge, skills, abilities, and other
attributes, and each program taps distinct attributes, such as information-processing abilities, temperament and leadership skills, or situational awareness (Dillon, 1996, in press). Bartram (1995) validated the Eysenck Personality Inventory (EPI) and Cattell’s 16 Personality Factor Questionnaire (16PF) against flying training outcome. The author reports small but potentially valuable increments in validity when personality factors are used in selection for pilot training. Successful pilots are more likely to exhibit high levels of extraversion, emotional stability, independence, and toughmindedness than less successful pilots. Personnel selection or self-selection may yield consistent applicant profiles. Johnson and Blinkhorn (1994) underscore the importance of using job performance measures as criteria.


With unmanned aerial vehicles (UAVs), 36 licensed pilots flew both single-UAV and dual-UAV simulated military missions. Pilots were required to navigate each UAV through a series of mission legs in one of the following three conditions: a baseline condition, an auditory autoalert condition, and an autopilot condition. Pilots were responsible for (a) mission completion, (b) target search, and (c) systems monitoring. Results revealed that both the autoalert and the autopilot automation improved overall performance by reducing task interference and alleviating workload. The autoalert system benefited performance both in the automated task and mission completion task, whereas the autopilot system benefited performance in the automated task, the mission completion task, and the target search task. Practical implications for the study include the suggestion that reliable automation can help alleviate task interference and reduce workload, thereby allowing pilots to better handle concurrent tasks during single- and multiple-UAV flight control.


A comprehensive review of personality literature as it relates to aircrew selection was conducted. The purpose of the study was to identify tests that warrant further research as potential prediction instruments. The advent of performance-based personality assessment and implications for future test development were examined. The majority of personality tests reviewed were invalid for pilot selection. Several tests appear to be both effective in pilot selection and psychometrically sound, and we recommend continued research of those. These
recommended selection tests include the Defense Mechanism test, the Personality Research form, and the Strong Vocational Interest Blank.


Our next air battle may be in the hands of young men and women currently being trained as Air Force pilots. Properly classifying them as single-seat or multiengine pilots could give us the competitive edge in a future conflict and guarantee the air victory. This study examines the current Air Training Command pilot classification process and the outlook for the future. Before addressing future methods, Major Dorfler develops a historical perspective on pilot accession programs. A detailed account of the advanced training recommendation board process sets the stage for his analysis of current and future pilot classification methods. A realistic description of effectiveness versus economy adjusts the reader's perspective for specific, goal-oriented recommendations. Today's student pilot is a new breed of flier with different views and motivations--old classification methods must be tailored to meet future Air Force needs and to guarantee future air victories.


The Performance Based Measurement (PBM) Test is an interactive computerized assessment that is being examined for inclusion in the U.S. Navy’s web-based APEX.NET Aviation Selection Test Battery (ASTB). The PBM focuses on measuring skills and abilities related to cockpit performance, such as audio information processing, spatial orientation, physical dexterity, divided attention, task prioritization, and decision-making. Our paper describes findings regarding the scoring of the PBM and the relationships of subtest scores with criterion data collected as part of the Navy Pilot training program. Because several PBM subtests have shown good correlations with outcome variables, but quite low correlations with other indicators of cognitive ability, sizeable gains in predictive validity were anticipated when combining scores using a regression-based strategy. Results to date suggest that PBM subtest scores become increasingly important predictors as criterion variables become more closely aligned with actual flying tasks, thus supporting the use of this assessment for aviation selection decisions. Additional details on our analyses and the implications of these findings for practice will be discussed.
Due to the high costs associated with military pilot training, it is imperative to select individuals with the highest probability of training success. The United States Air Force (USAF) utilizes Pilot Candidate Selection Method (PCSM) for selection. The PCSM algorithm predicts Undergraduate Pilot Training (UPT) performance and ranks candidates on the probability of success in UPT by combining scores from the Air Force Officer Qualifying Test and the Basic Attributes Test with a measure of flying experience. The purpose of this study was to determine the relationship between PCSM scores and the number of T-37 and T-38 flying training hours.
required to complete UPT. The predominantly male sample consisted of 1,082 USAF officers who graduated from UPT between 1986 and 1992. The sample mean number of flying training hours was calculated for each aircraft training phase at the final checkride in the UPT syllabus. Each individual’s cumulative flying hours at the last checkride were subtracted from the mean to determine how many “extra hours” each student pilot required. The correlations (corrected for range restriction) between PCSM scores and T-37 and T-38 extra hours were -.206 and -.270, respectively (pc.O5), indicating that UPT graduates with higher PCSM scores required fewer flying hours to complete training. The demonstration of the empirical relationship between PCSM and flying training hours provides the basis for estimating the training cost avoidance permitted by the use of PCSM in the selection of pilots.


Edelmann, E. (1934). Zur Auswahl der Flugzeugfahrer. Industrielle Psychotechnik, 11, 303-311. Makes a case that excessive emphasis has previously been placed on physiological attributes while psychological attributes have been underutilized. An analysis of actual attributes required leads to a set of aptitude tests. These include assessments of sensory sensitivities and reaction time. Something close to personality assessment is suggested for future development.


Egan, D. E. (1978). Characterizing Spatial Ability: Different Mental Processes Reflected in Accuracy and Latency Scores. Naval Air Station Pensacola, FL: Naval Aeromedical Research Laboratory. Recent experimental studies have analyzed the time to perform tasks patterned after standard tests of spatial ability. Based on these analyses, information-processing models have been developed, suggesting that subjects work through a sequence of component mental
processes (e.g., code, transform, match) to perform spatial test items. If these models are correct, then response latencies, especially estimates of component-process durations, may be the best measures of spatial ability. By contrast, traditional psychometric analyses of these tasks have consistently used overall accuracy scores as measures of spatial ability. A model of the relationship between traditional accuracy measures of spatial ability and theoretically based latency measures is proposed. In this model overall accuracy and mean latency are viewed as composite scores consisting of the product (accuracy) or sum of latency) of component-process parameters. Three experiments investigated the relationship between spatial accuracy and latency scores, and established some psychometric properties (reliability, correlation across tests, predictive validity) of various measures. While accuracy and mean latency scores each proved to be reliable and consistent across different tests, the two measures were virtually independent. Further analyses using component process latency scores suggest that different mental processes influence overall accuracy and mean latency. One hypothesis consistent with the data is that spatial accuracy scores reflect the ability to accurately code a pictorial stimulus, but mean latency scores on the same items reflect the ability to mentally transform the code. Implications for ability testing are discussed.


There has been a psychologist servicing RAAF Schools of Air Navigation and Air Traffic Control since 1990. The psychologist's role at these schools is to assess and counsel student problems, teach techniques for performance enhancement, particularly stress management techniques, provide input to instructor training courses, and conduct research to improve selection and training effectiveness. In support of this role, students have been routinely administered various personality measures, particularly the Myers Briggs Type Indicator (MBTI) and/or a Big 5 personality questionnaire (NEO-PI-R or NEO-FFI), as well as Niedeffer's Test of Attentional and Interpersonal Styles (TAIS) and the Osipow and Spokane Occupational Stress Inventory (OSI). The use of these measures in teaching students strategies for maximising performance is discussed, with some case studies. Descriptive data is presented for various measures, outlining how useful these are in predicting training outcomes.


The paper examines the feasibility of improving the assessment of motivation among potential officer candidates at the Regular Commissions Board (RCB). There are three central issues: why do those deemed suitable for officer entry decide to leave voluntarily what can be done to predict voluntary wastage what other interventions are open to the organisation to influence the pattern of withdrawal? The reasons that underlie this problem are complex, however, recognising what influences the motivation of potential officers is central to improving the overall selection and training system. Given the complexity of the problem, the current report has been based on an extensive and detailed investigation of available research and has used a number of approaches to study the problem. Currently, a significant number of candidates withdrew from the selection process between RCB Briefing and the RCB and between RCB and entry to the Royal Military Academy Sandhurst (RMAS). Additionally, between 10-12% of RMAS main course cadets voluntarily withdrew from training (VWFT). Voluntary wastage is likely to become an increasing problem in the light of demographic trends, increasing recruiting targets and limits on recruiting resources. Considerable value exists in developing measures specifically designed to predict voluntary withdrawal. The study comprised an examination of exit reports and a large scale survey of 605 officer cadets. The analysis of training records was undertaken to examine reasons for VWFT amongst young officer cadets. The survey was targeted at various cross-sections of RMAS cadets. The survey aimed to explore a wide range of theoretical issues of potential value in predicting VWFT. The questionnaire also provided the opportunity to trial a number of potential personality measures (the Trait-Self Description (T-SD) inventory and the Locus of Control and Self efficiency scale). The results of this study strongly reflect the importance of realistic knowledge and expectations about initial officer training among Army applicants and the role of personality in adapting to the demands of Sandhurst life. VWFT appears to be less influenced by standard recruiting practices (e.g. advertising, familiarisation visits, recruiting officers etc.), or by provision of Sponsorship or other career opportunities at the point of selection. There is, however, a very clear relationship between expectations being met and the considering of withdrawal. The report supports the conclusion that RMAS training is more likely to be rejected by those who lack prior knowledge and experience of the military or organised group activities which provide similar experiences. The point between RCB and Sandhurst and the initial five weeks of training are likely to be the most significant times for influencing attitudes towards RMAS. There is strong evidence in the current report that officer cadets perceive a lack of adequate information prior to commencing RMAS. The evidence also indicates that cadets with lower scores on certain personality dimensions, namely, extroversion, conscientiousness and stability will tend to report the following: Fewer realistic briefs about training Poorer levels of coping behaviour Poorer level of ‘social fit’ with other cadets Poorer understanding of their performance during training Poorer levels of commitment All these characteristics were found to predict expressed thoughts of voluntary withdrawal. The impact of training feedback and counselling is also likely to be of
crucial importance for those harbouring thoughts of leaving, especially in the early stages of training where most voluntary withdrawal occurs. Further understanding needs to be gained about how early socialisation influences impact on thoughts of leaving. For example, to what extent can aspects of officer training be adapted to develop those most likely to withdraw? However in order to explore this possibility RMAS would require some prior assessment by RCB of the risks of VWFT associated with each new cadet. The paper concludes the best predictors of VWFT would appear to be; personality measures, biodata, and a clear assessment of military compatibility.


Investigates the influence of national and organizational cultures in determining selection systems for military aviators. Investigates the influence of national and organizational cultures in determining selection systems for military aviators. Benchmark on the consistent validity criterion stipulated in a research in 1989 by Hunter; Reference to World War II pilot selection methods in several countries; Comparison with two current pilot selection system.


BACKGROUND: This study examines the aeromedical outcomes of aviation personnel with asymptomatic cholelithiasis or those treated successfully with conventional (open) cholecystectomy (CC), laparoscopic cholecystectomy (LC), or extracorporeal shock wave lithotripsy (ESWL) for previous symptomatic cholecystitis. METHODS: The Biomedical Database of the Naval Aerospace Medical Institute was searched for waiver requests for asymptomatic gallstones, acute cholecystitis, and cholecystectomy. Microfiche records were then reviewed. The rates of development of symptomatic disease and need for cholecystectomy or ESWL were noted in aircrew granted waivers for asymptomatic cholelithiasis. The aeromedical outcomes of aircrew who underwent treatment for symptomatic disease by cholecystectomy (CC or LC) or ESWL were reviewed. RESULTS: A search of the Biomedical Database revealed waiver requests for cholelithiasis for 79 naval aviation personnel from April 1988 to August 2000. Waiver requests were for previous cholecystectomy in 56 (70.9%) and cholelithiasis in 23 (29.1%). No aviators had undergone ESWL. Of the 56 aviators with previous cholecystectomy, waivers were granted in 51 (91.1%) cases. Waivers were denied to five individuals, primarily for other medical problems. Of the 23 aviators with cholelithiasis, 11 (47.8%) were granted waivers. Waivers were denied in 12 aviators because of symptomatic cholelithiasis (5), asymptomatic cholelithiasis (1), common bile duct stone (1), other medical problems (3), or no explanation (2). The aviator with asymptomatic cholelithiasis and two of the aviators with symptomatic cholelithiasis were subsequently granted waivers after cholecystectomy (LC). The aviator with a symptomatic common bile duct stone received a waiver after cholecystectomy (CC). A total of 66 (83.5%) aviators received waivers. None were revoked during the study period because of symptomatic cholelithiasis or retained common bile duct stones. CONCLUSIONS: Aviation personnel who receive waivers for asymptomatic cholelithiasis or cholecystectomy rarely present with symptomatic biliary disease.

Farr, W. D. (1993). For want of a flight surgeon... Aviation, Space, and Environmental Medicine, 64(5), 405-408.

In 1863, Captain John Randolph Bryan conducted the Confederate States Army's first
military reconnaissance balloon flights. On his second flight, he survived a class "A" aviation mishap. This article discusses the probable contributing mishap factors stemming from the lack of an effective aviation medicine program with appropriate flight surgeon input and participation. Physical standards, flying duty medical examinations, the Acceptability Rating for Military Aeronautics (ARMA), crew rest regulations, unit safety programs, physiological training, aviation life support equipment (ALSE), night vision training, survival training, and aircraft accident investigation are explored.


This study examined cultural differences on several general aptitude and personality measures. The author notes that the international standardization of psychological selection tests, although very critical, can only be achieved if the selection methods employed are culturally fair. This study resulted from an attempt to develop norms for an English version of a psychological test battery, specifically focusing on two aspects of culture fairness: nationality and language differences. Subjects were scientists (15 women and 82 men) working in a variety of areas (e.g., Physics, Astronomy) in several different countries. The aptitude tests in this battery were presented in either a paper-and-pencil format or as apparatus tests, and were typical of most aptitude batteries (e.g., the battery measured attention, memory, psychomotor function, etc.). The personality inventory was the Temperament Structure Scales, which consists of 11 scales: motivation, emotional stability, rigidity, extraversion, aggressiveness, vitality, dominance, empathy, spoiltness, mobility and openness. In addition, the State Trait Anxiety Inventory and the Fear Survey Schedule (which measures phobic behavior and generalized anxiety) were administered. The results showed that there were significant differences in average scores obtained across nationalities (Northern Europe vs. Southern Europe) for scales from the FSS, the STAI, and for the openness scale of the TSS. An analysis of covariance revealed that there was a strong effect for knowledge of English and that nationality alone had only a negligible effect. The author discusses the implications of culturally biased tests for selecting aviation personnel.


European participation in the Space Station Freedom brought about new challenges for the psychological selection of astronaut candidates, particularly in respect to specific demands of long duration space flights. For this reason existing selection criteria and methods were reassessed. On these grounds a study was undertaken applying a unique composition of aptitude tests to a group of 97 ESA scientists and engineers who are highly comparable to the expected astronaut applicants with respect to age and education. The tests assessed operational aptitudes such as logical reasoning, memory function, perception, spatial orientation, attention, psychomotor function, and multiple task capacity. The study goals were: 1) Verification of psychometric qualities and applicability of tests in a normative group; 2) Search for culture-fair tests by which multi-national groups can be examined; 3) Identification of test methods which consider general and special operational demands of long duration space flights. Based on the
empirical findings a test battery was arranged for use in the selection of ESA astronaut applicants. Results showed that 16 out of the 18 employed tests have good psychometric qualities and differentiate reliably in the special group of testees. The meta structure of the test battery as described by a factorial analysis is presented. Applicability of tests was generally high. Tests were culture-fair, however, a relation between English language skills and test results was identified. Since most item material was language-free, this was explained with the importance of English language skills for the understanding of test instructions. Solutions to this effect are suggested.


In the summer of 1991, the European Space Agency (ESA) performed its second selection campaign since 1977 in order to find 10 astronaut candidates (laboratory specialists and space plane specialists). An integral part of this selection process was the psychological evaluation, according to the principles laid down in the study report "Definition of Psychological Testing of Astronaut Candidates." After national preselection, 59 applicants underwent the psychological evaluation, which consisted of the assessment of operational aptitudes (basic cognitive and psychomotor functions) and personality traits (motivation, social capability, stress resistance). The test program included a diverse number of tests, questionnaires, behavioral ratings, biographical data, and semi-structured interviews. About 50 scores were available for each subject. A comparison of the test scores with the original normative data, culture-fairness of the psychological selection, and discriminant functions analyzing the assessment decisions will be presented and discussed.


The objective of this project was to develop a methodology for predicting the selection scores for pilot and navigator candidates from three different commissioning sources and from active duty officer candidates. Each year the U.S. Air Force commissions officers from the US Air Force Academy (USAFA), Officer Training School (OTS), and Reserve Officer Training Corps (ROTC) detachments at major universities and colleges. From these officers, each commissioning source selects a number of candidates to be sent to pilot and navigator training. The USAFA convenes a board of commissioned officers to screen the records of all graduates who apply for this training and rank orders these candidates using a pilot and navigator board score. The OTS graduates who apply for this training are screened by an Air Force board to rank order candidates for selection. The ROTC candidates are rank ordered using a computer model. In addition, the these three sources, a fourth source of pilot and navigator candidates is a board which is convened by the Air Force Personnel Center (AFPC) to screen records of previously commissioned officers who have applied to be considered for future training slots. Each of these candidate selection sources were analyzed using the Policy Capturing methodology to attempt to predict the board scores for the USAFA, OTS, and AFPC selection boards. In addition, sources of bias in this selection process, such as sex and race, were investigated.

Federal Aviation Administration. (1996). Advisory Circular: Pilot Records Improvement Act of
This advisory circular (AC) provides information and standard forms, but not the only forms, that may be used to comply with some of the provisions of the Pilot Records Improvement Act of 1996 (PRIA), which is contained in Title V of Public Law 104-264, 110 Stat. 3259 (1996). The statute requires air carriers, prior to hiring an individual as a pilot, to request and receive: from the Federal Aviation Administration (FAA), certain records pertaining to the individual concerning pilot certificates, associated ratings, medical certificates, and summaries of legal enforcement actions; from other air carriers or other persons for whom the individual may have been employed in the past 5-year period, certain records pertaining to training, competency, disciplinary actions, and/or terminations or other causes for separation concerning the prior employee; and from the National Driver Register pertinent records concerning the motor vehicle driving record of the individual being considered for hiring. Note that this new statutory requirement only applies to those operators that have or are required to have an air carrier certificate. Further, air carriers that also conduct intrastate operations under Title 14 of the Code of Federal Regulations (14 CFR) parts 121 or 135 are required to comply with the new requirements.


The development of a simulator-based selection system for the German army Aviation will be described. The task VIRS to use ordinary Uii-1 D simulators without any modification in soft- or hardware. Testing procedures had to be as simple as possible so as to make it possible for every instructor-pilot to run a subject through the program by simple following written instructions. Performance measurement had to be unambiguous and practical. A four-days-one-hour-per-day program was developed which leads flight naive subjects in 14 steps up to the handling of the simulator using pitch and stick while flying a complex maneuver (vertical-S) with all six degrees of simulator motion being active. Performance measurement is learning-objective oriented and highly objective. 44 subjects have been tested so far. All-subjects were able to handle the simulator with motion system on. Results vary over a wide range. A high degree of standardization among instructor pilots has been achieved. The system is ready for operation.


In Germany, aviation psychology looks back on an impressing history which had its beginnings way back in 1916 as some mythical accounts would have it. Although it is untrue that the late "Red Baron" made the acquaintance of aviation psychologists, it is certainly true to say that all German military pilots since the end of WW I have been confronted with aviation psychology in one way or another, if not with an actual aviation psychologist, then at least with aviation psychology methods and instruments. As a general rule, such instruments would include paper and pencil tests, and boxes with all kinds of levers, buttons, lights and bells. In the sphere of aviation, psychology was essentially synonymous with pilot candidate selection. Presumably
this is also true for other countries where aviation psychology is practiced. On the other hand, aviation psychologists were surprisingly hesitant in touching two other important areas of aviation, namely - pilot training - psychological support for aviators.


This project investigated the relationship between Air Force Officer Qualifying Test (AFOQT ) composite scores and student performance in Air Force air weapons controller training. The purpose of this study was to examine the feasibility of using AFOQT scores as one of the selection criteria for entry to the air weapons controller field. A analysis of training performance data for 968 students showed a significant positive correlation between AFOQT Academic Aptitude composite scores and successful completion of training . Thus, a range of possible minimum cutoff scores on this composite was recommended as a selection criterion.

Data concerning the appropriate cost of attrition rates ( Fiscal Year 1982 ) for the five air weapons controller training organization and the expected effects of establishing various minimum selection criteria were also presented . A separate data analysis of demographic factors, including age, and source of commission showed no consistent relationship between these variables and student performance that would be useful in establishing criteria based on these variables . The decision to use AFOQT scores as one of the selection criteria is now under management review at Air Force Headquarters.


This article summarizes the work conducted by various government agencies to improve the selection of aviation cadets during the period around World War II. Three different predictors were tested using three different samples, each of whom joined the service under somewhat different circumstances (e.g., before or after the war). The first predictor was the Wonderlic Personnel Test (PT), which was later to become the Aviation Classification Test (ACT). It was a group-administered intelligence test (112 items) that tapped vocabulary, following directions, arithmetic reasoning, etc. The Mechanical Comprehension Test (MCT) was the second predictor and it is a 76-item test that measures knowledge of "barnyard physics." Finally, the Biographical Inventory (BI) was used, which has 150 items that measure background variables, interests, habits, preferences, etc. The BI and the MCT were later combined to form an overall Flight Aptitude Rating (FAR). This paper contains some of the first published validity data for these tests, which are still in use today (e.g., the BI and MCT). A variety of criteria were examined including training outcome (pass/fail) both for flight school and ground school, reason for failure in flight training and number of flight hours (although validities for number of flight hours are not reported). The validities for the PT were generally low across all three samples, although validities for predicting performance in ground school were moderately high (.20 to .31; sample sizes were very small; Ns ranged from 24-45). The validities of the MCT for predicting flight/ground school training were moderately high (ranging from -.25 to .33 across the three samples). Finally, the BI did not predict ground school failures (or "Other Types of Failures") very well, but it did do a good job of predicting flight training failures (validities ranging from .29 to .34 across these samples).
SVT (by electrocardiographic [ECG] documentation or suggestive symptoms). RESULTS: The mean age was 34.3 years (range 17-56 years). Forty-two (42/238, 17.6%) had SVT (WPW syndrome) and 196 of 238 (82.4%) had the WPW ECG pattern only. The mean follow-up of 21.8 years (range 2-41 years) was obtained on 228 of 238 (96%) for a total of 4906 patient-years. Sudden cardiac death occurred in 1 of 228, an incidence of 0.0002 per patient-year (95% confidence interval 0.0-0.001). SVT occurred in 47 of 228 (20.6%) or 0.01 per patient-year. One hundred eighty-seven (187/228, 82%) initially had the WPW ECG pattern only; 28 of 187 (15.0%) reported SVT during follow-up. Forty-one (41/228, 18%) initially had WPW syndrome; 19 of 41 (46.3%) reported additional SVT during follow-up. CONCLUSION: Sudden cardiac death risk was low (0.02%/patient-year) in this WPW population. The SVT incidence was 1% per patient-year. Referral bias and some characteristics of the unique military aviator population may partly account for these low event rates. However, these results may be more applicable to unselected populations than are tertiary referral-based studies.


During the past ten years the writer and various collaborators have been engaged in developing and utilizing a method that has been named the "critical incident technique." It is the purpose of this article to describe the development of this methodology, its fundamental principles, and its present status. In addition, the findings of a considerable number of studies making use of the critical incident technique will be briefly reviewed and certain possible further uses of the technique will be indicated. The critical incident technique consists of a set of procedures for collecting direct observations of human behavior in such a way as to facilitate their potential usefulness in solving practical problems and developing broad psychological principles. The critical incident technique outlines procedures for collecting observed incidents having special significance and meeting systematically defined criteria. By an incident is meant any observable human activity that is sufficiently complete in itself to permit inferences and predictions to be made about the person performing the act. To be critical, an incident must occur in a situation where the purpose or intent of the act seems fairly clear to the observer and where its consequences are sufficiently definite to leave little doubt concerning its effects. Certainly in its broad outlines and basic approach the critical incident technique has very little which is
new about it. People have been making observations on other people for centuries. The work of many of the great writers of the past indicates that they were keen observers of their fellow men. Some of these writers must have relied on detailed notes made from their observations. Others may have had unusual abilities to reconstruct memory images in vivid detail. Some may have even made a series of relatively systematic observations on many instances of a particular type of behavior. Perhaps what is most conspicuously needed to supplement these activities is a set of procedures for analyzing and synthesizing such observations into a number of relationships that can be tested by making additional observations under more carefully controlled conditions.


Scores obtained at eight different stages of practice on the Complex Coordination Test together with scores on 18 reference tests were subjected to a Thurstone Centroid Factor Analysis. Nine meaningful factors were identified in the experimental battery. The results indicated considerable, but systematic, changes in the factor structure of the Complex Coordination Test as practice on the task was continued. The test became less complex (factorially) as practice was continued. Moreover, there was a change in the nature of the factors contributing variance at early and later stages of practice. Implications of the findings are related to certain problems of learning theory, psychomotor test development, and criterion analysis.


Assessing work requirements across the universe of job classifications represents a major problem confronting those responsible for personnel decisions. The dimensions on which work vary are numerous, insufficiently defined, and complicated by continuous automation and job redesign. In some cases, the establishment of dimensions, categories, or grouping of job requirements is an arbitrary process originating from practical necessity rather than from systematic empirical or then ethical rationale. It is clear that the absence of a complete taxonomic structure of human work classification reduces the effectiveness of personnel decisions regarding success of an individual worker within the demands of the job.


The cognitive paper and pencil tests known as the Aircrew Aptitude Test Battery (AATB), currently being used for pilot and navigator selection at the CF Aircrew Selection Centre (CFASC), are becoming dated. These tests were developed to meet selection requirements during World War II (Ferguson & Sloan, 1954; Knoll, 1950; Noble & Manning, 1948), later validated for Canadian use (Lockwood, 1965; McInnis, 1965), and more recently, psychometric characteristics were examined in James (1986). Though proving successful as an aircrew selection measure over the years, the tests are based on the traditional requirements of aircraft systems operating almost fifty years ago. There are a number of reasons for updating the current aircrew selection test battery. Contemporary aircraft have become increasingly sophisticated, with technological advances resulting in the automation of many functions previously performed by aircrew. However, the complexity of modern aircraft, which utilize advanced visual displays and computer technology, place significantly higher cognitive workload and information processing demands on aircrew. Furthermore, new testing methods have been introduced since the AATB was first used. It is desirable that these new methodologies be employed in the design and evaluation of the new tests. During the 1970s, the
US Air Force Human Resources Laboratory developed and tested an experimental cognitive test battery designed to improve the selection of aircrew officers. The tests that showed promise as good predictors of pilot (Hunter & Thompson, 1978) and navigator (Valentine, 1977) training performance were introduced at CFASC in early 1983 as the Experimental Aircrew Selection Battery (ASB-X). A French-language version of the ASB-X was developed and experimentally introduced at the CFASC in 1984. The ASB-X, comprised of eight multiple-choice tests, is administered to CFASC aircrew candidates as part of their regular selection processing.


The Air Traffic Controller Aptitude Test (ATCAT) has been used since 1985 as an experimental Air Traffic Control (ATC) selection measure. This paper presents the history, content and previous research on the ATCAT and related U.S. test on which the ATCAT is based. Item level data was gathered on 155 students and criterion data was obtained for 1.12 students who completed the ATC course. As was expected from U.S. research, psychometric analysis revealed that the ATCAT is a highly speeded test with upwardly biased reliability measures. The test operates similarly for both anglophone and francophone groups. Validation analyses produced non-significant results for the anglophone and combined groups. The francophone correlation was unexpectedly significant; however, this was found to be due to the effects of the ATC Technical Vocabulary Course. It was recommended that the experimental use of the ATCAT he discontinued. A new ATC measure, the Terminal Option Controller Test (TOCT), has been identified and will be introduced as an experimental selection measure. Strategies to increase validation sample size will be explored.


The Aircraft Landing (AL) test is a pilot selection device in which attempts are made at simulated approaches and landings of an aircraft on a runway until a criterion skill level is achieved. The test was designed with principles from the information processing model of skilled performance in mind, and two concepts from the model were tested: hierarchical mechanisms and feedback. Using Canadian Forces pilot trainees, validities up to 0.45 (n = 104, no previous flying experience) and 0.49 (n = 26, previous flying experience) were obtained against a criterion score based on flying tests at 7 and 12 h in a light aircraft. It was concluded that the practical utility of the test was highly satisfactory and that its validity could be increased by modifications. The pattern of results provided some support for the theoretical predictions from the information processing model, but methodological difficulties prevented a clear-cut interpretation.
Advantages of this approach over an abilities model are discussed.


This is a review of the Martin Scorsese motion picture, The Aviator. This motion picture is a biographical treatment of Howard Hughes in the narrow time frame of 1928 to 1946. It emphasizes his developing psychological disorders along with his innovations in aviation.


Measures of situation awareness (SA), or what operators know about their immediate situation, are reviewed. Three major approaches to SA assessment are considered: explicit, implicit, and subjective rating. Explicit measures require operators to self-report material in conscious memory. Implicit measures assess the influence of relevant events on subsequent task performance. Subjective ratings require operators to assign numerical values to the self-assessed quality of their SA. These three measurement approaches are evaluated in terms of their reliability and three kinds of validity: construct, content, and criterion. Several problems requiring further research are identified and discussed. In particular, reliability and content validity continue to present serious difficulties, some of which suggest that new approaches to SA measurement may still be needed.


Both explicit and implicit measures of situation awareness (SA) were evaluated in a series of experiments in order to assess their reliability and two kinds of validity: criterion and construct. In all of the experiments, subjects performed a simulated combat task in which they had to monitor the positions of enemy, friendly, and neutral objects. In addition, subjects had to attack and defend themselves against enemy objects. A memory probe procedure was used to explicitly assess two components of SA: location and identity awareness. In addition, a signal detection analysis was used to provide an implicit measure of SA. Test-retest correlations indicated that the location awareness measure was much less reliable than the others. Criterion validity was evaluated by correlating the SA measures with probability of a kill in the combat task. Although the SA metrics seemed to be fairly good predictors of kill probability, the best predictor was a measure of behavioral workload. Predictions of multiple resource theory were used to evaluate construct validity. In particular, it was predicted that difficulty in maintaining identity awareness would not affect location awareness, and this prediction was largely supported. Nevertheless, other aspects of the data seemed to contradict current versions of multiple resource theory.


The situation awareness (SA) and mental workload of 56 subjects were evaluated as they monitored one or more attributes of six objects moving systematically over a rectangular grid. Subjects were assigned to one of seven groups depending upon whether they were to monitor object locations (location task), object colors (color task), whether the objects flashed (flash task), or some combination of these three. Both task performance and subjective ratings were
used to assess subjects' awareness of the three object attributes. In addition, subjective ratings of mental workload were collected. All subjects performed the monitoring task under four different conditions formed from the factorial combination of 1) the probability that objects of a certain color would flash and 2) whether object colors remained consistent or changed during the course of a trial. The results pointed to the usefulness of both flash and color task performance as measures of SA. Subjects were very poor at the location task, suggesting either their location awareness was poor or the location task is not a good measure of that awareness. Subjective ratings proved useful but occasionally dissociated from task performance. One possibility is that subjective ratings reflect rational inferences by the subjects rather than the outcome of their introspections.


This paper briefly describes the development and content of the 1992 revision of the Navy and Marine Corps Aviation Selection Test Battery (ASTB). The ASTB is the primary instrument for selecting personnel into the student naval aviator (pilot) and student naval flight officer (navigator) program for the U. S. Navy, U. S. Marine Corps, and the U. S. Coast Guard. The first Naval aviation selection test was implemented in 1942 with revisions occurring in 1953 and 1971. Specifics of these early versions have been documented elsewhere (Griffin & Mosko, 1977; North & Griffin, 1977; Baisden & Holcombe, 1991) and will not be addressed here. In 1984 a program was begun to revise the Naval aviation selection test. There were five primary reasons for initiating a new test: (1) changes in the demographics of the applicant population (e.g., an all volunteer force, educational curricula); (2) changes in Naval aviation training (e.g., increased use of simulators) and operational aircraft (e.g., glass cockpits); (3) possible compromises in the test since its revision 13 years earlier; (4) decreased predictive validity; and (5) changes in federal guidance regarding employee selection procedures.


Situational Awareness (SA) is considered essential to safe and expert pilot performance. The computer-based WOMBAT™ test provides a multi-tasking environment designed to assess a set of abilities thought to be important for the maintenance of SA in complex environments. There is some evidence to suggest that performance on this test is related to elite pilot performance. This paper reports the results of a study assessing the predictive validity of the test
with regard to the training performance of ab-initio pilot students. Performance measures included official Flight Test scores as well as the number of flight hours to criterion performance. An analysis of data provided by approximately 30 student pilots was undertaken. The results did not support the position that WOMBAT scores have predictive validity in relation to the employed criterion variables. However, situational awareness data derived from a subsequent simulator study did show the expected relationship. The results of these studies are discussed with respect to the theoretical and practical considerations pertaining to selection, training and performance. Consequent recommendations center around the alignment of selection, training and performance measures. It is argued that the use of detailed and explicit developmental markers can result in both a robust means of assessing predictive validity and increased efficiency and effectiveness of flight training.


The authors state that much data has been collected on the student military aviator, but significantly less attention has been focused on the operational military pilot. This study compared the performance of three groups (jet aviators, a general population normative sample and a normative group of college-educated males) on several psychological tests. Of 298 pilots reporting to several Carrier Replacement Air Wing squadrons, 288 participated. Subjects completed a personal history questionnaire, the Maudsley Personality Inventory and the Edwards Personal Preference Schedule (EPPS). None of the inventories are described in any detail in the article. The jet aviator group differed significantly from the EPPS General Adult Male norm group on all scales except Intraception. This group differed significantly from the EPPS College Male norm group on all scales except for Need for Achievement and Exhibition. The EPPS has also been given to several classes of student flight surgeons and, with a few conceptually meaningful exceptions, the pilots in this study tended to score similarly to the student flight surgeons. The differences were that the student flight surgeons scored higher on Nurturance and lower on Dominance. As compared to the general male population, the jet aviators scored higher on the Heterosexuality, Dominance, Change, Achievement and Exhibition scales and lower on the Nurturance, Abasement, Deference, Order and Succorance scales.


One of the pivotal contributions of Military Psychology to improving military personnel selection has been the development and utilisation of structured, objective tests of ability and aptitude. In contradistinction to the cognitive domain, until recently there has been little success in the application of non-cognitive constructs to the selection of pilots. Researchers have suggested that less intrusive measures than those currently employed may be required to assess the ‘deeper’ aspects of the non-cognitive domain, such as the ‘dark side’ of personality. In this paper, data from a study of a novel application of Computer Aided Text Analysis (CATA) to the selection of RAAF pilot trainees is further investigated. Given previous findings that suggest CATA has the potential to be a useful tool in pilot selection, this study elaborates reasons for pilot training failure beyond the bivariate pass/fail criterion previously employed. The use of a bivariate pass/fail criterion introduces the potential for attenuation of prediction given the possible inclusion of failures for cognitively based reasons. In this study, the training suspension reports (TSR’s) were manually content analysed and ‘failure themes’ extracted. The TSR’s were encoded and results subjected to a Principal Components Analysis to determine if a latent component structure could be observed. From the component structure matrix a number of scores were calculated for each applicant. Themes extracted from a 20 minute communication skills essay were then used to investigate subsequent pilot training performance. The results hold relevance for the use of CATA as a military selection tool, specifically for pilot and more generally as a means of assessing non-cognitive domains in personnel selection.


Alcohol-related disorders are the most prevalent psychiatric conditions in the aviation population. Efforts to effectively screen aviators for these disorders are continually sought, as under-diagnosis may negatively impact aviation safety. This study evaluates screening tools that have been validated in non-aviators in terms of their utility for aviator patients. Methods: There were 111 male aviation patients (27 ± 7 yr) referred for psychiatric evaluation at the Naval Aerospace Medicine Institute who completed the Self-Administered Alcohol Screening Test (SAAST), the Alcohol-Use Disorders Identification Test (AUDIT), and the Common Alcohol Logistical Scale-Revised (CAL-R) prior to evaluation by a staff psychiatrist or psychologist. Results: There were 40 patients who were qualified psychiatrically with no diagnosis and 49 patients who were disqualified for psychiatric reasons due to a non alcohol-related diagnosis. The remaining 22 patients were disqualified for psychiatric reasons with an alcohol-related diagnosis. The optimal aviator cut-off scores were consistent with those of the general population, although the cut-off score used for the SAAST was set at the published sub-threshold.
level to provide greater sensitivity. The sensitivity/specificity values for the SAAST, AUDIT, and CAL-R were 59%/94%, 46%/96%, and 68%/81%, respectively. Conclusion: The psychometrically sophisticated CAL-R is sensitive, specific, and has good negative predictive value, although its use requires a psychologist and its availability is limited. The SAAST and AUDIT can be administered by a flight surgeon or aviation medical examiner (AME). Given the higher sensitivity of the SAAST it may be the most beneficial if administered first. The AUDIT can be used as a follow-up diagnostic test given its higher specificity.


This study used the Minnesota Multiphasic Personality Inventory (MMPI) to compare the scores of 15 male Army helicopter pilots to 16 male non-pilot Army officers. Differences in mean standard scores obtained by these two groups on the MMPI scales were compared using t-tests. Results showed that pilots scored significantly higher than non-pilots on several scales: Depression, Hypochondriasis, Hysteria, Social I-E (all significant at p < .05) and Psychopathic Deviate (p < .01). However, the authors used t-tests to compare scores across the two groups on 13 separate scales. Thus if a Bonferroni correction was used to counteract the increased probability of committing a Type I error that comes from making this many comparisons (.05/10 comparisons = .005 [this alpha level does not include the Lie, F and K-correction MMPI scales]), the pilots would have only scored significantly higher on the Psychopathic Deviate scale.


Sorts assessment instruments as either job representational or job analagous and as either empirical or analytical. Describes several tests and suggests the use of emotional behavior.


Since World War II, there has been a continuing effort to improve the tests used to select aircrew (10). Aptitude tests and biographical inventories have been updated periodically, and new tests have occasionally been added. Despite approximately 40 years of effort, the pilot composite of the United States Naval Aviation Officer Selection Battery has an uncorrected
predictive validity correlation of approximately +.15 to +.25 with a pass/fail criterion for undergraduate pilot training. The Air Force Officer Qualifying Test has predictive validities that are typically in the same range (7). Because of the escalating costs of training aircrew, it has become increasingly important to improve the predictive validity of aircrew selection batteries. At present, the general consensus of the selection community is that the existing paper-and-pencil tests fail to test adequately four major areas of individual differences that could increase the predictive validity of aircrew selection batteries: psychomotor skills, information processing abilities, higher-order cognitive processes, and personality (2,5,7,9,10). The lack of psychomotor tests in the existing aircrew batteries is an historical anomaly; during World War II, both the Navy and the Army Air Corps aircrew selection batteries included extensive apparatus tests to evaluate psychomotor skills. In the early 1950s, apparatus tests were eliminated from both batteries because of problems with calibration and reliability. Subsequently, researchers assumed that any psychomotor tests would present similar problems. More recently, advances in microprocessors have eliminated calibration and reliability problems and have made large-scale testing feasible. Recent Air Force studies (7,9) of two computer-generated psychomotor tests demonstrated that scores from the two psychomotor tests made unique contributions to prediction beyond that contributed by the existing paper-and-pencil aircrew tests. Because of these encouraging results, these two tests have been added to the Air Force aircrew selection battery.


This report describes the validation of an automated aircrew selection test battery that measures cognitive processes, psychomotor skills, and time-sharing abilities. Results indicate that performance-based test measures can be used to predict flight training performance.


This is the first of three brief papers that summarize the history of aviation medicine in the Royal Air Force. Just as the generals and politicians were slow to appreciate the potential of the airplane, so was the medical establishment slow in understanding that the flight environment involved medical and physiological challenges. This note outlines the development of research to support British military aviators up to the formation of the Royal Air Force in 1918.


Incorporation of a dichotic listening task of selective attention in the pilot selection test battery of the Israeli Air Force was studied using a group of 2000 flight cadets. In this test, subjects are presented with 48 auditory messages. Each message is composed of strings of words and digit names. Different strings are simultaneously presented to the two ears. Subjects are required to detect digit names in the relevant channel and to reconsider channel relevance upon indication. Three types of selective listening errors are recorded: omissions, intrusions, and switching errors. Flight cadets who had completed a two-year training program had significantly lower error scores on all attention measures. In addition, these measures had low correlations with all other tests of the pilot selection battery. Thus, attention capabilities appear to be an independent dimension that enhances the predictive validity of the present test battery.


It is the purpose of this paper to report certain aspects of a study conducted by the Aviation Branch of the American Institute for Research under the auspices of the National Research Council Committee on Aviation Psychology. Funds for the project were furnished by
the Civil Aeronautics Administration. This study, completed in November, 1947, was undertaken (1) to study current methods of selecting and evaluating the airline pilot and (2) to determine the critical requirements of his job. It was intended that the data obtained in this investigation be used as a basis upon which to develop improved procedures for selecting, training, and certifying airline pilots. At present the American Institute for Research is utilizing the data as a basis for devising a radically new type of flight examination for pilots seeking the Airline Transport Rating certificate. This latter project is under the same sponsorship as the study to be described in this paper. In the first phase of the study the general procedure followed was to survey the available sources of information pertaining to present methods of selecting and evaluating airline pilots. In the second phase of the project the procedure was to survey sources of information about the critical requirements of the airline pilot's job, an attempt being made to answer the question: "What behavior and characteristics are required for handling the job safely and effectively?"


This paper: 1) describes the need for a human-factors taxonomy; 2) identifies existing taxonomies from the scientific, training, test and evaluation, and mission modeling disciplines; 3) lists the rules used in combining these taxonomies into a single, coherent taxonomy; 4) presents the taxonomy at its top levels; and 5) provides a source for obtaining a copy of the AIAA Human-Factors Taxonomy standard.


We compared the cancer incidence of male United States Air Force (USAF) aircrew (342 cancers, 532,980.97 man-years) with non-flying Air Force officers (827 cancers, 1,084,370.08 man-years) between 1975-89. METHODS: Incident cancer cases for both aviators and non-flying officers were obtained from USAF hospitalization records. Age-adjusted standardized incidence ratios (SIR's) were calculated for aircrew using data from the National Cancer Institute's Surveillance Epidemiology and End Results (SEER) program. Aviator age-adjusted cancer rate ratios were also obtained using non-flying officers as an internal comparison group. RESULTS: We observed statistically significant excesses of aircrew cancers for all sites, testis, and urinary bladder. All other aviator cancer classifications were not significantly different from the comparison cohort; most notably, cancers of the colon and rectum, skin (both malignant melanoma and non-epithelial), brain and nervous system, Hodgkin's Disease and leukemias. CONCLUSION: Previous studies of commercial pilots that demonstrated excesses of these cancers may have been biased by the use of external comparison groups. We used an internal comparison population to reduce selection bias, information bias and confounding. From these data we detected notable excess aircrew cancer risk for cancers of the testis, urinary bladder, and


Grether, W. F., Melton, A. W., & Kossman, C. E. (1944). Development of the S.A.M. steadiness aiming test, Form B, for use in the selection and classification of aircrew personnel. Randolph Field, TX: 27th AAF Base Unit, AAF School of Aviation Medicine. The S.A.M., Steadiness Aiming Test, Form B, proved to be a mechanically satisfactory measure of hand steadiness. The test had excellent reliability. Validity of the test, however, was very low for prediction of success in elementary pilot training. The addition of a verbal stress element to the S.A.M. Steadiness Aiming Test did not result in an increase in validity.


To address the selection-related question, “What does the personality profile of the Army aviator of today look like?” 75 experienced Army aviators attending advanced leadership training completed the Revised NEO Personality Inventory, with scores depicting the five personality factors of: neuroticism, extraversion, openness, agreeableness, and conscientiousness. To address the classification-related question, “Are there certain personality profiles that distinguish among attack, scout, cargo, and utility pilots?” factor scores and their subsumed facet scores were compared across respondents representing the four mission platforms. Overall sample profiles and score differences among platforms are presented.

The U.S. Army Research Institute for the Behavioral and Social Sciences Rotary Wing Aviation Research Unit (ARI RWARU) administered the Revised NEO Personality Inventory to 217 student Army aviators awaiting Initial Entry Rotary Wing training. Scores reflected the incoming aviators’ standings on five personality factors: neuroticism, extraversion, openness, agreeableness, and conscientiousness. The male student factor and facet scores were then compared with a sample of male career Army aviators. Personality differences and similarities between the two samples are discussed as laying the foundation for longitudinal research.


Attrition in undergraduate naval aviation training represents a costly problem. An average of 25% of student naval aviators fails to complete training. This study reports an effort to develop automated single- and multiple-dichotic listening and psychomotor tasks, which have the potential to reduce aviator attrition through improved selection and may also be useful in initial pipeline classification. Statistical analysis of Study I, comparing a forward and backward series of automated dichotic listening (DLT) and psychomotor (PMT) tasks, indicated that a backward-direction orientation associated with the psychomotor tests resulted in increased difficulty for all PMT measures and two of three multitask DLT measures. The correlational estimates of test-retest reliability for the multitask DLT and PMT measures were adequate for both series of automated tasks but slightly higher (r - .80 DLT, r - .90 PMT) for the backward series. There were significant correlations between the DLT and PMT tasks, for each series of single- and multitask measures. The relation between such seemingly different tasks is difficult to understand since the DLT is an auditory cognitive processing task, and the PMT is an eye, hand, foot coordination task. However, the significant correlations were both smaller and fewer in number for the backward series of automated tests. Study II was a correlational evaluation between the new automated multitask measures and old nonautomated tasks with demonstrated validity for the prediction of primary flight performance. The correlations between corresponding tasks of the new automated and old nonautomated tasks averaged .60 for the DLT measures and .66 for the PMT tasks. The results of Study III indicated that certain automated DLT and PMT measures were significantly related to primary flight grades (PFG) in Navy flight training. For the backward series of tasks, all DLT and PMT measures were significantly correlated with PFG. However, only two DLT and two PMT measures of the forward series were significantly related to FFG. No significant correlations were found between the automated DLT and PMT tests and the pass/fail criterion. The absence of a suitable number of flight failure attritions was discussed as a possible reason for this result. A regression analysis for the backward series of test measures and primary flight training criteria indicated that a psychomotor stick and rudder measure and the FAR selection test were significantly related to PFG (R = .53, F (2,85) = 16.56, p < .000 1). There were no significant correlations between the automated DLT and PMT measures and prior flight hours. These results indicate that a series of automated DLT and PMT tasks are suitable replications of an older version of nonautomated tasks. The advantages of the automated tasks are that they require less administrative support and provide automatic scoring of performance. The backward series of automated tasks, which was correlated
more strongly with criterion performance, should be administered to a large sample of student naval aviators to determine if the tests can account for additional variance in the prediction of flight training performance beyond that of current selection tests.


A series of automated psychomotor and dichotic listening tasks, which require little administrative support and provide automatic scoring of performance, has been developed. The automated tasks account for additional variance in predicting Navy flight training performance beyond that of current selection tests.


Evaluates the prediction of naval aviator flight training performance in whiting Field, Florida. Use of multiple regression and an artificial neural network; Conduction of cognitive and perceptual psychomotor tests; Efficiency in predicting flight grade training performance.


The numerous Navy research efforts to identify motivational predictors of student naval aviator and/or student naval flight officer attrition have met with little success. In spite of the failures associated with the application of personality or motivational measures to naval aviation populations, there is little doubt that much of the attrition in naval aviation training is of a motivational origin. As a result, research personnel continue to seek motivational measures that may identify those individuals most likely to succeed in training. The present report describes an evaluation of the Omnibus Personality Inventory (OPI) as a predictor of student motivational attrition in naval aviator flying training. Initial validation results suggested that certain OPI scales were predictive of student naval aviator and student naval flight officer success in naval flying training programs. A cross-validation analysis was conducted to determine the stability of these findings. The analysis indicated that significant cross-validity relationships existed for current selection tests but not for OPI predictor measures. It must be concluded that the OPI is not sufficiently related to student aviator training performance to be of value in the prediction of aviator motivational attrition. These results support previous conclusions that future aviation selection research should be directed toward the identification of performance oriented, non-paper-and-pencil measures as motivational predictors.


Focuses on the historical perspective of the use of psychomotor, perceptual-cognitive paper-and-pencil and tests for the selection of pilot trainees by the military services in the U.S. Combination of paper-and-pencil and automated psychomotor tests for initial selection; Use of automated cognitive and psychomotor tests; Predictors of pass-fail criteria. Focuses on the
historical perspective of the use of psychomotor, perceptual-cognitive paper-and-pencil and tests for the selection of pilot trainees by the military services in the U.S. Combination of paper-and-pencil and automated psychomotor tests for initial selection; Use of automated cognitive and psychomotor tests; Predictors of pass-fail criteria.


Attrition in undergraduate naval aviation training represents a costly problem. An average one in four student naval aviators fails to complete training. This study represents an attempt to reduce aviator attrition through improved selection. A multitask experiment simulating certain motor control and communication requirements characteristic of flight was conducted to determine relative strengths of several performance measures as predictors of primary flight training success.


A difficult aspect of predicting fleet pilot performance is acquiring meaningful and reliable, inflight criteria. Without such criteria, performance assessment is both theoretically and realistically impossible. This study was an attempt to predict Air Combat Maneuvering (ACM) performance using performance-based laboratory tests and to evaluate the VF-43 adversary squadron’s grading of inflight ACM performance in the Fleet Fighter ACM Readiness Program at Naval Air Station Oceana. The purpose of the latter effort was to select convenient and reliable criteria for ACM performance assessment and use in the validation of the laboratory tests. In an initial evaluation (Study I), F-4 pilots performed in Fleet Fighter ACM Readiness exercises and completed performance-based perceptual motor and multitask tests. Results indicated that dichotic listening test measures, obtained during multitask conditions, could be used to reliably predict ACM inflight criteria. Results of a larger sample of F-14 pilots (Study II) indicated that an overall ACM grade (OAG) assigned by VF-43 adversary personnel can be predicted reliably by an objective kill difference composite score and three subjective measures: situational awareness, mutual support, and energy management. These four measures accounted for 78% of the variance with the OAG. A correlational analysis suggests that the VF-43 grading process is reliable and consistent. Additional results were obtained on the relation between the Naval Aerospace Medical Research Laboratory vision tests and ACM criteria (Study III). Contrast sensitivity measures were significantly related to a mean time-to-first-kill measure. Visual acuity and accommodative flexibility measures were significantly related to the initial sighting (tally - ho) and visual identification (VID) of adversary aircraft on an instrumented range. Age and/or experience in ACM may be an important variable in relating vision tests to pilot performance. It is recommended that: (1) improved performance-based tests should be administered to a sample of Navy pilots performing in Fleet Fighter ACM Readiness Evaluations to replicate initial test results; and (2) an overall ACM grade regression equation should be applied to a supplementary sample of pilots performing in Fleet Fighter ACM Readiness exercises to confirm the reliability and validity of the VF-43 adversary squadron’s grading process.
Attrition in undergraduate naval aviation training represents a costly problem. An average of 25% of student naval aviators fails to complete training. This study reports an effort to develop automated single- and multiple-dichotic listening and psychomotor tasks, which have the potential to reduce aviator attrition through improved selection and may also be useful in initial pipeline classification. Statistical analysis of Study I, comparing a forward and backward series of automated dichotic listening (DLT) and psychomotor (PMT) tasks, indicated that a backward-direction orientation associated with the psychomotor tests resulted in increased difficulty for all PMT measures and two of three multitask DLT measures. The correlational estimates of test-retest reliability for the multitask DLT and PMT measures were adequate for both series of automated tasks but slightly higher ($r = .80$ DLT, $r = .90$ PMT) for the backward series. There were significant correlations between the DLT and PMT tasks, for each series of single- and multitask measures. The relation between such seemingly different tasks is difficult to understand since the DLT is an auditory cognitive processing task, and the PMT is an eye, hand, foot coordination task. However, the significant correlations were both smaller and fewer in number for the backward series of automated tests. Study II was a correlational evaluation between the new automated multitask measures and old nonautomated tasks with demonstrated validity for the prediction of primary flight performance. The correlations between corresponding tasks of the new automated and old nonautomated tasks averaged .60 for the DLT measures and .66 for the PMT tasks. The results of Study III indicated that certain automated DLT and PMT measures were significantly related to primary flight grades (PFG) in Navy flight training. For the backward series of tasks, all DLT and PMT measures were significantly correlated with PFG. However, only two DLT and two PMT measures of the forward series were significantly related to FFG. No significant correlations were found between the automated DLT and PMT tests and the pass/fail criterion. The absence of a suitable number of flight failure attritions was discussed as a possible reason for this result. A regression analysis for the backward series of test measures and primary flight training criteria indicated that a psychomotor stick and rudder measure and the FAR selection test were significantly related to PFG ($R = .53$, $F (2,85) = 16.56$, $p < .0001$). There were no significant correlations between the automated DLT and PMT measures and prior flight hours. These results indicate that a series of automated DLT and PMT tasks are suitable replications of an older version of nonautomated tasks. The advantages of the automated tasks are that they require less administrative support and provide automatic scoring of performance. The backward series of automated tasks, which was correlated more strongly with criterion performance, should be administered to a large sample of student naval aviators to determine if the tests can account for additional variance in the prediction of flight training performance beyond that of current selection tests.


Several studies have suggested the possibility of predicting operational performance in fleet aviation environments. This report concerns the use of an automated performance-based test battery, involving cognitive and psychomotor functioning, to predict the operational performance of fighter pilots. Two groups of pilots who were completing fleet replacement squadron (FRS) training for the F/A-18 were tested on this battery. The older and more experienced pilot group got higher FRS grades than did the other group; test performance between these two groups was not significantly different. Those few significant correlations found between the test measures and the FRS grades were illogically patterned and of insufficient quantity or strength to demonstrate any reliable predictive ability. This could have been due to the homogeneous nature of each of these subject groups in terms of piloting skills and abilities. Keywords: Pilot selection; Pipeline assignment; Psychomotor ability; Cognition; Dichotic listening; Multitask performance; Fleet aviation performance; Jet fighters; Flight training. (jg)


Since 1947, the Navy has relied on the Aviation Qualification Test (AQT) of general ability and the Flight Aptitude Rating (FAR) -- a composite of mechanical comprehension, spatial aptitude, and biographical tests--to select naval aviators. While this selection system has served the Navy well, the failure rate of pilot selectees has remained consistently high (although certainly less than the attrition rate before using a selection test battery), averaging 20-25% over the last 20 years. The cost of these failures represents millions of Navy training dollars lost each year. As a result, researchers at the Naval Aerospace Medical Research Laboratory have attempted repeatedly to improve selection test batteries. While previous efforts have failed, new research results appear to have the potential to improve our selection of pilots and naval flight officers.


The purpose of this study was to develop empirical keys to predict a variety of training outcomes using the Strong Vocational Interest Blank (SVIB) and the Officer Biographical and Attitudinal Survey (OBAS). Subjects were 593 pilot candidates who entered pilot training through Officer Training School (OTS). It is interesting to note that UPT trainees from OTS represented the largest percentage of eliminees. The SVIB has 54 occupational keys and five supplemental keys. The OBAS is a 116-item inventory which contains a variety of background and attitudinal items. It also contains the importance-possibility scale, which measures the importance of certain career needs and how likely it is that those career needs can be met by the Air Force. The predictive utility of the Air Force Officer Qualifying Test (AFOQT) score was also examined. Four criteria were used: total elimination (UPT pass/fail), Flying Deficiency Elimination (FLY), Self-Initiated Elimination (SIE) and Motivational Deficiency Elimination (SIEMOA). Regression analyses were conducted to examine the prediction obtained using the keys and the number of pilots correctly classified as passing or failing UPT using several different prediction models. The OBAS keys demonstrated significant cross-validity only for UPT pass/fail (R = .13, p < .05) and the FLY (R = .14, p < .05) criteria. Seventeen of the SVIB occupational scales demonstrated significant zero-order validities with UPT pass/fail resulting in a multiple correlation of .45 (not cross-validated). Three different models combining a variety of these predictors were tested. Model 1 consisted of the 17 SVIB scales, the four OBAS keys and the AFOQT (cross-validity = .14, p < .05). Model 2 consisted of just the 17 SVIB scales and the AFOQT (cross-validity = .111 ns). Finally, Model 3 consisted of the four OBAS keys and the AFOQT (cross-validity = .14, p < .05). With respect to classification of pass or fail, Model 1 was the most efficient, resulting in an overall correct classification of 71 percent. The authors conclude that non-cognitive data have practical value in the selection of pilot trainees.


Psychological diagnostics are often charged with predicting how well a person will
perform under LOAD in a precarious situation. It is all the more paradoxical that the data upon which such prognoses have traditionally been based are invariably won under neutral, if not to say sterile, conditions - a far cry from the actual situation under question. Sports and Traffic psychologists have undertaken to make amends for this shortcoming by simulating the real-life conditions in the laboratory. This strategy, however, has its price; for in so doing, the psychologist must make compromises in objectifying specific behavioral dimensions separately. As the laboratory simulation approximates the actual demands upon everyday performance for the behavior studied, the single, unique capability, dimensions become confounded, hopelessly intermingled. The effect is that prognoses regress to the range of chance.


A major focus of research in crew resource management (CRM) has been the association between personality and effective small-group leadership. In addition, the airline industry attempts to assess personality and temperament by placing pilot candidates through a two or three stage interview process. In an ongoing effort at Auburn University to develop a comprehensive a Pilot Candidate Selection Model, the authors have studied the relationship between the professional pilot's approach toward people as resources to accomplish goals and perception of effective leader behavior. Christie and Geis' Mach V scale (as a construct of skill in small group leadership) and a modification of the Bernard Bass Leadership Questionnaire (as a construct of leader perception) were employed in the survey of 98 Air Force officers 31 of which were serving actively as rated pilots. Research indicates a significant relationship in success in leadership of small groups and the Mach V scores. Using the measurements of the modified Bass questionnaire and the Mach V inventory as predictor variables, the authors
derived a discriminant function identifying pilots with ninety percent accuracy. Results indicate the potential use of the Mach V and the Bass Leadership Questionnaire as supporting instruments in the screening of candidates for commercial pilot training. The authors recommend extending current research to Part 141 flight training centers to validate the potential of the Mach V and the Bass instruments in evaluating students applying for commercial pilot training.


The purpose of this study was to re-explore possible factors that influenced the subjects’ visual-spatial information processing as they took mental rotation tests. Two types of mental rotation tests were designed to measure subjects’ visual-spatial abilities. One type of mental rotation task required the subject to distinguish figure’s rotation angle between <90 and > 90 degrees. All 177 subjects of four groups varying in flight experience (0 hr, 8-23 hrs, 82-85 hrs and instructor-pilots) were assessed with 80 stimuli. The results of the comparison among four groups indicated that the accuracy ratio was the most reliable factor in measuring individuals’ visual-spatial ability. The increase of complexity level could be considered to improve this experiment. Another type of mental rotation task required the subject to distinguish between an object and its mirror image. Different complexity levels of stimuli were used. All 70 subjects of two groups (passed and failed 12-15 hour flying training program) were assessed with 240 stimuli. The componential analysis was used to analyze the information process of the mental operation. The results indicated that the task’s internal validation was good and the estimation of the higher level of complexity stimuli was significant. The parameter pool of pilots selection module examined the parameters that estimated by the componential analysis. The results indicated that mental rotating tests, using higher level of complexity stimuli, could be considered as part of the test for pilot selection battery.


There are a variety of on-going attempts to generate unmanned aerial vehicle (UAV) technologies to exploit the advantages that these semiautomated and automated airborne platforms promise to render. (Although we refer specifically to UAVs here, our arguments apply, in principle, to all remote vehicles whatever their medium of operation. The principles themselves also extend to other forms of nontransport-based entities.) With regard to such operations, the collective community is searching for the ratio between operator(s) and vehicle(s) that will prove most efficient and effective.


The Selection of Air Force and Navy flight personnel is a progressive process, commencing before the enlistment of the candidates (Phases 1 and 2) and continuing after the normal military training (which lasts for approximately one year) into Phase 3. The first Phase is a general screening of such factors as Intelligence and Leadership qualities, carried out in the respective Officer or NCO Selection Centres. The second Phase is a preliminary flight-aptitude screening, using Computer-based psychological tests, grading candidates as broadly ‘Suitable’ or ‘Unsuitable’. The third Phase is more precise, making a final decision as to candidate suitability and further predicting what particular activity each candidate would be best suited for (e.g. Jet, WSO, Prop, Helicopter or Navigator). It consists of 3 weeks Navigation/Academic instruction, 1 week FPS 80 Selection and for those who have survived thus far, 5 weeks Flying instruction on light prop aircraft, including 18 flying hours. FPS 80 is the abbreviation for the Flight Psychological Selection System of the Aviation Psychology Section, Aerospace Medical Institute of the German Air Force. As the need was identified to improve the effectiveness and reliability of the Selection System, FPS 80 was conceptualised. It was then designed and a detailed Functional Specification was prepared, from which the required Hardware and Software was commissioned. FPS 80 was installed in July of 1987, from which time it was further tested and standardized. It was introduced as part of the selection process on the 1st April, 1990. In this paper, we will concern ourselves with a description and statistical evaluation of the FPS 80 Selection system.


For years pilot selection has focused primarily on the identification of individuals with superior flying skills and abilities. More recently, the aviation community has become increasingly aware that successful completion of a flight or mission requires not only flying skills, but the ability to work well in a crew situation. This paper describes the development, validation and computerization of a situational judgment test, called the Situational Test of Aircrew Response Styles (STARS), which targets the interpersonal skills necessary to function effectively in a crew situation. Recent research, especially in the crew resource management (CRM) area, highlights the importance of interpersonal skills and certain personality traits for effective pilot performance. This crew resource management research was motivated by analyses of the causes of aircraft accidents which showed that the majority of these accidents involved crew errors (Foushee, 1984). Further analysis indicated that aspects of the crews' interpersonal interactions such as breakdowns in coordination and communication most frequently played a causal role in these accidents (e.g., Cooper, White, & Lauber 1980). Helmreich (1987) hypothesized that pilot performance is determined by ability, personality, and attitudes, and suggested that since the first two variables are very difficult to change, crew resource management training should focus on changing attitudes. Thus, training can be viewed as a way to promote awareness of group dynamics, bring about attitude change, and improve interpersonal skills, but it does not change the underlying traits that have been shown to be related to crew resource management. However, selection based on personality traits and/or relevant interpersonal skills may enhance crew resource management performance, above and beyond what could be expected from crew resource management training alone.

This report describes research done by the Human Resources Research Organization Force on the first phase of project RELAY, The Impact of Military national Aspirations and Development of Skills. The overall purpose of the research is to increase the degree of benefits of an occupational nature achieved by military personnel, during and after active duty, through improvements in the interface and civilian occupational manpower systems. The research efforts are expected to contribute to this goal by providing information needed for better manage linkages and interchanges between the military and civilian manpower systems.


The Air Force Officer Qualifying Test (AFOQT) is an aptitude measure used to select officers, pilots, and combat system operators. This technical report reviews research that answers many common questions about the AFOQT, including whether the test is fair, whether it is biased against minorities or women, whether it is too expensive, whether it should be replaced, and whether it predicts the performance that is important to the Air Force. In addressing these questions, we do not produce original data analyses. Instead, we present a synthesis of the existing knowledge about the AFOQT and other selection tests and examine its implications for the future of the AFOQT.


This paper addresses the use of the Schmidt-Hunter meta-analysis procedure to determine
the generalizability of the Air Force Officer Qualifying Test (AFOQT) validities across Air Force specialties (AFSs). Meta-analyses were conducted on all available Academic Aptitude composite validities aggregated and disaggregated into major occupational subgroupings. The results suggest that although the validity of the Academic Aptitude composite of the AFOQT may vary across AFSs, the AFOQT is of general value in officer selection.


Crew resource management (CRM) researchers and practitioners typically take a training approach to improving crew performance. Since the late 1970s, CRM training programs have become an increasingly important part of the commercial aviation industry. Most of the major commercial airlines have either developed their own CRM training programs or adapted existing training programs to suit their own needs (see Prince, Chidester, Bowers, & Cannon-Bowers, 1992). While CRM has existed in “pockets” of the military for years, because of the multitude of commands and locations, a broad-based application has been slower to evolve. The Air Force has recently published an Air Force Instruction (AFI 36-2243) that, for the first time, establishes broad requirements for developing and managing CRM training programs, and requires CRM training for all Air Force aircrew members (Secretary of the Air Force, 1994). Research has shown that CRM training does change attitudes linked to airline incidents and accidents, including attitudes toward communication and coordination, attitudes toward command responsibility, and recognition of stress effects (e.g., Irwin, 1991). In addition, Hehreiche, Foushee, Benson, and Russini (1986) demonstrated a link between CRM attitudes and line flying performance. However, Prince et al. (1992) cite evidence which suggests that the personalities of trainees set limits on training effectiveness and attitude change. Thus, recent research in the area of CRM points to the importance of interpersonal skills and certain personality traits for effective pilot performance. Training can be viewed as a way to promote awareness of group dynamics, bring about attitude change, and improve interpersonal skills, but it does not change the underlying traits that have been shown to be related to CRM. Selection based on personality traits and/or relevant interpersonal skills could greatly enhance CRM performance, above and beyond CRM training alone.


Helme, W. H., Willemin, L. P., & Day, R. W. Psychological factors measured in the differential officer battery. Alexandria, VA: Behavioral and Systems Research Laboratory. The report is one of a series of major publications marking the culmination of the OFFICER PREDICTION research program and the impact of the findings on BESRL's ongoing and newly formulated program on officer evaluation and career development. The first publication in the series derived important dimensions of officer leadership behavior. In the present study, the major psychological factors derived from officer responses to tests of the experimental differential officer battery (DOB) are identified and delineated; the experimental predictor scores resultant from a reduction of the large number of measures obtained are described.


Helton, K. T., & Street, D. R. (1992). The Five-Factor Personality Model and Naval Aviation Candidates (pp. 16). Pensacola, FL: Naval Aerospace Medical Research Laboratory. As personality testing has improved, various models for constructing and interpreting aviation selection tests have been proposed. Of particular interest to our study is the use of the five-factor personality model to naval aviation selection test interpretation and development. The five personality factors are conscientiousness, agreeableness, openness, extraversion, and neuroticism. Therefore, we conducted a joint factor analysis on the Pilot Personality Questionnaire (PPO) and the Edwards Personal Preference Schedule (EPPS) taken by 158 Navy and Marine Corps student aviators. A principal component analysis (PCA) and a factor analysis (FA) with varimax rotations produced a robust five-factor solution. On the basis of content analysis, the FA factors obtained in our study coincided with the five classic dimensions of the five-factor personality model. Although investigations of personality in pilot selection have yielded mixed results, the finding of a five-factor solution in our study suggests that the five-factor personality model may be useful in personality testing in aviation selection decisions.


Hendrikson, I. J. M., & Elderson, A. (2001). The use of EEG in aircrew selection. *Aviation, Space, and Environmental Medicine, 72*(11), 1025-1033. The value of the electroencephalograph (EEG) as a screening device in aviation medicine is questioned, because few subjects are disqualified on grounds of an EEG exam. At the Netherlands Aeromedical Institute, pilot applicants are rejected with a diagnosis of epilepsy or
with severe EEG abnormalities (including epileptiform patterns where epilepsy is highly suspected). Although several studies have shown a low incidence of epileptiform EEG abnormalities in candidate pilots, subjects with an epileptiform EEG have a substantially increased risk of sudden incapacitation during their flying careers. In this review, we calculate the probability that a candidate with epileptiform EEG, but no history of epileptic seizures, will develop seizures during his flying career. This probability is about 25%, more than 12 times higher than for subjects with normal EEG and no history of epileptic seizures (2%). Subjects with epileptiform EEGs not only have increased risk of future epileptic seizures, but additionally it is recognized that epileptiform EEG discharges may be associated with episodic functional impairment, which can be a danger when a subject is flying. Taking this into account, one should consider rejecting all candidates with epileptiform EEGs in the future. This is at the expense of a small group of subjects with false-positive EEGs, but we believe that concern for public safety must override other considerations in these rare cases. To improve the understanding of the usefulness of the EEG in pilot screening procedures, an international classification and coding system should be developed, so that data from different countries can be compared.


From the psychologist's point of view an outstanding feature of the war has been the demand for highly specialized abilities and aptitudes in addition to general military virtues. No where has this been more true than in the Air Service and no where have the problems of selecting individuals and assigning them to the work for which they are best fitted been more urgent. As late as April 1917, the Aviation Section of the Signal Corps had but 52 trained flyers. The Air Service with upwards of 16,000 flyers is thus a new product. When war was declared the need for selecting thousands of men for training as pilots, observers and balloonists was met by the creation of Examining Boards and Physical Examining Units. Rigorous physical and medical standards were adopted. While the mental, moral and professional requirements were equally exacting the standards were necessarily vague and general. We were instructed to select men of good education and high character, men who were in every way qualified and fitted to become officers of the U. S Army — a rather intangible set of specifications. We were constantly enjoined to remember that the flying officer was not to be an "aerial chauffeur," but a "twentieth century cavalry officer mounted on Pegasus."

Henneman, R. H. (1946). Proficiency measures for fighter pilots at the operational level of training in the Army Air Forces. American Psychologist, 1, 293.


The training effectiveness literature has paid little attention to the potentially dynamic interaction of individual differences with various phases of training in determining ultimate training success. This study investigates the role of individual differences in explaining the transfer of learning from 1 phase of training to another among pilot trainees in a multistage, aviation training program. Using 3 of the Big Five factors (Conscientiousness, Emotional Stability, Openness to Experience), the results showed these measures to contribute to the prediction of the number of hours it took for trainees to attain their private pilot's license. Significant interactions between some of these measures and success on an earlier, simulator phase of the training program were also found. The results are discussed in terms of both the role of individual differences in training research as well as the broader issue of transfer of learning between phases of training.


This report conducts limited exploratory research into the possible relationship between birth order and an above average aptitude for aviation training. It reviews some of the basic elements of the science of ethology, birth order traits and possible correlations to military aviators. The study concludes that actual research into the Navy and Air Force aviator populations should be conducted to establish what, if any, relationship exists. Improved selection criteria might be established.


This chapter documents the 80-year history of military pilot selection, both theory and methodology of a variety of different countries with a focus on the U.S. The authors reviewed a great deal of research on psychomotor skill/quickness and intelligence/aptitude, but only the portions of the chapter dealing with "personality/character" are reviewed. Self-selection and rigorous physical examinations played a major role in early military pilot selection, with only the most adventurous people volunteering. During the period prior to 1919, the measurement of emotional coolness under pressure was the focus of pilot screening. Later, ingenuity and courage were added in screening for air-to-air combat, and interviews and observation were the main approaches for tapping these characteristics. From 1919 to 1938, successful combat pilots were characterized as quiet, methodical men, not given to emotional excitement, who were able to inhibit instincts of self-preservation. At this time, Germany pioneered the use graphology to form
judgments of pilot characterological suitability (Wyatt & Teuber, 1944). The U.S. Army and Navy used interviews to judge psychiatric well-being, while Germany continued to emphasize observation of facial expressions and emotional reactions to stresses. From 1938 to 1945, the U.S. Army Air Force found that commercially available paper-and-pencil personality tests did not contribute to pilot selection, while European countries "emphasized screening pilots for the proper character" (at that time a term referring primarily to motivational potential, personal presence/domination and emotional calmness under stress). Interviews were common during this period and were usually guided by an applicant's answers to a paper-and-pencil biographical inventory. A great deal of research dealing with personality was conducted during the period from 1945 to 1970. Some of this research showed that personality inventories could accurately identify psychiatrically unsuitable applicants, but did not distinguish between successful and unsuccessful aviation cadets. Sells (1956) found that motivational factors were more predictive of long-term success as a pilot than of training performance. Prior to this, some countries were already screening based on nonclinical personality traits. Both Israel and Denmark screened for nonauthoritarian, egalitarian leadership style. Many European countries were also either using or evaluating the Defense Mechanism Test, which is a psychoanalytic projective instrument designed to measure stress tolerance.


The testing procedure for applicants for operational service as fighter pilots is aimed at selecting candidates with; high theoretical capacity, high manual capacity, high mental capacity and high development potential as to competent fighter pilots. At the recruitment centre the applicants undergo a stepwise selection procedure. The first day consists of computerized group tests. If the results indicate potential the applicant is approved for a second day of individual tests. The second day ends with interviews by a psychologist and a pilot officer. At this stage the psychologist considers all input concerning the applicant and presents a synthesis to the daily selection board resulting in an eligibility rating. Those approved by the board are sent on to medical examination. Once a year a final selection board is held where psychologists, flight surgeons and pilot officers together decide on a final ranking order. There is a close connection between selection and training, and psychologists do follow-up studies on all stages of the pilot training as a way of getting continuous feedback from the system and vice versa, thus keeping updated with the progress of each individual and the demands that are placed upon them during training. The level of attrition during training is low, which is a result of successful selection where the individual is viewed from different professional angles in an overall perspective, and in combination with a training philosophy focused on supporting and not de-selecting.


Hocking, B. (2003). Training airline cadets from over 35 cultures: Some lessons learned. Paper presented at the 12th International Symposium on Aviation Psychology, Columbus, OH. Shows a syllabus hour versus module line and lines for cadets from different cultures, including those who eventually fail out. Shows that trend (slope) for those who fail is accelerating away from syllabus hour line. For those who make it, their lines lie on top of the

Meta-analyses and recent large-scale primary studies concerning the importance of intelligence indicate that for almost all jobs general mental ability (g) alone predicts performance well. However, there is a controversy concerning the question whether specific abilities (s) are needed to predict job or training performance. In the present study performance test data of 5223 applicants from the DLR program for selection of ab initio air traffic controllers at DFS Deutsche Flugsicherung GmbH were analyzed. Results of different approaches based on exploratory factor analysis and structural equation modeling to estimate g- and s-intelligence components were compared. In addition to this, the criterion-related validities of different g- and s-measures were tested using training performance criteria from 282 DFS trainees. It is argued that the preference for an intelligence model depends in part on the utilized theoretical approach and the objectives of the diagnostician (description vs. prediction).


To ensure that pilots possess the necessary level of competence for effective teamwork during line operation, some airlines have introduced special test methods into their selection procedures that allow measuring different sub-components of Social Competence before a pilot applicant is being employed. Costs and benefits of these measures vary to some degree. For a German airline, we have conducted a validation study (N=292 ab-initio pilots) with several of these measures, including the Interpersonal Competence Questionnaire (Buhrmester, Furman, Wittenberg & Reis, 1988), the Social Skills Inventory (Riggio, 1989), the Temperament Structure Scales (Maschke, 1987), and Assessment Center methods (Hoeft & Pecena, 2004). Different sub-facets of Social Competence are described, which exhibit sufficient reliability and generality to be considered as predictors for pilot selection. The findings of this study reveal significant correlations between some personality scales and aspects of Social Competence. However, correlations with concrete behavior ratings in simulated social situations are low. On the other side, Assessment Center ratings based on behavior observations correlate substantially with the overall success of a candidate throughout the selection procedure. Questionnaire data contribute little extra variance to this equation. Results are discussed with reference to aspects of social desirability as well as costs and benefits of the different approaches to measure Social Competence in pilot selection.


The purpose of this study was to examine whether physical fitness is an important component in the selection process of pilot candidates to the Israel Air Force (IAF) flight school. METHODS: There were 223 male pilot candidates who volunteered to participate in the study. All subjects were tested 1 - 12 wk prior to a week-long "bootcamp" for aerobic power (Astrand bicycle test), anaerobic power (vertical jump test), and percent body fat. In addition, an activity profile was established based on an activity history questionnaire. All fitness measures were correlated to a performance score based on the IAF selection criteria measure for each candidate. RESULTS: Candidates who were accepted to flight school had a higher aerobic capacity, anaerobic power output relative to body weight and a lower percent body fat than candidates who were not successful. Significant correlations were seen between the performance score and aerobic power ($r = 0.31$), anaerobic power ($r = 0.17$) and anaerobic power relative to body weight ($r = 0.21$). Linear regression analysis showed that aerobic power explained 9% of the variance in the performance score, while anaerobic power explained an additional 3%. The results of this study suggest that physical fitness has a positive influence on the success of pilot candidates in gaining admittance to the IAF flight school.


The invisible college of psychologists who do research with measures of normal personality now largely agrees about the structure of personality; this group also agrees that competently developed personality measures are valid predictors of real world performance. Outside that college, however, there is still considerable skepticism regarding the meaning and validity of these measures. This article attempts to summarize the data needed to answer the most frequent questions about the use of personality measures in applied contexts. Our major conclusions are that (a) well-constructed measures of normal personality are valid predictors of performance in virtually all occupations, (b) they do not result in adverse impact for job applicants from minority groups, and (c) using well-developed personality measures for preemployment screening is a way to promote social justice and increase organizational productivity.


This study examined the ability of expert clinical psychologists to predict advanced flight training success using an experimental battery of psychological tests. The tests were similar to those used by clinical psychologists, but had been adapted for group administration. The researchers used an extreme groups design. Fifty successful (e.g., passed flight training, pilot stanine of six or greater) and 50 unsuccessful (e.g., did not pass training, pilot stanine less than six) subjects were randomly selected from a much larger pool of 1504 test subjects. Subjects completed the Background Information Form (a biographical history inventory), the Ink-Blot test (group adaptation of the Rorschach), Feeling and Doing (a psychosomatic inventory), What Is He Saying (a sentence completion test), the L-D test (a group version of the Szondi test) and a group administered version of the Draw-A-Person test. Nineteen psychologists were given profiles of all subjects containing information from all of these tests. They were asked to judge whether a subject would pass or fail flight training. They made these judgments on a global level using all available information, and also make ratings of how confident they were in their judgments. A subset of these psychologists also made pass/fail judgments after examining each of the tests one at a time. Results showed that none of the psychologists were able to predict the pass/fail criterion at an accuracy level much better than chance using the global approach. Even when the researchers only examined those judgments that the psychologists were most confident in, only two were able to predict at a level better than chance. The results were very similar for the subset of psychologists who made pass/fail predictions after examining each test individually.


Reviews aviator selection assessment in several countries. In 1935, the psychological laboratory at The Imperial University of Japan in Tokyo carried out work in aviator selection of aviators, including measurement of reactions to positional change. In Budapest, tests were developed to measure intelligence, observation and reporting ability, various psychophysical attributes, and others to select officers. German testing emphasized assessments of handwriting, speech, and vocational tests.


Chapter 1 summarizes efforts conducted in relation to the first task, reviewing the existing Army Aviation accession process and relevant literature. The overall goal of this was to collect information that could be used to produce a rational decision on a specific selection strategy and produce a rational decision on a specific testing strategy (from the Statement of Work for this contract, p. 3). Chapter 2 details the job analysis undertaken to define the characteristics required for successful performance both in helicopter training and on the job. Chapter 3 describes the development and pilot testing of the predictors, or prototype instruments, and Chapter 4 describes the development of criterion measures to be used in the validation effort. Chapter 5 reports on the preliminary validation data collection and describes the results of both the predictor and the criterion data analyses. The final chapter, Chapter 6, presents the results of the validation analyses.


AR 611-110 Selection and training of Army Aviation officers (1970).


AR 611-110 (1972).

AR 611-110 Selection and Training of Army Aviation Officers (1972).


As part of the mission of the Personnel Research Division of the Air Force Human Resources Laboratory, continuing studies have been conducted to investigate ways to improve the selection procedures for admission to Undergraduate Pilot Training (UPT). This research has included attempts to improve the existing paper-and-pencil selection measures, investigations into the use of new solid-state psychomotor apparatus tests. And evaluations of learning ability through the use of a light-plane simulator. This report will outline the research that has been performed by the Personnel Research Division in these areas and will indicate some of the tasks that remain to be addressed.


Hunter, D. R. (1987). *Automated aircrew aptitude assessment: Historical perspective*. Paper presented at the 27th Annual Military Testing Association Conference, Ottawa, ON, Canada. Presentation retrieved from A recent review of the English-language literature from 1910 to the present identified over 200 studies dealing with aircrew selection (Hunter, in press). This considerable volume of research reflects the continuing concern of the military services with aircrew selection and classification. Many instruments have been evaluated over the years, with varying degrees of success.


Two measures of pilot risk perception are described. One measure assessed pilots’ perception of the level of risk experienced by other, fictional, pilots, and the second measure assessed the pilots’ perceptions of the level of risk they would experience if they were personally involved in a set of scenarios. Analyses are reported for factor scores derived from the 2 measures. Analysis of variance demonstrated significant differences in the risk ratings for the
pilot certificate groups with the more advanced certificate holders (i.e., commercial and airline transport) reporting lower levels of perceived risk. Construct validity was assessed using only private pilots (N = 369). Correlations between the factor scores and measures related to the constructs generally supported the construct validity of the risk perception measures. Inaccurate risk perception, measured as the discrepancy between the perceived risks of flying and driving, was found to be a better indicator of involvement in hazardous aviation events than any of the factor scores. It is suggested that the risk perception measures be used by other investigators to assess the contributions of these constructs to accident involvement in comparison to the contribution of other constructs.


Until World War II, the Army Air Corps counted on its stringent qualification requirements and low production goals to screen its pilot candidates. During World War II, the Army Air Forces needed men to fill its requirements for 100,000 aircrew positions, and thousands of candidates went through the training process. Qualification requirements relaxed initially and became more rigorous as the need for pilots changed during the course of the war, but no true flight screening program existed until the Korean War with the advent of the Revitalized Pilot Training Program in November 1952. Demand for more pilots and high attrition rates during the Korean War, which were prevalent during World War II as well, combined with tight defense budgets to force the Air Force to turn to some sort of flight screening to reduce attrition rates. For most of the next decade, Air Training Command (ATC)
continued to run a light plane screening program; but the introduction of the T-37 and the all-jet training program in 1958 encouraged Air Force officials to view light plane screening as counterproductive. It ended two years later. However, the war in Southeast Asia increased the demand for pilots again, and ATC reintroduced light plane screening, which continued in various forms until insurmountable problems with the T-3A prompted the end of the program in 1997. Inevitably, attrition rates rose, ensuring the return to a new program, Introductory Flight Training. By 2002, however, the hunt was on for a replacement program to provide a higher degree of standardization and uniformity. As the Air Force faces an era of stressed budgets, filling its ranks with those who will earn their wings is imperative. A flourishing flight screening program is as important today as any time in the Air Force’s history. As Air Education and Training Command embarks on yet another revision, returning to the philosophy of flight screening before flight training, it is instructive to examine how the command got to where it is today. Ultimately, concern with the monetary and personnel costs associated with high attrition rates guarantees that the Air Force will continue to use some sort of flight screening to identify pilot candidates whose probability to earn their wings is high—the very people who form the core of the Air Force’s combat capability.


The purpose of the study was to determine if student scores on the Pensacola Motion sickness Questionnaire (MSQ) would supplement current multiple prediction formula in predicting completion of flight training and/or voluntary withdrawal from this training. The MSQ score was found to be significantly related to both completion of flight training and voluntary withdrawal from training. With respect to this latter criterion, the MSQ was the single most valid predictor available. When included in the multiple prediction formulae, the MSQ significantly increased the multiple validity for predicting both criteria.


A review of the literature on pilot selection and training perceptual-motor processes, and cognitive processes was conducted. The objectives of this review were: (a) to identify perceptual-motor and cognitive tasks that demonstrated reliable individual differences in performance and (b) to identify perceptual-motor abilities and cognitive processes of demonstrated importance to successful piloting behavior. From this review, a set of tasks were identified that tapped the abilities and processes important to piloting and that showed evidence of producing reliable individual differences in performance. A conceptual framework makes
explicit the link between the tasks selected and the requirements for successful pilot performance. This resulted in a large number of tasks which were candidates for inclusion in a pilot selection task battery. Psychometric and pragmatic criteria were applied to the tasks in the candidate pool, resulting in the identification of 15 tasks for inclusion in the final task battery. These tasks span the perceptual-motor and cognitive domains with special emphasis on Attentional and decision-making performance. The paradigms for the selected tasks are specified in detail to allow for the development and implementation of the tasks on a computer system, and a number of implications for validation of the battery are provided. The tasks included in the final battery are Perceptual Speed, Complex Coordination, Compensatory Tracking, Kinesthetic Sensitivity, Route Walking, Selective Attention, Time Sharing, Encoding Speed, Mental Rotation, Item Recognition, immediate/delayed Memory, Decision Making Speed, Probability Estimation, Risk Taking, and Embedded Figures.


The Canadian Armed Forces' aircrew production system is similar to that of other military forces, in that it involves interrelated programmes of recruitment, selection, classification, and training. The object of the system is to produce aircrew pilots and navigators who meet specific performance standards and, moreover, to produce these highly-skilled personnel in specified numbers at minimal unit cost. The importance of economic factors in such a production system are emphasized by the fact that, as an individual progresses through the system from the recruiting level to the operational training units, the total monetary investment in a trainee increases cumulatively and the loss of a single trainee becomes increasingly important as he nears graduation. At the present time, aircrew officers are probably the most costly product which the Canadian Armed Forces are required to manufacture within their own resources.


The Army Research Institute Aviation Research and Development Activity (ARIARDA) successfully implemented the Multi-Track Test Battery and associated classification functions in 1988. The battery and functions have been used to assign more than 4,000 flight students to their combat skills aircraft. The subsequent program determined the applicability of the battery to prediction of student performance in flight training. Performance evaluation in primary training consists of four-flight phase grades and 12 academic-phase grades. In addition to these, primary overall average grade and primary overall flight grade were predicted using forward stepwise multiple-regression procedures. Stepwise multiple-discriminant analysis was used to investigate two additional measures--flight deficiency training setback and flight deficiency attrition. The
capability of the battery to predict primary training grades is demonstrated. Results of discriminant analysis of setbacks and attrition should be viewed with caution.


The Army Research Institute Aviation Research and Development Activity successfully implemented the Multi-Track Test Battery (MTTB) and associated classification functions in 1988. The battery and functions have been used to assign flight students to their combat skills aircraft. The present program determined the applicability of the battery to prediction of student performance in flight training. Performance evaluation in training consists of flight phase grades and academic phase grades. In addition to these grades, Overall Average Grade and Overall Flight Grade were also predicted using Forward Stepwise Multiple Regression procedures. Stepwise Multiple Discriminant Analysis was used to investigate two additional measures, flight deficiency training setback and flight deficiency attrition.


Initial Entry Rotary Wing Multi-Track (IERW-MT) flight training was initiated by the U.S. Army Aviation Center (USAAVNC) in 1988. Under this program flight students are assigned to complete their training in one of four rotary wing aircraft types by the eighty fifth day of flight school. The Army Research Institute Aviation Research and Development Activity (ARIARDA) developed a method to optimize aviator candidate classification for best probability of success. This method was implemented by USAAVNC in May, 1988. In this initial validation study 686 IERW-MT graduates, assigned to training tracks by the ARIARDA developed procedures, provided data to determine if the discriminant functions to be used would accurately match students with their actual assigned aircraft, and to determine if the functions could be enhanced by including other available performance data. This effort also explores the potential of the classification test battery for prediction of flight deficiency training setbacks, track flight performance, and common core training performance. The results strongly support the utility of the classification procedure for track assignment and indicate its applicability to performance prediction.


In late 1986, the U.S. Army Aviation Center (USAAVNC) redesigned the Initial Entry Rotary Wing (IERW) course for aviator candidates. The new training is called IERW Multi-
Track (IERW-MT) and became operative in May 1988. The research problem for the U.S. Army Research Institute Aviation R&D Activity (ARIARDA) was to develop tests and procedures for selecting aviator candidates for one of four helicopters prior to training day 100. ARIARDA simultaneously pursued two avenues of research. On the one hand, available test instruments were considered and evaluated for their potential to discriminate among aviators. On the other hand, groups of Subject Matter Experts (SMEs) developed criticality-rated aviator candidate abilities and traits for specific operational helicopters. Extensive literature reviews and liaison with sister services and other agencies were accomplished. Four test instruments were evaluated for use. The underlying abilities, traits, and skills these batteries purported to measure matched the abilities, traits, and skills identified as necessary by the SMEs for each of their helicopters. Upon selection of the subtests contained in the ARIARDA experimental test battery, high-time aviators were given the experimental battery to develop scoring profiles for specific aircraft and to generate the data for the statistical analyses that resulted in the Preliminary Multi-Track Classification Algorithm.


In 1986, the U.S. Army Aviation Center (USAAVNC) redesigned the Initial Entry Rotary Wing (IERW) course of flight training. The new training course, IERW Multi-Track (IERW-MT), replaced the primary helicopter trainer, TH-55, with the UH-1. At training day (TD) 101, students would continue training by being tracked into one of four operational helicopters: UH-1, AH-1, OH-58, or UH-60. In 1987, the Army Research Institute Aviation R&D Activity (ARIARDA) was tasked to develop a process for assigning students to one of the four helicopters. The process required matching the skills, abilities and traits of individual students to the skills, abilities and traits necessary to successfully complete flight training in a particular helicopter. The assignment process also had to provide the highest potential for a successful aviation career in operational units. The assignment process requires application prior to TD 100, normally at about TD 85. Since only 18 months were available to conduct the research, ARIARDA simultaneously pursued two avenues of research. On the one hand, available tests were evaluated for their potential to discriminate among experienced aviators as well as students. On the other, highly experienced groups of aviators, using Small Group Analysis (SGA) techniques, identified the critically rated aviator skills, abilities and traits necessary for flying each helicopter.


Refractive surgery to visually rehabilitate refractive errors of the eye continues to evolve at a significant pace and is here to stay. The surgical manipulation of the cornea by carefully planned incisions, as in radial keratotomy, represented the first procedure to evolve for the correction of ametropia and is an area of continued active development and improvement. However, many concerns mitigate against this procedure in the aeromedical arena. More recently, photorefractive keratectomy using laser technology to ablate and recontour the corneal surface has emerged as a viable modality. This paper explores the aeromedical factors surrounding this new revolutionary procedure and discusses the issues relevant to evaluating its applicability to the modern military aviator as well as reviewing results of the latest clinical trials currently in progress. The goal is to provide the aeromedical community with the fundamental information required to formulate aeromedical decisions and policy-making in regard to a new procedure that is certain to have tremendous impact on the selection of future aircrew candidates.


The GoldenEye-50, developed by Aurora Flight Sciences, was selected by the Defense Advanced Research Projects Agency as one of multiple candidates to provide the basic platform for the Organic Air Vehicle II program for expected integration into the U.S. Army’s Future Combat System program. The GoldenEye-50 is a transportable (approximately 18 lb) unmanned aerial vehicle (UAV) with vertical take-off and landing capability. It is designed to carry a payload to support reconnaissance and chemical detection missions and can transform from a hover-and-stare mode to wing-born flight as needed. In support of the Human-Robot Interaction Army Technology Objective, an assessment of the operator control unit (OCU) was conducted during a technical demonstration of the GoldenEye-50 held at Fort Knox, Kentucky, from 9 to 12 May 2005. The authors’ primary objective was to reveal human factors engineering issues associated with the design of the OCU interface and to learn the tasks of a UAV operator, particularly with multi-mode flight capability of vertical and horizontal flight. From the observational data, potential issues and recommendations for the final OCU interface design accepted by the Army are presented.


Presented at the 27th Annual Military Testing Association Conference, Ottawa, ON, Canada.

Given the increasing complexity of modern military aircraft and the high costs associated with pilot training, the Canadian Forces (CF) has been examining methods of improving the assessment and selection of pilot candidates. Consistent with this objective, the CF has developed and is evaluating the Canadian Automated Pilot Selection System (CAPSS), an automated computerized test of complex psychomotor coordination based on flight simulation technology (James, 1985a; 1985b; 1985c). This paper describes CAPSS and outlines plans for its validation.


Perhaps the best key to psychological problems in aviation is offered by the opening sentence in a foreign flight-manual which reads: "It must be remembered that flight is an unnatural activity in man." True when it was written, it is more than ever true today. Today one must help to select and maintain men who will operate machines in three dimensions, machines that will climb too fast, accelerate and decelerate too rapidly, rise too high, operate under great extremes of temperature, fly when men cannot see, and cover an enormous amount of space during the period of a simple reaction-time. In wartime this is further complicated by the fact that enemies on the ground, on the sea, and in the air are quite likely to make lethal gestures during any of the maneuvers named.


This paper reviews issues relating to the psychological testing of aircrew. The paper aims to clarify issues relating to the validity, reliability, and value of the psychological testing of aircrew. It is first suggested that the entire domain is characterized by terminological and methodological confusion. The economic and other benefits of psychological testing are contrasted with the potential risks, including abuse and the application of tests in circumstances for which they were never designed. Reference is also made to European cultural differences which may potentially impact the practical realities of psychological testing. (Author)


Despite his passive-sounding name the Royal Navy Observer is a very active and important member of a helicopter crew. He acts as a navigator and tactical coordinator in 2- and 4-man helicopters. He operates variable depth sonar, radar, electronic warfare equipment, and weapons systems such as homing torpedoes, depth charges, and air to sea missiles. The Royal Navy has met various problems in recruiting suitable personnel, that is, in providing this important "man in the back seat." The majority of Observers join the Service on a Short or Medium Career engagement, and only a relatively small number come from amongst Full Career Officers. The Short/Medium Career Commission for aircrew in fact provides the majority of both Pilots and Observers in the Royal Navy. For Short/Medium Career Officers the training pattern until recently has been 8 months basic Officer training, twenty four weeks Basic Flying Training (100 flying hours in a fixed wing aircraft), Advanced Flying Training (27 flying hours in a helicopter), followed by 50 flying hours of Operational Flying Training. Over a seven year period Short/Medium Commission Officers have shown a 28% wastage rate during Officer training, of which half was voluntary. During Basic Flying Training (BFT) there was 29% wastage (40% of those entering), and 7% (16% of those entering) at Advanced Flying Training (AFT). The overall wastage has therefore been approximately two-thirds of those recruited. From the above-figures two major problem areas can easily be identified: voluntary wastage in Officer training and wastage in Basic Flying Training. Failure at this latter stage is predominantly (90%) ascribed to problems in the air (as distinct from ground school). Often trainees are described as "leaving some of their brains behind on the ground."


The increasing specialization of rotary wing missions and aircraft has precipitated a reanalysis of traditional strategies for assigning student aviators to one of four rotary wing missions: cargo, utility, aeroscout, or attack. Although previous research has suggested that certain abilities are appropriate for inclusion in a classification algorithm, there are no data to indicate that there are differences in the ability requirements (types or levels) for the four missions. This paper describes the results of several analyses designed to compare the ability
requirements of the four missions. Ability rating data, obtained from subject matter experts for each mission, were transformed using the Method of Successive Intervals (MSI) to remove systematic biases identified in the raters' distributions. Analyses of the transformed data indicate that there are no differences in the types or levels of abilities required to perform the most demanding tasks for each mission.


Applied research in psychology not only has contributed directly to societal advances but often has fostered basic research as well. Prominent examples are the programs directed by Yerkes in World War I to develop the Army Alpha test and several programs in World War II, including The American Soldier that assessed soldiers' attitudes during the war; a program for selecting agents for the Office of Strategic Services; and the Aviation Psychology Program to select and classify applicants for flight training in the Army Air Forces. Highlights of these programs are presented here, with special attention given to by far the largest, the Air Forces program. After World War II, many of the hundreds of psychologist veterans became prominent research psychologists. Most became university professors. Among those who continued to work in applied settings was John Flanagan who had served as Chief of the Army Air Forces Psychology Branch. (After the war, Saul Sells succeeded Flanagan as Head of the Aviation Psychology program.) Flanagan founded the American Institutes for Research (AIR), the earliest mission of which was to select flight personnel for civilian airlines. Another part of the AIR mission was to enhance civilian air safety by assuring the widespread use of Flanagan's critical-incident procedures (or near-accident reports) that now serve to reduce accident rates in a variety of industries as well as in aviation.


Kalagian, S. P. (1964). *Personnel aspects of the Army aviation program*. Paper presented at the 10th Army Human Factors Research and Development Conference, Fort Rucker, AL. From a humble beginning in 1942 when light single engine civilian Cubs were pressed into service for reconnaissance and observation missions, Army aviation has experienced a
continuous growth toward a current requirement based on organizational structures in excess of aviators (1). These requirements do not include the airmobile test activities now in being at Fort Benning and Fort Riley nor the unprogrammed requirements that have been generated to support counter-insurgency operations in the major trouble-spots of the world.


The Air Force Human Resources Laboratory (AFHRL) is conducting a multi-year research and development (R&D) program to improve procedures for selecting candidates for USAF Undergraduate Pilot Training (UPT). The principal goal of selection procedures for UPT is to screen out those candidates with low chances of completing training and/or becoming successful operational pilots. This is more important now than ever because the costs of each UPT eliminee have risen dramatically (approximately $64,000) and the operational mission in which USAF aircrews are employed has increased in complexity and difficulty. Therefore, improvements in identifying candidates who have the requisite abilities for success will both reduce the wasted costs associated with eliminations and help ensure that operational—pilots are capable of meeting the rigorous demands of today's military aerospace environment.


A multi-year research program to improve the selection of pilot trainees and the classification of student pilots for either fighter or heavy aircraft training is described. A battery of experimental tests measuring psychomotor skills, personality traits, and cognitive abilities is being given via computer prior to training. The subjects' performance in training and operational flying is then tracked and analyzed. The preliminary results and future directions of this ongoing program are discussed.


Beginning in 1955, the Deputy Chief of Staff for Personnel established requirements for the development of instrument to select officers as fixed-wing pilot trainees and enlisted men as arrant officer candidate rotary-wing pilot trainees. In 1963, the requirement was expanded to
provide for a consolidation of the separate program. Procedure: To meet the initial requirement, research programs were conducted involving the experimental testing of 2000 enlisted men, 1200 officers, and 1200 ROTC cadets. Particular attention was given to the development of measures to select enlisted personnel for rotary-wing training, including preflight (OCS) to prepare graduates for warrant officer commissioning. Pending completion of a long-range research effort, partial results were utilized to develop interim test batteries for operational use. Finally, current operational data and previous research findings were combined to provide a basis for a comprehensive selection program. Selection tests initially developed by the Air Force and modified for Army use were effective in predicting fixed-wing training success for officers and ROTC cadets. Selection tests developed by the U. S. Army Personnel Research Office were effective in predicting the success of enlisted applicants for warrant officer candidate preflight and rotary-wing training. A comprehensive set of Flight Aptitude Selection Tests (FAST) was developed which provides effective measurement of both fixed-wing and rotary-wing aptitude for applicants to warrant officer candidate aviation training and of both fixed-wing and rotary-wing aptitude for applicants to officer aviation training.


The Army ROTC Flight Instruction Program was authorized by regulation in 1956 in order to provide basic ground and in-flight fundamentals to meet minimum requirements of the Federal Aviation Agency (FAA) and to qualify students for FAA private pilot certification. The objective was to create a reserve pool of qualified pilots who can be utilized in the event of a national emergency. ROTC flight training may further serve as a selection device and as useful preparation for the Active Army's flight training programs. The Army Fixed-Wing Aptitude Battery, AFWAB-1, was administered experimentally to samples of students applying for ROTC flight instruction during the years 1956-57, 1957-58, and 1958-59. The battery was then evaluated for effectiveness in discriminating between successful and unsuccessful trainees. On the basis of this research, AFWAB-1 was adopted for ROTC use in 1961.


Disqualifying criteria for aircrew in Europe (JAR-FCL 3) are, besides a diagnosis of epilepsy after the age of 5 and a history of episode(s) of disturbance of consciousness, epileptiform paroxysmal electroencephalographic abnormalities and focal slow waves. Intermittent photic stimulation (IPS) provokes in about 0.5% of healthy subjects (range 0-2%) a photoparoxysmal response and is most often the only abnormality (70-90%). The literature is scarce and shows great diversity in methodology. Standardized IPS with simultaneous video will not only allow collection of sufficient data for proper epidemiological studies, but can also reveal
clinical and often unnoticed or misinterpreted signs and symptoms like myoclonia, loss of
consciousness, and occipital seizures with visual auras. The pilot (sleep deprivation, strong
sunlight) and the traffic controller (stress, monitors) are more prone to visually induced seizures.
Furthermore, the increasing exposure to potentially seizure-triggering visual stimuli might have
its impact in a more indirect or cumulative way.

discharges on complex tasks and cognition: relevance for aircrew and air traffic
controllers. Epilepsy & Behavior, 6(1), 31-34.

Subtle seizures consisting of brief alteration of consciousness with or without
automatisms may go unnoticed in daily life, but can be detected more easily with
electroencephalographic (EEG)/video recordings. Generalized and partial epileptiform EEG
discharges can nevertheless be subclinical (subclinical epileptiform discharges, SEDs). When
appropriate complex tasks are presented, it has been shown that even very short SEDs of 0.5
second disrupt cognition. In daily life this has been shown during automobile driving: half of the
subjects showed significant deviations in lateral position of the car during SEDs and made more
errors in an attention task while driving. Individual differences in the cognitive effects of SEDs
are, however, striking and may be partly due to interaction between level of performance and
frequency of spontaneous EEG discharges, as has been shown in another driving study: about
75% of subjects showed suppression of SEDs by driving, which is a combination of sensory,
mental, and motor activity. Not only can SEDs negatively influence performance, but in some
cases mental activities can provoke epileptiform discharges. It is important to realize that these
mechanisms exist and that only detailed EEG studies can clarify these issues. In air traffic
controllers, brief alterations of consciousness and cognitive impairment have occurred but cannot
be accepted for safety reasons; therefore, Eurocontrol has used the EEG as a screening tool since
1995.

Department of Defense Human Factors Engineering Technical Advisory Group
the Department of Defense Human Factors Technical Advisory Group 52nd Meeting.

Medicine, 14(6), 383-385.

trend. Maxwell Air Force Base, AL: Air Command and Staff College, Air University.

psychomotor and psychosocial skill demands of UCAV operators. Paper presented at the
Second Annual Conference: Unmanned Combat Aerial Vehicles II (UCAV's), London.

Cognitive, Psychomotor, and Psychosocial Skill Demands of Uninhabited Combat Aerial


The opportunity to appear before this body suggested making a report upon some one of the phases of army work with which I have been connected, or of presenting certain conclusions and principles which have been developed or strengthened by my total War Department experience I have chosen the latter course because I believe my report will be of more value if I present some of the principles of selection underlying war procedure which are equally applicable to peace conditions. My experience in connection with tests for illiterates, trade tests, and the selection of ground and flying school men, entirely supports the conclusions which will be reported, but I shall choose another field, that of selection of officer material from S. A T C. units, to illustrate the points made, because, in my judgment, the principles adopted in the selection and classification of S. A. T. C. men are in advance of those used in any other field of classification. The problem of classification in the Students Army Training Corps followed directly from its main purpose which was to provide for the securing of material fit for officer training.


Kennedy, R. S. (1975). Motion sickness questionnaire and field independence scores as predictors of success in naval aviation training. Aviation, Space, and Environmental Medicine, 46(11), 1349-1352.

The present report has shown that a motion sickness questionnaire can be used to predict susceptibility to motion sickness or flight training success, depending on the items scored. There is a discussion of the theory that motion sickness results from conflicting perceptual inputs. This theory is related to aircraft operating conditions. Scores on a personality test which appear to be
related to similar perceptual phenomena are related to aviation success. One phenotype, field independence, seems to be promising in this regard. In addition to use of this finding in aviator selection, it is felt that studies of this trait, as it relates to an ability to reconcile conflict and to motion sickness insusceptibility, should be conducted.


We stated in our Phase I proposal that it was "...our intention to be fully responsive to this solicitation and to create '...a new test battery (and new tests)...' by focusing in Phase I on ability related aptitudes to determine whether these computerized tests add new variance to the existing tests, as well they might." In Phase I we tested an automated battery which combined tests from three different computerized test domains with which we had extensive experience: 1) cognitive tests; 2) temporal acuity (visual temporal information processing) tests; and 3) psychomotor tests including video games. These tests were all: written in "C", validated against simulated aviation and space shuttle landing tasks, could each be fielded immediately so that plans for networking and field testing could incorporate actual lessons learned from empirical evidence both about content of behavioral measures, and also about the difficulties to be encountered in attempting to test at remote sites. Data were collected on 120 aviation candidates. We examined the psychometric properties of measures from the automated battery, and related those measures to tests and composites from the Aviation Selection Test Battery (ASTB). Analyses indicated that: a) Tests in the new automated battery were generally of adequate to high reliability. b) Correlations between the automated battery and ASTB variables were, with a few exceptions, not large but in the expected directions, suggesting that tests in the automated battery do not duplicate the variance measured by the ASTB, and could thus potentially augment the predictive power of the ASTB. c) Relationships among the Spatial Apperception Test (SAT) and several of the temporal acuity tests indicate that temporal factors may be involved in SAT performance. At the same time, there is considerable reliable variance unique to each domain, again implying the potential for augmentation of the ASTB by measures from the automated battery. Serendipitously, because of a computer malfunction, data on the automated battery were collected on two different computers. In the course of analysis, it was determined that data from the two machines differed on several of the tests in the automated battery. The implications of these differences for future plans networked, distributed selection testing were discussed, and it was suggested that possible effects on testing outcomes resulting from computer characteristics should be given additional emphasis, both in the Phase II effort which would be proposed by Essex and in the Navy's current plans for implementation of networked selection.


A study was conducted to determine whether a number of computerized tests would add new variance to a paper-and-pencil selection battery, the U.S. Navy's Aviation Selection Test Battery (ASTB). The computerized tests came from three different domains: 1) cognitive tests; 2) temporal acuity (visual temporal information processing) tests; and 3) psychomotor tests including video games. These tests were all written in "C" and had been validated against simulated aviation and space shuttle landing tasks. Data were collected on 120 aviation candidates. Analyses indicated that: a) Tests in the new automated battery were generally of
adequate to high reliability; b) Correlations between the automated battery and ASTB variables were, with a few exceptions, not large but in the expected directions, suggesting that tests in the automated battery do not duplicate the variance measured by the ASTB, and could thus potentially augment the predictive power of the ASTB; and c) Relationships among the Spatial Apperception Test (SAT) of the ATSB and several of the temporal acuity tests indicate that temporal factors may be involved in SAT performance. At the same time, there is considerable reliable variance unique to each domain, again implying the potential for augmentation of the ASTB by measures from the automated battery.


A critical task for most organizations is the ability to select qualified employees for vacant positions. Consequently, much effort has been spent developing tests that predict future job performance. One of the more popular selection methods is the biodata questionnaire. Biodata questionnaires are self-report multiple-choice instruments that measure prior behavior and reactions to life events relevant to job performance. Me&analyses of the selection literature by Dunnette (1972), Hunter & Hunter (1984), Schmitt, Gooding, Noe, & Kirsch (1984), and Reilly & Chao (1982) reveal that biodata effectively predicts a wide variety of performance criteria (e.g., ratings of overall performance and advancement potential, commendations, sales volume, bonuses), yielding validities in the .39s to .40s.

King, J. E. (1946). *The modification-revision method in psychomotor measurement, a minor study in aviation psychology*. Chicago, IL.


The Minnesota Multiphasic Personality Inventory (MMPI), long a psychometric staple, has not been readily compatible with the third Diagnostic and Statistical Manual of Mental Disorders (DSM-III and DSM-III-R; 1, 2). Morey, Waughr and Blashfield (8) rationally! empirically constructed MMPI personality disorder scales to assess these DSM-III Axis II conditions, but adequate outpatient validation remains to be accomplished. The present study, based on 104 male aviators referred to a consultation service, found significant (p < 0.001) positive correlations in nine of 11 personality disorder scale comparisons between the MMPI and the Millon Clinical Multiaxial Inventory (MCMI), a test that attempts to more closely correspond to the criterion of the DSM-III/DSM-III-R. The Antisocial and Compulsive personality disorder scales failed to significantly correlate. The MMPI personality disorder scales; however, did not significantly identify which subjects had psychiatrically noted maladaptive personality traits.


The US Army Research Laboratory is developing the Job Assessment Software System (JASS). JASS is a computer program to define and measure human aptitudes required to do a job. A person familiar with the job (designer, worker, trainee, etc.) uses JASS by answering "yes" or "no" to a series of questions. The answers identify the aptitudes required. For each required aptitude, JASS presents a 7-point scale to rate the aptitude level. JASS is based on the work of Dr. Edwin Fleishman (Fleishman & Quaintance, 1984) of George Mason University. This paper will describe JASS and its recent applications and will also describe current and anticipated efforts to further develop and improve the JASS tool.


The Computerized Adaptive Screening Test (CAST) is used by Army recruiters to predict
prospective applicants' subsequent performance on the Armed Forces Qualification Test (AFQT). A modified version of the CAST software was used in 60 recruiting stations across the country from January through December 1985 to collect CAST item-level performance information. Screening test data were matched to applicant records from Military Entrance Processing Stations to obtain ASVAB scores and relevant demographic information. The cross-validated, corrected correlation between CAST and AFQT scores is .83. CAST's ability to predict important AFQT performance categories and Army Aptitude Area scores was also examined. Alternative subtest lengths were evaluated and item bank characteristics were described.


In research on bomber crew training and proficiency, an effort has been made to develop economical, reliable, and valid measures of attitudes exhibited in various situations, and to determine the relationship of such attitudes to crew proficiency. The present study as made to assess the stability of crew attitudes from the training to the combat situation. The study was designed also to provide additional information concerning: (a) relationships found in an analysis of 1952 combat crew attitudes and crew performance, and (b) the combat attitude scales used in that study.

Early feedback from the operation of Uninhabited Air Vehicles (UAVs) indicates that improvements in the operator interface aspects of these emerging systems would reap significant gains in system performance and effectiveness. This applies to both effective control of UAVs as well as management of data and dissemination of the associated information. The Canadian Forces (CF) is pursuing the introduction of UAVs, and while such platforms may provide an enormous amount of data, the management of data to support effective human decision making is still an issue. Various levels of automation have been suggested as a way of addressing the problem including Intelligent/Adaptive Interfaces (IAIs) for decision support. IAIs are intended to manage information dynamically and provide the right data and information to the right people, at the right time, to support effective decision making. The work reported in this paper investigated the efficacy of IAIs in an operational situation. The selected environment involved UAV operations in support of counter terrorist activities with the IAI modelled as part of the UAV tactical workstations of a modernized CP140 aircraft. In order to produce an analysis of UAV operations which are relevant to CF UAV implementation plans, a one hour mission scenario was developed which reflected a portion of the upcoming Canadian Forces Experimentation Centre (CFEC) Atlantic Littoral Intelligence, Surveillance and Reconnaissance Experiment (ALIX) program. In order to facilitate the development of a performance model implemented in an Integrated Performance Modelling Environment, a Hierarchical Goal Analysis and Operational Sequence Diagrams were prepared for the scenario. The model was run in two modes: one assuming the operators used a conventional interface and the second assuming interface automation using an IAI. The difference between mission activities without automation (IAI OFF) and with automation (IAI ON) was reflected in the time to complete critical task sequences and in other measures of performance. It was concluded that the operational control of UAVs from an airborne platform was a complex task and workload increased during use of multiple disparate UAV assets. The use of a workstation, which incorporates an IAI mode, permitted operators to continue working under high time pressure, resulting in upper level goals being achieved in reduced time.


An objective of this paper is to provide a literature review economic and non-economic reasons for attrition from military and civilian jobs, and to develop and calibrate a multivariate model of attrition. The stipulated model relates attrition from first job to civilian and military pay, educational levels of servicemen, socioeconomic status of parents, job satisfaction, age, race and perceived locus of control of an individualistic serviceman. The study used data from the National Longitudinal surveys of youth from 1966 to 1979. It concludes that attrition rates in the military can be reduced by increasing military wages, particularly in the lower rungs of the pay ladder, matching job requirements with education levels and increasing job satisfaction in the Military Occupational Specialties of the separtees.


Ninety multi-engine instrument rated pilots were assigned to no motion, sustained linear scaled down analog motion, and washout motion in a GAT II simulator for determining the effects of degree of motion upon the predictive validity of flight simulators. Five instrument and five contact maneuvers were flown in the simulator followed by flight in a Piper Aztec aircraft. Performances were recorded by two observers and the interobserver reliability coefficients were 0.962 and 0.919 for instrument maneuvers and 0.879 and 0.613 for contact maneuvers in the simulator and aircraft, respectively. The condition of no motion resulted in greater error than the other two groups in the simulator, but there were no significant differences in the aircraft. Correlations of aircraft performance from the simulator maneuvers were 0.763 (no motion), 0.911 (sustained motion), and 0.651 (washout motion). Simulator motion did not result in better aircraft performance, and higher predictive validity was found with very basic sustained motion.

In this article, the authors propose adding the aircrew psychomotor test (APT) to the Air Force Officer Qualifying Test (AFOQT; Form N) in an effort to reduce the number of cadets who fail to complete flight training. The APT consists of two psychomotor tasks. For the first task, subjects are required to track a target using two hand controls. The second task introduces the use of foot pedals. The subjects are required to use the hand controls and the foot pedals to keep a cross centered on the crosshairs that split the screen in half. Approximately 200 cadets from the Air Force Academy's classes of 1978 and 1979 were administered the AFOQT and the APT. They were ranked according to their scores and all who failed to score above the cut-off score were identified. All subjects were allowed to participate in flight training. After completion of flight training, their training performance was compared with their scores on the AFOQT and the APT. Using the APT as the screening device, 9.3 percent of the pilot-qualified cadets in the class of 1978 would have been denied access to pilot training, but 83 percent of these cadets would have succeeded; 6.0 percent of the class of 1979 would have been denied training, but 91.6 percent of these cadets would have succeeded. The correlations between the APT scores and UPT pass/fail were -.04 (class of '78) and -.09 (class of '79). Results of the analyses using the AFOQT were approximately the same as those obtained using the APT. Thus, neither the AFOQT or the APT discriminated very well between those who passed flight training and those who did not.


This paper presents an overview of the field of aviation psychology from its pre-World War I beginnings to current problems and opportunities for today's aviation psychologists. This brief survey covers some of the more significant early activities of aviation psychologists in the United States and Europe.


One hundred and fifty persons answered a questionnaire measuring self-esteem, propensity for feedback, self-efficacy, and certain demographic information. Subsequently, the students completed the Basic Flight Instruction Tutoring System (BFITS), a series of fully-automated criterion-referenced lessons designed to teach a person how to fly an airplane. BFITS provided feedback whenever student performance on monitored variables approached the limits of acceptable performance. After BFITS the students entered the traditional flight training program. Flight time prior to the first solo flight, landings before first solo, and total time to the private pilot certificate were obtained. Individual needs for feedback were significant factors of performance in the BFITS training and its transfer to the aircraft. Because of the interaction between individual propensity for feedback and training performance an individually adaptive feedback methodology is proposed.


Approximately equal number of male (45) and female (43) Air Force Academy cadets learned a complex aerial maneuver (chandelle) on a desk-top flight simulator. These cadets had participated in a previous experiment (Koonce & Berry, 1980) where they were given a battery of tests which tapped several cognitive and perceptual-motor areas. Of special interest in the
The present study was the relationship of the various cognitive factors to the rate of acquisition of the complex aerial maneuver. Prior to the introduction of the complex maneuver all subjects were trained to criterion level on four basic instrument flight maneuvers (Koonce & McCloy, 1980). Then they learned how to fly the complex maneuver with the number of trials required to reach criterion performance as the dependent variable. Results indicated cognitive factors were very significant in predicting complex maneuver performance. Additionally, individually tailoring the regression equations by sex as opposed to utilizing a general overall regression equation greatly enhanced predictive capability.


The paper deals with individual differences of fighter pilots and flying control officers according to tension of central regulation mechanism during an experimental test simulating an operating activity. It shows and describes the human genotype dependent quality of the central nervous system structure to process information flows of any intensity and density. It develops a selection procedure for flying control officers working in dynamic information flow conditions according to individual human central nervous system ability to process information and resist information stress.


The authors note that a wide range of factors influence the number of flight training hours necessary for obtaining a pilot's certificate. Sometimes a student reaches a learning plateau, and it is necessary for that student to change instructors before he or she is able to complete their training. The purpose of this study was to explore whether a significant relationship exists between differences in personality characteristics of instructors and those of their students and the number of flying hours required to complete flight training. The Myers Briggs Type Indicator (MBTI) was administered to 22 male and 10 female students and their instructors after completing flight training. Correlations were then computed between the difference scores of students and instructors on each of the MBTI scales and the number of flying hours required to complete training. No significant correlations were found for the Sensing/Intuition, Thinking/Seeing and the Judgment/Perception scales for either men or women. The correlation between number of flight hours and the difference score on the Extrovert/Introvert scale was .37 a < .05) for men. A moderate correlation was found for women between flying hours and the difference score on the Sensing/Intuition scale, but this correlation was non-significant (mostly likely due to a lack of power).


This report describes the job analysis performed by The U.S. Army Research Institute for the Behavioral and Social Sciences Rotary Wing Aviation Research Unit (ARI RWARU). It was part of a larger research project to develop and validate a selection system for U.S. Army rotary wing aviators, called Selection Instrument for Flight Training (SIFT). The activities performed by Army aviators and the personal attributes required to perform those activities were examined. This job analysis helped identify predictor measures subsequently used to validate the prototype SIFT test battery.


This study examined personality differences between student naval aviators, student naval flight officers and a normative college sample using the Tri-Dimensional Personality Questionnaire (TPQ). The TPQ is a 100 item true-false questionnaire designed to assess several adaptive stimulus-response characteristics of personality proposed by an integrated biosocial theory of personality. It measures three dimensions: Novelty Seeking (activation), Harm Avoidance (inhibition) and Reward Dependence (maintenance) [these dimensions correspond to the dopamine, serotonin and norepinephrine neural systems, respectively]. The authors hypothesized that personality characteristics, psychomotor skills and cognitive abilities play a significant role in the selection of aviation as a career goal. Thus, there should be both
similarities and differences in personality characteristics between pilots and flight officers. One hundred twenty-nine aviation candidates from the U.S. Navy flight training program were used as subjects. Seventy-nine were student naval aviators and 50 were student naval flight officers. Three hundred twenty-six college students comprised the normative group. A one-way analysis of variance indicated that there were significant differences between the groups in Novelty Seeking and Harm Avoidance. Specifically, pilot candidates scored higher in Novelty Seeking than the normative group, and they scored lower than flight officer candidates and the normative sample in Harm Avoidance. The authors conclude that student naval aviators and student naval flight officers are likely to pursue novel and unfamiliar experiences and to appear calm and uninhibited under normal circumstances. They characterize the results of their study as being a "snapshot" taken at that point of time between finishing basic officer training and the beginning of flight training.


The Tridimensional Personality Questionnaire and the Hand Test were administered to 129 naval aviation student (79 student naval aviators and 50 student naval flight officer) candidates while they were waiting to begin basic flight training. Results indicated that, although pilot and flight officer candidate scores departed significantly in the same direction from normative data, there were few significant differences between the 2 groups on personality dimensions. However, both groups were significantly different from normative data on some dimensions. Possible interpretations of the data are discussed.


This study evaluated a behavioral-based personality instrument currently used by a branch of the United States Armed Forces in their experimental aviation selection battery. The instrument, the Dot Estimation Task (DOT), was designed to measure compulsiveness versus decisiveness. The reliability and validity of this experimental instrument was evaluated. One hundred and fifty-three university undergraduate students were administered the DOT and either of two paper-and-pencil compulsivity instruments in a counterbalanced design. Four weeks later, 90 subjects were retested on the DOT and the alternate compulsivity instrument. The results indicate that the DOT has a test-retest reliability of .64 but has no relationship to either compulsivity measure.


The Basic Attributes Test (BAT) is a multiple aptitude computer-based battery designed to measure individual differences in psychomotor coordination, cognitive abilities, personality, and attitudes. The Air Force plans to operationally implement the BAT as a pilot candidate selection instrument in the near future. Scores from the Air Force Officer Qualifying Test, BAT, and biographical information will be combined in a new pilot candidate selection composite to
predict undergraduate pilot training outcomes. Although much useful research has been done in
the BAT battery, the need for additional psychometric research to improve test scoring
procedures and predictive efficiency was identified. The purpose of this study was to investigate
(a) the internal consistency of item-level test scores, (b) the effects of alternative scoring
procedures (e.g., treatment of outliers, data transformations, alternate scoring algorithms) on
internal consistency and validity, and (c) the factor structure of the BAT. Results showed that (a)
internal consistencies of most BAT scores are acceptable, indicating that the constructs are being
measured reliably, (b) neither censoring outlying data points nor transforming data had a
significant impact on internal consistency or validity of BAT scores, (c) few alternative scoring
procedures improved BAT score validity, (d) test scores relate to a meaningful factor structure,
and (e) BAT scores can be combined into an efficient model for the prediction of undergraduate
pilot training performance.

and psychomotor test data. Military Psychology, 8(1), 43-58.

Authors of many statistical texts and review articles have pointed to the possible adverse
effects that outliers can have on the calculation of sample statistics and have suggested several
methods for detecting and treating outliers. We investigated two different methods-data
censoring and transformation-for treating outliers in aptitude test data at the item level and total-
score level and their effects on the internal consistency and predictive validity of six
computerized tests being evaluated by the U.S. Air Force. Results from our sample of more than
2,000 pilot training candidates indicated that neither outlier treatment method at either level of
analysis had significant effects on the tests' internal consistencies or predictive validities.
Possible reasons for these findings include the frequency with which outliers occur and the
robustness of linear modeling methods.

and Company Limited.


Psychological Association.

Lane, N. E., Oberman, A., Mitchell, R. E., & Graybiel, A. (1966). The thousand aviator study:
Smoking history correlates of selected physiological, biochemical, and anthropometric
measures. Naval Air Station Pensacola, FL: U.S. Naval Aerospace Medical Institute.
The Pensacola Thousand Aviator Study began in 1940 with the examinations of 1056
student aviators and flight instructors on a variety of physiological, psychological, and socio-
economic parameters. Follow-up examinations on the group were conducted in 1951, 1957, and
963. During the 1963 follow-up, smoking history information on 675 subjects was obtained by
questionnaire and confirmed by interview, together with concurrent data from clinical
examinations, laboratory tests, anthropometry, and personal history variables. Two smoking
variables were created, Cigarette Amount (CA) and Cigarette Years (CY), each on a scale of 1 to
5 points. From the concurrent data, 62 variables were selected for relevance and general interest
to be examined in relation to smoking. Twenty-four of the 62 variables had significant correlations (p < .05) with CA, and 16 showed significant relationships to CY. Findings are related briefly to previous research and problems of cause-effect isolation are mentioned. It is concluded that results in general support of previous findings on smoker-nonsmoker differences.


This article examines whether measures of certain personality constructs can predict pilot involvement in accidents. The author makes the assumption that there is not a unitary set of noncognitive measures that make one more likely to be involved in accidents. Rather, there may be some personal characteristics that, depending on the circumstances, predispose one person more than another to engage in behaviors that increase the probability that he or she will be involved in an accident. Specifically, the author attempts to answer the question of whether or not there are personality, motivational and/or other psychological variables associated with behavioral responses that, under various circumstances, are more or less likely to result in aircrew accidents. An extreme groups design was utilized including 89 F-4 fighter aircraft pilots who were involved in at least one class A mishap and 89 F-4 pilots who were considered to be accident-free (control group). Form A of the 16 PF was used to assess personality. This instrument measures 16 source or primary factors and eight secondary factors (in addition to including several validity scales). The impact of life changes on temporary accident proneness was also measured using the Holmes-Rahe Social Readjustment Rating Questionnaire (SRRQ). Accident involvement was a dichotomous variable, either "crash" or "safe." Relationships between the variables were examined using set correlations. None of the eight 16 PF secondary factors or the SRRQ scales yielded significant relationships with the crash criterion. For the 16 PF primary factors, pilots who crashed scored significantly higher than those who did not crash on conscientious and self-sufficient. They also scored significantly lower on trusting, naive and relaxed/tranquil. The multiple correlation between the predictor set (i.e., the significant 16 PF scales) and "crashing" was .52. The results were not cross-validated, but the estimated cross validation coefficient was .23. The author interprets the results as supporting the need to conduct "limited domain" research because past research has convincingly debunked the notion of a "grand theory" of accident proneness.


This report is part of an overall project to identify an evaluate alternative models for
selecting and classifying soldiers into military occupational specialties. Tasks 1 and 2 of the project are documented in this report. More specifically, this document contains descriptions of the selection and job assignment processes in the Army, Navy, Marine Corps, and Air Force in terms of their intended design and operational reality. Subsequent to the description of current military selection and classification systems, alternative models are presented and evaluated along numerous qualitative dimensions.


Lessard, P. B. (1981). A simulated aircraft landing test as a pilot selection test. Paper presented at the 23rd Annual Military Testing Association Conference, Arlington, VA. Pilot selection in the Canadian Armed Forces has been a continuing concern especially since attrition rates for trainees has been as high as fifty percent on a few courses. In the reported study, the utility of a custom-developed, microprocessor-driven aircraft landing test WT), was examined in terms of its added value to the current pilot selection battery. A CROMENCO II microcomputer was programmed to simulate the landing of a light aircraft where visual stimulation was presented on a CRT and the candidates used a “ joystick” and throttle to perform three landing tests. Several dependant measures were automatically recorded every 500 m sec. This ALT was a further development of earlier research conducted on another system. (Fowler,
One hundred and fifty male candidates applying for military flying training were tested on the ALT as well as on the current test battery which consists of a psychomotor test in an aviation tester, pencil-and-paper tests which tape verbal and quantitative aptitudes and a memory test. The candidates also completed a measure of cognitive style, the Group Embedded Figures Test, and selected scales from Jackson’s Personality Research Form. It was found that performance on the ALT was independent of performance on the current test battery as well as performance on the Group Embedded Figures Test and the selected scales from Jackson’s Personality Research Form. Future research into the use of the ALT as a selection device will take place once trainees have completed the primary flying training course.


The author discusses the increasing dependability and reliability of aircrafts and how this has made the role of the human operator much more critical. One effort to identify individual attributes that are important to aviation safety has been the delineation of five hazardous thought patterns (HTPs): anti-authority, impulsivity, invulnerability, macho and resignation. These HTPs are thought to have the status of constructs and are seen as mediating the link between basic psychological processes and irrational pilot judgment. Even though there is basically no validity research on these HTPs, they are a core element of the judgment training currently being conducted by several U.S. and Canadian airlines. The purpose of this study was to examine the construct validity of these HTPs using the 16 PF and Rotter's Locus of Control scale. Thirty-five male pilots participated (4 held commercial ratings and 31 had private licenses). Subjects completed three scales from the 16 PF (impulsivity, superego strength and integration self-concept control) and the Rotter Locus of Control scale. They also completed a self-assessment inventory designed to measure the five HTPs. When completing this inventory, subjects represented with 10 flight scenarios that describe errors in pilot judgment and asked to rank (from most probable to least probable) the reasons why they might have made the error. Results showed that fourteen of the subjects did not have a predominant HTP, but the percentage of pilots showing a single dominant HTP was similar to that found in later studies (e.g., Lester & Connolly, 1987). Of the remaining subjects, 43 percent primarily exhibited the invulnerable HTP, impulsivity was dominant in 20 percent and macho was dominant in 14 percent. The remaining HTPs were predominant in only a few subjects. An analysis of variance (ANOVA) revealed a significant relationship between the three HTPs and the integration self-concept control scale and Rotter's locus of control scale. Pilots with the invulnerable HTP scored significantly lower on the integration scale and pilots with the macho HTP scored significantly higher on the integration scale. Also, pilots with the dominant macho HTP were significantly more internally controlled than either the invulnerable or impulsive pilots. The authors conclude that this study provides some support for the construct validity of the HTPs, but notes that their sample was extremely small.


The authors state that errors in pilot decision-making (PDM) can be accounted for by five hazardous thought patterns (HTPs): anti-authority, impulsivity, invulnerability, macho and resignation. However, the evidence for these HTPs is largely anecdotal and their exact nature is
not very clear. For example, some researchers consider them to be attitudes, while others consider them to be personality variables. This study was conducted with five purposes: (1) establish base rates for the different HTPs; (2) examine the relationships among the HTPs; (3) determine how the HTPs are related to pilot personality; (4) determine whether the HTPs are related to involvement in aviation accidents and incidents; and (5) examine the utility of two forms of the Pilot Decision Making Questionnaire (PDMQ). One hundred fifty-two males in their late teens or early twenties (all pilots) were used as subjects in this study. They had an average of 190 flight hours. Subjects completed the Rotter Locus of Control Scale, four scales from the 16 PF (Emotional Stability, Surgency, Conscientiousness and Integration) and two forms of the PDMQ. In the PDMQ Form A, subjects are presented with scenarios and asked to indicate whether the pilot's behavior was "very much like me" or "not at all like me." This is designed to be a measure a subject's overall propensity toward irrational judgment. The PDMQ Form P describes irrational decisions and presents the subjects with reasons (each representing a HTP) for why that pilot behaved as he or she did. Subjects are asked to rate how likely it was that each of the reasons would have caused the pilot to behave in that manner. This was designed to measure the strength of each of the HTPs. These data were analyzed and subjects were classified according to their predominant HTP (39%-Invulnerable, 24%-Impulsive, 19%-Macho, very few-Resignation, 0% Anti-Authority). HTP intercorrelations ranged from .21 to .53 and all were positive. Subjects who displayed the Macho HTP were significantly more internally controlled than subjects who displayed the Invulnerable or Impulsive HTPs. They were also more Conscientious than those classified as Invulnerable. No other significant differences among the groups were found. Subjects were separated into groups with "better" and "poorer" judgment on the basis of their scores on the PDMQ Form A. Pilots with better judgment were more internally controlled and were better integrated than those with poorer judgment. No relationships were found between scores-on the PDMQ Form A and involvement in aviation accidents/incidents.


This article examines the relationships between behavioral attitudes and accidents for aviators and enlisted air wing support personnel aboard an aircraft carrier. One hundred fifty-six aviators and 879 support personnel participated in this study. These subjects completed a 22-item attitude questionnaire which asked them to rate the extent to which they agreed or disagreed with
each statement. The questionnaire was factor analyzed (varimax rotation) and items loading most highly on each factor were summed to form six scales: logic, adventurousness, discipline, concern with self and brashness. The accident criterion was the number of personal injuries (or deaths) which were sufficiently severe to require a visit to sick call. Results showed that the Adventurousness scale was significantly correlated with personal injuries among the enlisted air crew personnel ($r = .12$, $p < .01$) and with aircraft accidents among aviators ($r = .25$, $p < .01$). None of the other five factor-analytically derived scales were correlated with accidents for either sample.


A common goal of many clinical research studies is the development of a reliable clinical decision rule, which can be used to classify new patients into clinically-important categories. Examples of such clinical decision rules include triage rules, whether used in the out-of-hospital setting or in the emergency department, and rules used to classify patients into various risk categories so that appropriate decisions can be made regarding treatment or hospitalization. Traditional statistical methods are cumbersome to use, or of limited utility, in addressing these types of classification problems. There are a number of reasons for these difficulties. First, there are generally many possible predictor variables which makes the task of variable selection difficult. Traditional statistical methods are poorly suited for this sort of multiple comparison. Second, the predictor variables are rarely nicely distributed. Many clinical variables are not normally distributed and different groups of patients may have markedly different degrees of variation or variance. Third, complex interactions or patterns may exist in the data. For example, the value of one variable (e.g., age) may substantially affect the importance of another variable (e.g., weight). These types of interactions are generally difficult to model, and virtually impossible to model when the number of interactions and variables becomes substantial. Fourth, the results of traditional methods may be difficult to use. For example, a multivariate logistic regression model yields a probability of disease, which can be calculated using the regression coefficients and the characteristics of the patient, yet such models are rarely utilized in clinical practice. Clinicians generally do not think in terms of probability but, rather in terms of categories, such as —low-risk” versus —high risk.”

controlled trial, based on a 14 hour light aircraft course. Tests were marked in the air at 9 and 14 hours by independent examiners. When all of the students had completed the RAF Basic Flying Training course one year later, their tests results were compared with raining outcomes. The results indicated a very high relationship between flying test marks and probability of success in later training. The marks awarded by the examiners were more predictive than the assessments of the flying instructors. Based on the results of the trial, the RAF established a Flying Selection Squadron and the paper presents some data from follow-up studies which were carried out as a validation of the selection procedure.


This study examined the validity of a broad set of predictors for selecting European managers for cross-cultural training program in Japan. The selection procedure assessed cognitive ability, personality, and dimensions measured by assessment center exercises and a behavior description interview. Results show that the factor Openness was significantly related to cross-cultural training performance, whereas cognitive ability was significantly correlated with language acquisition. The dimensions of adaptability, teamwork, and communication as measured by a group discussion exercise provided incremental variance in both criteria, beyond cognitive ability and personality. In general, these results are consistent with the literature on domestic selection, although there are some important differences.


This paper continues to develop a measure of decision making styles among U.S. pilots. Previous studies have examined this measure with some success (Lubner et al., 1999; 1997). The measure was designed to be incorporated into a set of Advanced Aeronautical Decision Making (AADM) training programs. These programs are based on techniques used by experts (Adams, et al., 1995). Having a brief, reliable, valid and adaptable test that could be self-administered should improve the acceptance, implementation and evaluation features of the AADM programs. Such a measure would also be useful for research purposes, and perhaps, for pilot selection.


Currently there is a lot of interest in integrating unmanned aircraft systems (UASs) into
the national airspace system (NAS). Well-established processes are currently in place to certify all aspects of the design, operation, training, and maintenance of the aircraft now in the NAS. The FAA is drawing on these experiences and processes to develop certification criteria for UASs as there are requests for them to enter the NAS. It is important to identify the differences between manned aircraft and unmanned aircraft systems and how those differences will impact operations in the NAS. As these differences are addressed, all users that will impacted in the system should be considered, as well as the infrastructure of the certification processes and regulatory requirements. Operations in the National Airspace System will inevitably change with the introduction of UASs. As an industry, we must utilize established processes and current tools to work together to successfully adapt to these emerging technologies.


In a present-day aviation, either civil or military, the flight simulators are more intensively used. The flight simulator allows for training of motor, perception and decision-making activities under the conditions close to natural. The favorable relation of costs and security of training follow-up with reference to the same activities tested on a real flight is the main cause of a wider use of the flight simulators (Svoboda, Heron, Weinberg, 1993). According to Orlansky & Chatelier (1983), the use of flight simulators enables to save almost 50 percent on training time as compared to the same conditions in which no simulators were used. Appropriate calculations show that training costs with the use of simulators are only 80% of those with the use of the planes. Operation and service costs pertaining to the flight simulators depending upon the type of training tasks pay off gradually within the time interval from 7 months to 2 years.


The aim of the paper was to determine a degree of relations between the real results obtained with flight simulator by candidates examined for military aviation and their psychological results. The stochastic relation in the form of linear function between studied variables was particularly investigated. The linear analysis of regression was used for that purpose. The authors indicated a significant relation between such psychological variables as:
temperamental traits, eye-movement coordination, relation assessment: velocity-distance, mental
skills and reaction time with choice and simulated flight results. Low percent assessment errors
(approximation) proves not only of the power, but also of a high and significantly higher than
zero coefficient of multiple correlation.


MacKinnon, D. W., Crutchfield, R. S., Barron, F., Block, J., Gough, H. G., & Harris, R. E.
(1958). An assessment study of Air Force officers: Part I: Design of the study and
description of the variables. Lackland Air Force Base, TX: Personnel Laboratory.

Poor pilot decision-making in deteriorating weather is the leading cause of a significant
percentage of fatalities arising from aviation accidents in the last two decades. Research has
identified psychological factors underlying pilot decision strategies as the primary reasons for
faulty decision making during inclement weather conditions. In the present review, the authors
attempt to evaluate existing experimental data on the cognitive and affective processes that
govern pilot decision making in changing weather conditions, with specific emphasis on
instances of VFR (visual flight rules) flight into IMC (instrument meteorological conditions). We
present a consolidated model of pilot decision processes and at each stage in the decision tree we
discuss the possible intrinsic and extrinsic factors that might affect the efficacy of cockpit
decisions. Based on this model, we examine interventions aimed at reducing the incidence of
sub-optimal pilot decision making under poor weather conditions. Suggestions for improving the
quality of aeronautical decision making through the use of technology and training are provided.

Mahoney, H. (1957). An analysis of current personnel selection for navigator training in the
USAF. Maxwell Air Force Base, AL: Air University, Air War College.

15th Annual Military Testing Association Conference, San Antonio, TX.

selection. Paper presented at the 42nd Annual International Military Testing Association
Conference, Edinburgh, UK.

With the advent of democracy in South Africa, the personnel practices of its National
Defence Force had to change. Prior to the 1994 national elections, the practice of appointing only
white candidate pilot was never challenged. An increasing pressure has however subsequently
been brought to bear on the organisation to address the issue of representativity amongst pilots.
Integration of the pre-election armed forces resulted in a degree of representativity amongst junior
and support ranks. However, the senior and specialised ranks remain relatively unrepresentative,
consisting mainly of white males. The process of recruiting suitable black candidates for
technical professions is made difficult by the fact that these individuals are in high demand by
the private sector and depressed by the relatively poor remuneration packages within the
government departments. One area in which representativity has become a politically sensitive issue is that of black pilots. This has resulted in a degree of suspicion and enmity towards the current pilot selection procedure. The Military Psychology Institute (MPI) is currently coming under pressure to change the manner in which pilots are being selected for the SANDF. The author of this paper will thus present a critical review of the black pilot(s) recruitment and filtering process for such appointments. In addition, discuss the management of pilot selection as a professionally and legally defensible practice addressing the core areas of who sits in the selection panel, tests batteries used for this and finally, a brief look into how the training is being executed.


This research analyses the attrition rates of various enlisted personnel groups within different Army job assignments. Three regression equations are developed in order to project the attrition rate of eight demographic groups to 76 Military Occupational Specialties (MOS). Education, sex, AFQT, along with MOS assignment are the independent variables. The rates generated by these equations show where important tradeoffs exist with respect to personnel allocation and the expected rate of attrition.


The Proficiency-Based Aviator Selection System (PASS) is the result of an operational feasibility program developed to determine whether a learning sample approach could be used to select candidates for rotary wing aviator training. PASS was based on the Automated Pilot
Aptitude Measurement System (APAMS), a five-hour learning sample of fixed wing piloting tasks, developed for the Air Force Pilot Selection Program. PASS utilizes a UH-1 Flight Simulator (UH-IFS) with a five-degree-of-freedom motion bas to present the syllabus and test materials. The APAMS syllabus was extensively modified to conform to rotary wing flight operations, and the UH-IFS software was reconfigured and formatted to meet the PASS requirements for training and performance measurement. Four voice synthesizers (VOTRAX) were interfaced to the UH-IFS to provide vocal feedback. The operational capability of the PASS was demonstrated in a test with 11 experienced rotary wing pilots and 11 candidates for rotary wing training. The primary audience for this report will be operational personnel in training selection and simulation, particularly for rotary wing aircraft.


Examines the test battery used for pilot selection of the Air Force in Norway. Evaluation of the cognitive and psychomotor abilities; Decline of predictive validity on personality and intelligence test; Comparison of the test results.


Martoccia, C. T., & Kelley, P. H. (1957). Some differences among Naval aviation cadets who attrited during presolo stage, later basic flight training, and advanced air flight training. Pensacola, FL: U.S. Navy School of Aviation Medicine.


Medical Disqualifications Among Female Army Aviator Training Applicants. Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.


A goal set in the 2004 Cognition in Space Workshop was to review the literature on predictors of individual performance in real-world tasks and develop a set of tests to be used during all NASA-funded research simulating space missions (e.g., bed-rest studies).” This literature review examines skill taxonomies, skill measures, and skill/task mappings.


The Ability Requirements Scales developed by Fleishman and his associates have been widely used as a job analysis technique. However, the ability ratings obtained with this technique may be confounded by rater bias effects. Fifteen instrument-phase instructor pilots rated the extent to which 32 basic abilities were required to perform 16 helicopter training tasks. Psychometric analyses showed significant differences in rater distributions. The method of successive intervals was used to transform the ratings to a common, equal-interval scale.
Analyses of the transformed ratings indicate that significant task-ability discriminations were made that produced a logical, interpretable pattern of ability and task interrelationships.


The increasing specialization of rotary wing missions and aircraft has precipitated a reanalysis of traditional strategies for assigning student aviators to one of four rotary wing missions: cargo, utility, aeroscout, or attack. Although previous research has suggested that certain abilities are appropriate for inclusion in a classification algorithm, there are no data to indicate that there are differences in the ability requirements (types or levels) for the four missions. This paper describes the results of several analyses designed to compare the ability requirements of the four missions. Ability rating data, obtained from subject matter experts for each mission, were transformed using the Method of Successive Intervals (MSI) to remove systematic biases identified in the raters’ distributions. Analyses of the transformed data indicate that there are no differences in the types or levels of abilities required to perform the most demanding tasks for each mission.


Unmanned aerial vehicles (UAVs) are quickly becoming a part of the national airspace system (NAS) as they transition from primarily military and hobbyist applications to mainstream flight applications such as security monitoring, satellite transport, and cargo hauling. Before the full potential of UAV flight in the NAS can be realized, however, FAA standards and regulations for UAV operations must be established. Given the experience of the U.S. military that mishap rates for UAVs are several times higher than for manned aircraft (Williams, 2004)—over thirty times higher, in some cases (Department of Defense, 2001)—the importance of carefully designed standards and regulations is clear. Issues related to human factors are likely to be of particular concern in establishing guidelines for UAV flight. As noted by Gawron (1998), UAV flight presents human factors challenges different from and in some ways greater than those of manned flight. These arise primarily from the fact that operator and aircraft are not co-located. As discussed in more detail below, the separation of operator and vehicle imposes a number of barriers to optimum human performance, including loss of sensory cues valuable for flight control, delays in control and communications loops, and difficulty in scanning the visual environment surrounding the vehicle. Unmanned flight also allows the possibility that a single operator might control multiple vehicles simultaneously, a task likely to impose unique and heavy workload demands. The goal of the current work was to examine the existing research literature on the human factors of unmanned flight, and to delineate issues for future research to address. The topics discussed below are divided into the categories Automation; Perceptual and Cognitive Aspects of Pilot Interface; Air Traffic Management Procedures; and Crew Qualifications. As will be clear, however, the issues covered within the various categories are highly interrelated. Answers to questions about crew complement, for example, will be contingent on the nature and reliability of automation provided to support UAV operators.
Likewise, decisions about interface design will depend on the extent to which flight control is automated, with manual flight mode demanding traditional stick-and-rudder controls and automated flight mode allowing for point-and-click menu-based control or other forms of non-traditional interface.


A two-part experiment was conducted to assess the hypothesized utility of various time-sharing measures as indicators of performance in the general aviation flight trainer. Equal numbers of males (28) and females (28) participated as subjects. Part one involved single and dual performance on a single-axis, compensatory tracker and a digit-cancellation, reaction time task. There were no significant gender differences on time-sharing measures. Part two indicated significantly better male performance on all simulator variables. Separate multiple regression equations were calculated for males and females, as well as overall equations including gender as a variable. Besides gender in the overall equations, measures of time sharing skill were the best predictors of simulator performance in all three types of equations. The regression equations based on gender differed in constituent predictor variables as well as weightings on similar variables. The results demonstrate the utility of time-sharing measures as predictors of complex-task performance. Additionally, they suggest the appropriateness of employing gender based predictor equations when establishing training or selection criteria for male and female complex-task operators.


It is well known that stress can have an adverse effect on performance and that individual differences in responses to stress are varied. This research explores the potential of developing an objective measure of stress resistance and the possibility of developing a laboratory model of the effects of stress on performance. The paradigm developed reveals that (1) emotional stimuli disrupt performance, (2) the disruption is exacerbated by time pressure and task difficulty, (3) repetition of the emotional stimuli (habituation) eliminates the disruptive effect, (4) it is not the emotionality of the stimulus, but rather the threat component that is critical to the disruptive effect. These results parallel effects in everyday life and suggest that the paradigm shows great promise for developing a measure of stress resistance and a laboratory model of the effects of stress.


The Flight Aptitude Selection Tests (FAST), the Army's primary selection batteries for Warrant and Commissioned Officer aviators, were made operational in 1966 and have proven to be effective selection instruments. However, because of changes in initial flight training programs and the current population of aviation trainees, a revision of the FAST is underway. This study was conducted to determine how critical the need for revision is.

McReynolds, J. (1954). Administration of the Aviation-Cadet Officer-Candidate Qualifying Test

Score achieved by new airmen with a specific time limit for each section of the test are compared with those of new airmen taking the test under the operational administrative procedure which does not involve the use of specific time limits. The groups, equated on all variables of the Airman Classification Battery, took the test in 2 sessions; the actual testing time for both groups was the same. There were significant differences between groups for 6 of the 13 subjects with all differences in favor of the group with specific time limits.


The Table Reading Test (TRT) assesses an individual's perceptual speed, or how rapidly one absorbs and processes visual search information. The focus of this study was to determine the validity of the TRT scores as a predictor of pilot training success. A total of 116 subjects enrolled in an aviation program at a major university in the southwest were tested between the Fall 2005 and Fall 2008 semesters, inclusive. Their TRT scores were tested for correlations to one of several flight/academic performance criterion. The results of the analyses found the TRT best predicts: Time to Solo (r = -0.228, p < 0.024), Time to Private (r = -0.754, p < 0.001), and GPA (r = 0.283, p < 0.002).


A new approach to selection research has been outlined, one that adapts the concepts of n-dimensional geometry to multivariate data. The MMPI was utilized as a measuring instrument and by means of a discriminant equation based on two clusters, 83 per cent of a sample of 935 NavCads were correctly categorized into pass and fail categories. It is anticipated that future selection batteries may be supplemented by personality measures.


The Euro-NATO Joint Jet Pilot Training (ENJJPT) program at Sheppard AFB conducts Undergraduate Pilot Training (UPT) for 13 NATO nations with a focus on producing premier fighter pilots. As ENJJPT transitions to the new T-6 Texan II, the leadership is examining if the
current assignment model meets the needs of the US Air Force for the US students. To assign US students, the Senior National Representative uses the Merit Assignment Scoring System (MASS) to rank order students and assign aircraft based on preference and availability of assignments. MASS accounts for every activity in pilot training as well as a subjective input from the instructors as to the overall attitude and performance of the student. The score obtained from the MASS is categorical by assigning a weighting to a particular category of performance. Currently, there is no direct link between the skills needed to fly modern fighter aircraft and the MASS. Additionally, many of the skills learned in pilot training span multiple categories and it is possible for a deficiency to be buried in the MASS score. The goal of this research was to identify the core skills required to fly the various fighter aircraft through the use of a Combat Air Forces (CAF) wide survey instrument, interviews, and working group inputs. An assignment model was created with a focus on assigning students based on skill strengths. After the core skills were identified and related to UPT events, a value hierarchy was created and a model developed to identify the best aircraft fit for a student based on their performance as related to the skill sets. This paper frames the issues, outlines the methodology used to define the skill sets, and discusses the development of the model. Finally, recommendations are made on future changes to MASS, the UPT student assignment process, and the pilot training syllabus.


The objective is to analyze and specify the fundamental flying abilities which comprise the training objectives of Undergraduate Pilot Training (UPT). The results of this analysis will be used as a foundation for structuring research on and recommendations for improvements in Air Force flying training programs. The Phase I effort both the surface analysis and the taxonomy structure. Examination of previous task and skill taxonomies failed to provide a usable basis for the present effort. The surface task analysis was developed on the basis of a breakdown of task elements according to the cue, mental action and motor action involved. The flying tasks analyzed were found to fall into three categories: fundamental transitions, composite transitions and continuous transactions. The surface task analysis was organized so the more complex flying maneuvers could be accommodated by a sequence of two or more of the three categories of task types identified. A cubic taxonomic structure was developed with cue, motor action and mental action dimensions. A set of classification rules were provided for locating any flying training task in a specific "pigeon hole" within the taxonomic structure. A procedure for evaluating the validity of the taxonomic system was established for use during Phase II of this program.
In June 1977, the U.S. Army Aviation Center (USAAVNC) instituted the 175/40 Initial Entry Rotary Wing (IERW) Flight Training Program. An integral feature of this program is a dual track in which students are selected for specialized tactical training employing the Aeroscout (OH-58) aircraft or for training in the Utility (UH-1) helicopter early in training. Success with the 175/40 program led to the concept of the Mission Track program in which students would receive specialized training in one of the four basic helicopter missions: Aeroscout, Attack, Utility, and Cargo. The intent is to replace the present IERW plenary rotary wing aviator with a "systems aviator" who will normally remain dedicated to a specialized aircraft system/mission for the balance of his military career. It is anticipated, that the systems aviator concept will result in more proficient aircrew performance in the tactical environment, thereby enhancing combat readiness in U.S. Army Forces Command units. The overall goal of this research effort was to initiate development of a Mission Track selection process. Four specific tasks had to be accomplished. First, the skills, aptitudes, and abilities (i.e., attributes) required for success in each mission had to be identified. Second, standardized tests and training course measures which assess those attributes had to be administered to IERW students. Third, preliminary selection algorithms had to be developed. Fourth, recommendations for a final Mission Track Selection Battery including suggestions for validation, had to be developed.


Miller, R. E. (1960). The Cadet Screening Test as a predictor of AFOQT aptitude composites. Lackland AFB, TX: Personnel Laboratory.


The Air Force Officer Qualifying Test (AFQQT) is used in various officer procurement and selection programs. Scores on this test are of significance in selecting officers for attendance at basic technical courses. A study of 975 reserve officers in seven different technical courses was therefore conducted to provide data on the predictive validities of AFOQT composite scores for final technical course grades. Satisfactory validity coefficients were obtained for the AFOQT aptitude composites against the course criteria. Most of the composites were valid for each separate criterion, and coefficients as high as .58 were obtained. Evidence was found that these validities persist in different samples of officers enrolled in the same course at different times. Validities of the AFOQT interest composites, however, were markedly fewer and frequently negative. The highest in terms of absolute value was .32.


Applicants for each Air Force Academy class take a battery of selection tests to establish their qualifications. Entering cadets take an additional battery consisting mainly of nonacademic experimental tests, developed as part of a program for the production of officer selection and classification devices. Both batteries are validated at the end of the fourth class year against academic and nonacademic criteria. In the class of 1964 the criteria were the Academic Standard Score, Cadet Effectiveness Rating (CER), Residualized Cadet Effectiveness Rating (with respect to physical aptitude), Extracurricular Activities Standard Score, Nonacademic Standard score, and Early Motivational Elimination. Using multiple regression techniques, it was found that
there are measures in both the selection and experimental batteries having validity for each of the criteria. Multiple correlations up to .01 were obtained with the Academic score as the criterion, and up to .51 with the CER. Validities are not significantly different from those observed in the class of 1963 for selection tests common to both classes. Previously observed fluctuating validities appear to have stabilized.


Objective: The Implicit Association Test (IAT), in combination with a battery of additional psychometric tests, was employed to examine the accuracy with which it predicts pilots' risk-taking behavior. Background: Risk management is an integral part of piloting. Many factors affect pilots' risk management, including individual differences. Therefore, employing a unique methodology from social cognition, the present study examined the influence of attitude, as measured implicitly through the IAT, personality, and flight experience variables on pilots' risk-taking behavior. Method: In addition to a simulated flight on a computer-based flight simulator, 35 pilots completed a battery of psychometric tests. Results: Among the 6 risk perception variables, 10 risk attitude variables, and 2 experience variables, only 2 variables were found to be significantly related with in-flight risk-taking behavior: everyday risk (risk perception) and the IAT effect (attitude). Of these, the IAT effect was the strongest predictor of flight behavior. Conclusion: The results indicate that implicit attitudinal measures, such as the IAT, provide a more accurate forecast of pilot behavior than do the more traditional explicit attitudinal or personality measures. Application: An implicit attitudinal measure can be proactively employed to identify pilots who are potentially more likely to engage in high-risk activities, hence permitting a more strategic approach to pilot training.


The psychiatric qualifications for Naval aviation duty are more stringent than those for general duty and disposition is tied much more closely to diagnosis. These facts apply not only to DSM-IIIR Axis I where diagnoses render an individual "Not Physically Qualified" (NPQ), but also to Axis II diagnoses, Within the Navy, a variety of recommended general duty dispositions may be made by the clinician if a personality disorder is diagnosed. The patient's command may then consider these recommendations in formulating an appropriate administrative action. Separation is either pursued or the patient returns to duty in his or her rated specialty. Within the aviation community, personnel must be found to be both psychiatrically physically qualified (PQ, no Axis I diagnosis) and aeronautically adapted (AA, no Axis II diagnosis) in order to
remain on active flight status. The most recent editions of the Navy Manual of the Medical Department (Chapter 15) and the Navy Flight Surgeons Manual state that any Axis II diagnosis is cause for rejection of candidates and disqualification of designated individuals. In addition, prominent personality traits which result in maladaptive behavior, emotional instability, or impaired judgment may render the patient "Not Aeronauticaly Adapted" (NAA) if safety of flight, crew coordination, or effective mission execution are effected. For all practical purposes, a finding of NAA is considered permanent, and it is for this reason that recent guidelines require that designated individuals be referred to the Naval Aerospace Medical Institute (NAMI) for final determination. Once referred, all patients take the MMPI and complete several clinical interviews. Normative data have been established which improve the sensitivity and specificity of the MMPI within the Naval Aviator and Naval Flight Officer communities. Efforts are currently underway to establish similar norms for the MMPI-2.


The U.S. Navy is revising the Navy/Marine Corps Aviation Selection Test Battery (ASTB); the existing version has been used since its previous revision in 1971. The ASTB consists of four paper-and-pencil tests: The Academic Qualification Test (AQT) measures
general intelligence. The Mechanical Comprehension Test (MCT) measures the ability to perceive physical relationships and solve practical problems in mechanics. The Spatial Apperception Test (SAT) measures the ability to perceive spatial relationships from differing orientations. The Biographical Inventory (BI) contains items concerning personal history, interests, and aviation knowledge. The MCT, SAT, and BI scores combine to produce the Flight Aptitude Rating (FAR). The ASTB is frequently referred to as the "AQT-FAR with its summary score consisting of two Stanine scores, e.g., 8/6, meaning an AQT Stanine of 8 and a FAR Stanine of 6. The existing BI was validated using Student Naval Aviators (SNAs); however, the BI has also been used to select Student' Naval Flight Officers (SNFOs). Hopson et al. (1978) investigated the relationship between pass-attrite during undergraduate flight training and AQT, SAT, MCT, and BI scores. Their results are reproduced in Table 1. For SNAs, all tests increased the multiple correlation (R); however, for SNFOs, the SAT and BI did not significantly increase R. The purpose of the present study was to develop and validate separate BI scoring keys for SNAS and SNFOs for the Revised ASTB.


This study investigates selection criteria for selecting female aviators for training in combat helicopters (AH-64, AH-1, OH-58D, and RAH-66). Selection for such training would occur as either a part of the multi-track program of instruction used in the current Initial Entry Rotary Wing flight training course, or as transition training for already qualified aviators.
Analysis included a review of: current Army Regulations governing prerequisites for combat helicopter training and combat helicopter maintenance test pilot training; Initial Entry Rotary Wing selection criteria for combat helicopter tracks (AH-1 and OH-58); Aviation Branch Personnel Manager interviews; Combat Helicopter Manprint/Anthropometric restrictions; Standards of medical fitness; Anthropometric standards; and previous reports on female performance in Initial Entry Rotary Wing training. Conclusion supports selecting females for combat helicopter training using the same selection criteria currently used for choosing males for such training. Study recommends additional research in aircraft accommodation measurements; social-psychological aspects; and physical body strength requirements. Helicopter pilot training, Selecting women aviators, Military helicopter pilot selection.


The purpose of the present study is to develop a Mission Track Assignment Battery (MTAB) that can be used to assign student pilots to different mission tracks for training depending on their demonstrated skills and abilities. The development of such a testing system will allow the Army to make more effective and efficient utilization of personnel. However, before the test battery could be implemented it was necessary to investigate the nature of the work performed in the different missions and to develop objective and valid instruments which measure the required abilities and predict pilot success and proficiency in training and in the field.


A test battery was developed to represent a broad range of abilities and skills that were
identified as important in piloting helicopters. The battery was developed on the basis of a
taxonomic approach to job analysis which linked pilot tasks to ability requirements for different
mission tracks. The tests developed were based on an earlier review, by ARRO staff, of the kinds
of tests likely to measure the abilities identified in the job analysis approach employed. The
present study included the development, programming and pre-testing of computer interactive
tests designed to measure abilities as underlying critical tasks in the various helicopter missions.
conditions of administration were developed and test reliability Tasks judged by expert pilots to
be critical for pilot effectiveness were identified as possible measures of performance in the
different mission tracks. These tasks can be translated into criterion measures and used in
validating the test battery developed. The purpose of the validation will be to empirically link
test scores with performance in these critical tasks representing the different mission tracks. The
findings would indicate if all the tests are needed or if empirical validities showed which limited
set of tests can be used to predict mission performance.

interactive tests for assigning helicopter pilots to different training missions. *Vertiflite*,
31(5), 48-51.
The use of helicopters continues to grow in such areas as law enforcement and military
operations. As the aircraft become more technically sophisticated and their missions more
complex, increasing demands are placed on the pilots. Consequently, there is a need to improve
the effectiveness of the selection and training of pilots to ensure that pilots have the skills and
abilities required to operate helicopters effectively and safely. Recently, there have been some
developments in the field of human performance measurement which should help to ensure that
pilots have the proper skills. More specifically, advances in computer technology such as the
success of the microcomputer have opened the way for broader and more cost-effective use of
computers in the selection and assignment of personnel. As a result, there has been a growing
effort undertaken by the military services and private corporations to make use of the computer's
capabilities in selection testing. Currently, the U.S. Army is developing computer-based tests for
selecting and assigning Army pilots to different helicopter training missions.

National Research Council (U.S.). Committee on Aviation Psychology. (1949). *Proceedings of a
Conference on the Selection and Classification of Enlisted Aviation Personnel, May 6, 1949.*
Washington: Division of Aviation Medicine, Bureau of Medicine and Surgery, United States Navy.

National Research Council Committee on Selection and Training of Aircraft Pilots. (1942). An
historical introduction to aviation psychology. Washington, DC: Civil Aeronautics Administration.

NATO Advisory Group for Aeronautical Research and Development. (1954). *Methods and
criteria for the selection of flying personnel, symposium held 23-25 February 1953, Paris, France.*

notes from the Aviation Psychology Laboratory; research report of the U.S. Naval School of Aviation Medicine.* Pensacola.


Operational populations exposed to extreme risk environments (EREs) might expect to sustain substantial losses, yet must be able to be relied upon to complete their mission or missions regardless. Existing human (personnel) reliability programs are inadequate to assure that personnel capable of meeting both the necessary security and operational requirements are available for response to such conditions. This study explores a number of issues to consider in building a robust human reliability program (HRP) structure capable of supporting single to multiple operational populations, scenarios, and missions, using any of several program structure formats. The HRP structure format may be used within a single agency or government-wide.


A comprehensive selection examination in the Republic of Singapore Air Force (RSAF) aims to minimise medical wastage of military pilots who have to function safely in the unnatural environment. Of the 8642 applicants examined, 657 (7.6%) were rejected for non-medical reasons before completion of medical examination. Of the remaining 7778, 58.7% passed the selection examination while 41.3% failed. Ophthalmological (34.3%), anthropometry (23.7%), and ENT (13.7%) conditions were the three major causes for failure (71.7%). Myopia and astigmatism accounted for 57.6% while squints accounted for 22.1% of the ophthalmological conditions. Amongst ENT conditions, 70.8% were for marked vasomotor rhinitis, sinusitis and nasal septum deviation with marked narrowing of one or both nasal passages, while 22.4% were for permanent abnormal hearing threshold shifts above the minimum standards.


Changes in the selection procedures for the U.S. Navy's landing craft air cushion (LCAC) vehicle operator training program are under development. Several cognitive, personality, and psychomotor selection tests are being evaluated. This study analyzed the performance of 36 LCAC operator trainees on an automated series of single and multiple psychomotor (PMT) and dichotic listening (DLT) tests that measure abilities involving eye-hand-foot coordination and divided attention. Point-biserial correlational analyses between test measures and training criteria resulted in a number of statistically significant correlations ($p < .05$). Additional analyses using the t test indicated that successful trainees performed significantly better than unsuccessful trainees on PMT measures ($p < .05$). In multiple task conditions, where PMT and DLT were administered simultaneously, only performance on PMT subtasks indicated significant differences between the two groups. These findings suggest that psychomotor tests have the potential to predict LCAC training program outcome.

To date, only limited entry requirements exist for selection of vehicle operators for the U.S. Navy landing craft air cushion (LCAC) vehicle training program. What these requirements should be has not been empirically determined, hence a research effort in this area is needed. An additional impetus for such research has been a series of costly accidents resulting from operator error. Our objectives were to develop a cognitive, psychomotor, multiple-task, and personality-oriented test battery having the potential to predict the training outcome of LCAC operators and serving as an LCAC-personnel screening system. Automated tests used included Dichotic Listening, Psychomotor, Manikin, One-dimensional Compensatory Tracking, Digit Cancellation, and Risk-taking. Significant predictors of training grade criteria included a multiple Dichotic Listening test. Training grade also correlated with the stick-rudder-throttle conditions of the Psychomotor task and with the One-dimensional Compensatory Tracking task when performed in combination with the Digit Cancellation task. Risk-taking tendencies were also significantly related to overall training grade. These findings suggest that components of the test battery have the potential to predict LCAC training performance.

Determinants of naval flight officer program entry and pipeline choices. *Aviation, Space, and Environmental Medicine (News Note)*, 59(1), 74-77.


As the final phase of a test-development project, 3 forced-choice self-report tests were administered to student groups to develop empirical scales predictive of peer-nomination personality ratings drawn from the same groups. Five factor scales were developed for each test. For the independent sample of a double cross-validation analysis, all of the scales correlated positively with their criteria. Combined scales for the 3 tests showed usefully high correlations. Multiple regression analyses demonstrated that scores from other personality tests administered to the students did not add appreciably to prediction of the peer-nomination scores. The self-report tests yield better prediction (.38-.47) of 3 of the peer-nomination factors (Extroversion, Conscientiousness, Culture) than of Agreeableness (.31, 1.34) or Emotional Stability (.26, .29). From a second administration of the self-report tests with instructions to fake responses favorable for acceptance for officer training, detection keys were derived that identified a high percentage of faked performances, and factor scales were developed which reduced the occurrence of faked scores in the extremes of the distributions.


A new technique for measuring individual differences in basic attention capabilities and the validity of these differences in predicting success in flight training were investigated. The testing system included a digit-processing, reaction-time task and a one-dimensional compensatory tracking task. Comparisons were made between separate and concurrent performances of these tasks, with both equal and shifting task priorities. Adaptive techniques were employed to obtain maximum performance levels for each subject in the single-task condition and to maintain dual-task difficulty within subjects. Consistent individual differences in basic attention capabilities were observed and several dimensions of attention capabilities are suggested. A preliminary validation study compared scores for flight instructors and student pilots. In addition, the student sample was dichotomized based on performance in training. There were reliable differences for both groups on dual-task performance efficiency.
Recent research in predicting aviator performance in flight training from measures of divided attention performance has emphasized the need for the development and refinement of testing methods, investigation of the range and consistency of the individual differences associated with the measures, and the relationship of these measures with score on existing paper-and-pencil tests currently used for selection of aviators. Gopher and North indicated that there was a wide-range of individual differences in single and dual task performance of one-dimensional, compensatory tracking and a discrete, digit-processing, reaction-time task. Furthermore, there were low and non-reliable correlations between single-task performances and generally low and non-reliable correlation between single and comparable dual-task score for each task. These results suggested that the chosen tasks represented independent performance abilities, and that single-task performance was independent from time-sharing performance. In addition, the individual differences demonstrated in time-shared performance were consistent across various experimental manipulations of task priorities in the study, indicating a high degree of reliability of these measures. The goals of this study were to investigate (1) test-retest reliability coefficients of single and dual-task performance measures over separate test days$ (2)$ the range and consistency of individual differences in the measures and degree of relationship of the attention scores to scores on a standard set of aviation selection tests, including the Academic Qualification Test (AQT), Mechanical Comprehension Test (MCT), Spatial Apperception Test (SAT), and Biographical Inventory (BI). The investigation of these relationships is at important initial step in planning subsequent investigations of the validity of these attention measures as predictors of performance in various phases of naval aviation training. Test-retest reliability is important in ensuring that changes in subject motivational levels, physiological states, or rate of skill acquisition do not bias the measurement of the skills being assessed from day to day. Relationships between single- and dual-task performance are important in determining the independence of time-sharing of the candidate from separate task performance capability. The correlation of attention measures with the paper-and-pencil tests will determine the feasibility of continuing the investigation of these measures as independent predictors of aviator performance.


Oberman, A., Lane, N. E., Mitchell, R., & Graybiel, A. (1965). The Thousand Aviator Study: Distributions and Intercorrelations of Selected Variables Naval Air Station, Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Oberman, A., Mitchell, R., & Graybiel, A. Thousand aviator study. Naval Air Station Pensacola, FL: Naval Aerospace Medical Research Laboratory.


Advanced U.S. Army technology and hardware systems place a higher cognitive demand on the individual soldier than ever before. Sophisticated weaponry and hostile mission environments of modern conflict threaten to overwhelm the capacities of the human operator. New selection and training instruments are being developed to (a) select people most likely to perform well under high cognitive demands, (b) identify weaknesses in people, and (c) alter or train the person to improve response to the increased cognitive work load. The primary goals of this Phase I SBIR effort were to develop a new conceptual model and to suggest new testing and training approaches to handle the cognitive complexity of many Army tasks. Such approaches may enhance the identification and training of people to perform cognitive tasks efficiently during conditions of extremely high work load. To begin this process, a general "nonlinear" model of performance was first developed by exploring performance theory; this theoretical orientation was then translated into practical assessment and training tools to select and enhance people likely to excel at tasks demanding particular combinations of skills. A nonlinear approach to combining these procedures into a practical "test battery" and a specific training approach based on this model were proposed.


The role of the modern pilot requires a high degree of situational awareness. This involves the ability to search for relevant information, assess opportunities and priorities, and maintain performance under stress. The PC-based WOMBAT[TM] test has been designed to measure individual aptitude to cope with such demands. In the first experiment performance on the WOMBAT test was compared with performance on a batter of tests of specific underlying abilities. In the second experiment the performance of elite soaring pilots was compared with that of matched pilot and control groups. The results support the theory that the WOMBAT test measures individual ability to maintain situational awareness and that this ability is found in high levels in elite pilots.

Is the difference between an accident and an incident merely a matter of luck, or are there significant differences in terms of flight crew performance and associated performance shaping factors (PSFs)? This study obtained self-report data from a sample of 1,144 New Zealand pilots. There were differences between accident- and incident-involved pilots in terms of age, flight experience, and involvement in hazardous events. Incidents were more likely to be attributed to failures to detect or diagnose information, whereas accidents were more likely to be attributed to failures to choose an appropriate goal or strategy. There was no difference in the quantity of PSFs associated with accidents and incidents. These self-report data are consistent with previous findings based on external coding of air accident reports.


A comparison of the validity of psychometric g and specific ability or job knowledge, s, for predicting pilot and navigator criteria was conducted. Psychometric g and s were estimated from the principal components of a multiple aptitude test battery. The criteria included passing-failing training, an overall performance composite, academic performance, and work samples of pilot and navigator tasks. Regression analyses conducted to evaluate the predictive efficiency of g and s demonstrated that g was the best predictor of all criteria and s contributed little beyond g.


The situation awareness (SA) and mental workload of 56 subjects were evaluated as they monitored one or more attributes of six objects moving systematically over a rectangular grid. Subjects were assigned to one of seven groups depending upon whether they were to monitor object locations (location task), object colors (color task), whether the objects flashed (flash task), or some combination of these three. Both task performance and subjective ratings were used to assess subjects' awareness of the three object attributes. In addition, subjective ratings of mental workload were collected. All subjects performed the monitoring task under four different conditions formed from the factorial combination of 1) the probability that objects of a certain color would flash and 2) whether object colors remained consistent or changed during the course of a trial. The results pointed to the usefulness of both flash and color task performance as measures of SA. Subjects were very poor at the location task, suggesting either their location awareness was poor or the location task is not a good measure of that awareness.
ratings proved useful but occasionally dissociated from task performance. One possibility is that subjective ratings reflect rational inferences by the subjects rather than the outcome of their introspections.


In November, 1980, the Army Research Institute through its Fort Rucker Field Unit and Florida State University initiated a three-year research project to evaluate the Revised Flight Aptitude Test (R-FAST) for possible sex and racial bias, and to develop experimental test items if bias in the present instrument was observed. This project is being conducted in four phases. The first phase included the evaluation of the R-FAST for possible bias. Parallel and experimental Item forms and test items are to be developed in the second phase. The experimental items are to be correlated with the parallel items and validated against a criterion standard in the third phase. The Performance-based Aviator Selection System (PASS), currently being validated at Fort Rucker, will represent the criterion in the third phase. The experimental items are to be correlated against final IERW grades in the fourth phase.


Most of the young men who want to join the Air Force as a pilot must become officers and have to pass the Central Personal Office, the only selection institution of the German Armed Forces to test the personal conditions of all applicants for all 3 branches of the Bundeswehr. For further information of the fundamentals and the diagnostic methods used for the selection of officer candidates please join the report of KLASSMANN. The holistic assessment concept of different qualification levels for an officer career requires different psychodiagnostic methods like psychometric measurements, written data (essays, questionnaires) and behavioral observations and assessments by a testboard (2 officers, 1 psychologist) in an interview, round-
table-discussion, short-cut lecture role plays and sports. The selection concept and the criteria which are run to accept or reject an applicant require for the fulfillment of the general officer assignments in different military fields, but do not regard to the particular requirements of the Flying Services.


Park, K. S., & Lee, S. W. (1992). A computer-aided aptitude test for predicting flight performance for trainees. *Human Factors, 34*(2), 189-204. Perceptual/psychomotor and cognitive tasks in a computer-aided aptitude test were studied to predict the success of a trainee in flight training. Pilots' tasks were tentatively classified into five categories: tracking, reaction, memory, estimation, and visual scanning. To investigate the performance of a trainee in these categories, 16 single tasks and 10 dual tasks were examined. In the factor analyses three common factors (tracking, reaction, and memory) were meaningfully extracted. To select significant tasks for predicting flight performance, we performed stepwise regression and discriminate analyses. In the regression analyses, memory tasks were most significant in predicting the flight performance of a trainee. In the discriminate analyses, tracking tasks were most significant for distinguishing the passing and failing groups.


Two hundred and three Air Force ROTC subjects were administered a large battery of printed and apparatus psychomotor reference tests from which 50 scores were taken. Following
administration of the reference tests, subjects devoted 17 sessions distributed over a six-week period to practice on a complex tracking task. The matrix of intercorrelations among these scores was factor analyzed and 15 ability factors identified. An analysis then was conducted of the extent to which variation in performance in tracking at the different stages of practice could be accounted for in terms of the identified ability factors. The ability factors accounted for only a small portion of the variance in tracking performance. Hypotheses are offered concerning the selection of a different set of reference measures which might be more effective. The analyses of these data do indicate, however, that the prediction of terminal tracking proficiency is better accomplished through a set of external measures than through initial scores taken directly from the tracking task. Early proficiency on the task itself was unrelated to terminal proficiency.


This report reviews the literature reflecting the employment of perceptual-psychomotor tests for selection of aircrew members since World War II and provides behavioral concepts for consideration as possible future test development areas. The review considers the use of flight experience as well as perceptual-psychomotor screening devices and comments on the results of the programs in which such experience is intentionally used. The fundamental importance of criterion definition to development and validation of selection devices is discussed. Recent research is reviewed leading to the derivation of behavioral concepts recommended for consideration as principles on which new perceptual-psychomotor tests may be based. The merits of simple tests as opposed to complex tests in which numerous facets of performance are concurrently assessed are considered and the latter approach is recommended. References are included in support of the review and critical items are annotated.


Aviation is a unique military occupation which requires a substantial amount of psychological resources from those who fly. As such, military aviation has a long, though somewhat uncomfortable relationship with psychological assessment. Psychologists were involved in aviation selection as early as 1919 (Henman) and we continue to be involved with aeromedical and selection research, as well as pilot selection, and clinical assessment of pilots for return to duty (retention) to this day. In some instances, however, psychologists are distrusted for fear that psychological information will be misused. This paper will discuss the issues of psychological assessment in the areas of aviator selection and retention. The issues involved with
the purpose of assessment, methods of assessment, and focus of assessment will be discussed.


This report presents a review of research in the aviator selection and general personnel selection domains. That information was used to identify knowledge, skills, attributes, and other factors that should be included in a job analysis focusing on the Army aviator job. It was further used to develop a recommended strategy for an Army aviator selection battery.


There is currently a surge in the utilization of Uninhabited Aerial Systems (UAS). Although the importance of the human in the system is often ignored with a focus upon the physical airframe, there are nevertheless numerous human factors issues that must be considered one of which is the training of operators. This paper will describe the inventory and assessment of existing U.S. military and civilian UAS operator training activities and programs conducted by the Arizona State University group of the UAV Alliance, Research, and Curriculum Development Partnership Program. The paper will then discuss various avenues of future research pertinent to operator training including what training backgrounds UAS operators should possess, issues in team training, and use of simulators.


Growing acceptance of a taxonomy of personality traits developed by the Air Force in the late 1950s (Tupes & Cristal, 1961) has prompted a reexamination of the utility of personality measures for aircrew selection and classification research. Candidate theories are identified and then evaluated according both to general scientific criteria and to specific operational criteria (e.g., Hall & Lindzey, 1978; Imhoff & Levine, 1981). The Five Factor Model (Goldberg, 1990; McCrae & Costa, 1985; Tupes & Cristal, 1961) is selected as the most suitable framework for guiding future Air Force research in the personality domain. Example items to measure relevant characteristics are proposed, as are directions for future research.

An evaluation of Army, Air Force and Navy pilot selection and classification research and operational selection procedures resulted in several conclusions. The services share significant commonalities. For example, all three services utilize paper and pencil tests for pilot selection that are augmented by computerized, performance-based selection or classification tests and the individual selection systems assess common personal attributes. All three services have documented that computerized testing improves accuracy for predicting success in flight training beyond that of paper and pencil testing alone. The working group found that the requisite skill, abilities and attributes for completion of flight training are similar for Army, Navy and Air Force. This is supported by the fact that the services test similar domains in their respective selection programs. Although no service currently has the resources to properly maintain a computer test battery while improving aviation selection technology, their pooled resources could support a single battery. It is recommended that the services consolidate administration, maintenance, and research and development of a common computer test battery to select and classify aviation candidates. This consolidation could provide a significant savings over the cost of each service conducting its own test maintenance and research programs.


Cadets were instructed to indicate whether each of the other members in their class was
"were careful" or "less careful" than themselves. The Carefulness Ratings (CR) were correlated with scores on each of the primary selection tests and grades in the U.S. Naval School, Pre-Flight, and the addition of CR to the validity of the Pensacola Student Prediction System was investigated. Carefulness Ratings had significant relationships to the majority of the primary selection tests and Pre-Flight grades currently employed as predictors. For this reason, despite a significant relationship of CR to the criterion of success/failure, its unique contribution was too small to be of practical value.


The Edwards Personal Preference Schedule, a forced-choice personality "need" inventory, was evaluated as a predictor of success in naval flight training. The EPPS failed to discriminate between student aviators who completed training successfully and those who dropped voluntarily or failed due to poor performance. The schedule showed little promise as a predictive instrument for flight training.


Academic grades received during primary training were evaluated as predictors of success in naval aviation training. The addition of primary academic grades to the current prediction formula resulted in a significant increase in its predictive effectiveness.


This paper presents a history of naval aviation psychology from the onset to the close of the Second World War. It is based upon archives from the files of the first Aviation Psychology Section in the Navy’s Bureau of Medicine and Surgery. The paper discusses the evolution of the first naval aviation selection tests and early organizational considerations which ultimately led to the formal establishment of aviation psychology in the Navy.


Professionals in personnel research have long been searching for a culture-fair predictor of Scholastic Aptitude which might indicate future-potential success in relatively complex technician-level jobs of Ss (those educationally disadvantaged, but with hidden and hard-to-measure potential) Guion (1965). Working with similar kinds of situations, Safford (1967) employed Dunn's Object-Sorting Task to predict academic performance of school children. Prediction is the prime concern of personnel research. In personnel psychology, the problem often boils down to the prediction of adequate performance on moderately complex, technician-level jobs. Typically, the familiar measures of "scholastic aptitude" (which admittedly tend to reflect a bias in favor of the length and location of schooling) have been the best single predictor of achievement, and supervisors' rating been the most often used criteria measure of job proficiency.


The purpose of this study was to validate selection performance standards for the screening of candidates for entrance into the US Navy and Marine Corps Unmanned Aerial Vehicle (UAV) Pioneer Pilot training program. A minimum Pioneer crew consists of an external pilot (EP), internal pilot (IP), and a mission commander/payload specialist (MC). The EP is responsible for take-offs, landings, and control of the vehicle when it is within visual range. The IP is responsible for control of the aircraft when it is beyond visual range. The MC is responsible for planning and execution of the mission, operation of the payload, and for information gathering during the mission. In the development and initial validation phases of this system, a task analysis was completed in training and fleet squadrons to identify both tasks that are critical for safe flight and skills required to perform piloting tasks. Specific computer-based psychomotor tests were chosen as predictor variables based on the task analysis and initial validation. In the present study subjects consisted of 39 students: 5 IPs and 34 Ground Control Station Operators (who received combined IP and MC training) for whom both psychomotor test battery scores and training outcome data were available. A single, four-component, unit-weighted, composite scoring algorithm was generated to indicate performance on the computerized test battery. This composite score was found to be a significant predictor of final average in primary UAV training ($r = .59$, $p < .001$). Mean composite scores also significantly differed between students who ultimately qualified as operators in their operational fleet units and those who failed to qualify ($t=-2.92$ (37), $p < .01$).


Poe, A. C. (1953). Effectiveness of the flight aptitude rating battery for the selection of naval aviation cadets. Pensacola, FL: Naval School of Aviation Medicine.


This chapter focuses on the selection and training of people who work in aviation specialties. Aviation work encompasses a full spectrum of activity from operators of aircraft (i.e., pilots), to flight attendants, dispatchers, flight controllers, mechanics, engineers, baggage handlers, ticket agents, airport managers, and air marshals. The topic covers a lot of territory. For manageability, we concentrated on three categories of aviation personnel: pilots and aircrew, maintenance technicians, and flight controllers. One problem shared by nearly all aviation specialties is their workload. Workload within most categories of aviation work has been increasing since the beginning of aviation. In the earliest days, available technology limited what the aircraft could do, similarly limiting the extent and complexity of aircraft operations. Pilots flew the airplane from one place to another, but lacked instrumentation to deal with poor weather conditions—conditions that were simply avoided. Maintainers serviced the airframe and engine, but both of these were adapted from relatively familiar, non-aviation technologies and materials. Flight controllers, if they were present at all, were found standing on the airfield waving red and green flags. Since those days, aircraft capabilities, aircraft materials, and aviation operations have progressed remarkably. The aircraft is no longer a limiting factor. Pilots, maintainers, and controllers are no longer pushing aviation technology to its limits, but are themselves being pushed to the edge of the human performance envelope by the aircraft that they operate, maintain, and control. To give an idea about the work for which we are selecting and training people, it may help to discuss the workloads that different specialties impose on aviation personnel. The following is a short discussion about each of the three selected aviation specialties and the workloads that they may impose.

Pomarolli, R. S. (1966). Perceptions and attitudes of aviators toward voluntary withdrawal from
flight training. Pensacola, FL: Naval Aerospace Medical Institute.


The Edwards Personal Preference Schedule, a forced-choice personality "need" inventory, was evaluated as a predictor of success in naval flight training. The EPPS failed to discriminate between student aviators who completed training successfully and those who dropped voluntarily or failed due to poor performance. The schedule showed little promise as a predictive instrument for flight training.


The current selection tool used by the U.S. Navy and Marine Corps for its aviation officer program is a paper-and-pencil test that measures academic aptitude. There are no other tests in operational use that measure the psychomotor skills and cognitive processing skills that aviators need to fly. The Naval Aerospace Medical Research Laboratory (NAMRL) is currently evaluating the use of such a test. The purpose of this study was to establish the relationship between the paper-and-pencil Aviation Selection Test Battery that is currently being used and a Computer-Based Performance Test (CBPT). A factor analysis resulted in a four factor model accounting for 66% variance. The four factors measured by the CBPT and ASTB include tracking, quantitative skills, dichotic listening, and spatial abilities. Neither of the test batteries fully loaded on all factors. This indicates that while there are some similarities between the two test batteries (i.e. math skills, spatial apperception), there are also some important differences (e.g. dichotic listening and tracking) between the two test batteries. These differences may prove to be beneficial for future aviation selection tests.


The concept of birth order and its effects have intrigued psychologists for at least a century. Beginning with Sir Francis Galton's study of intellectual eminence in 1874 (1), many studies have shown clear differences between first- and later-born individuals on a wide variety of variables. Among these are differences in intelligence college attendance, willingness to volunteer, schizophrenia, dependency, affiliation need, and various measures of performance. One difference, found by Schacter (2) and of particular interest to aviation psychologists, was for reaction to stress. In his laboratory studies, Schacter allowed students who were waiting for what was assumed to be mild or extremely painful shock either to wait alone or with someone. He found that his subjects who were first or only children showed a greater desire for affiliation under these conditions than later borns.


The research being reported in this paper was an effort by Division No. 6 (Aviation) of the Human Resources Research Office, to describe the ways in which ineffective aviators cope with combat stress. It was thought that by focusing attention on the behavior of men judged ineffective in combat, we could begin developing measures of overall aviation combat effectiveness. In designing a preliminary study of this kind, we chose a methodology that provided rich and detailed information of the combat environment of Vietnam and aviators' reactions to it.


Air Force commissioned officer and Officer Training School (OTS) pilot training candidates who do not have a Private Pilot's License are required to complete a Flight Screening Program (FSP). FSP is a 14-hour flying program in the T-41 (Cessna 172). All student sorties are graded by the Instructor Pilot (IP). Students performing especially poorly may be eliminated before completion of the program. After 12 flying hours, students are administered Final Evaluation Flight covering the basic flying skills taught. Students failing this evaluation may repeat it one time. Students achieving a satisfactory Final Evaluation Flight grade proceed to Undergraduate Pilot Training (UPT). All other students are eliminated from the pilot training program for flying training deficiency (FTD). Thus, the program acts as a screen for entry to WT.


Optimal visual acuity is a requirement for piloting aircraft in military and civilian
settings. While acuity can be corrected with glasses, spectacle wear can limit or even prohibit use of certain devices such as night vision goggles, helmet mounted displays, and/or chemical protective masks. Although current Army policy is directed toward selection of pilots who do not require spectacle correction for acceptable vision, refractive error can become manifest over time, making optical correction necessary. In such cases, contact lenses have been used quite successfully. Another approach is to neglect small amounts of refractive error, provided that vision is at least 20/20 without correction. This report describes visual findings in an aviator who was fitted with a contact lens to correct moderate astigmatism in one eye, while the other eye, with lesser refractive error, was left uncorrected. Advanced methods of testing visual resolution, including high and low contrast visual acuity and small letter contrast sensitivity, were used to compare vision achieved with full spectacle correction to that attained with the habitual, contact lens correction. Although the patient was pleased with his habitual correction, vision was significantly better with full spectacle correction, particularly on the small letter contrast test. Implications of these findings are considered.


Determination of the proficient performance of aircraft flying tasks continues to be a subjective judgment made by instructor pilots. Current practice in training squadrons consists of "flights" during which a subset of tasks from the training syllabus are performed a varying number of times by the pilot trainee at the discretion of the instructor pilots. During or shortly after each flight, the instructor pilot "grades" the pilot trainee on the tasks performed using a standard scale but also employing his own personal criteria. While instructors differ in their personal rating bias (hard-easy), they attempt to grade in terms of "average performance at this stage of training." It is usual for the pilot trainee to be exposed to several different instructor pilots. After a specified minimum number of flights, and a recommendation by an instructor pilot, the pilot trainee is scheduled for a final "check flight. His performance on selected tasks is graded by an instructor pilot acting in the independent role of "check pilot." Should the pilot trainee not perform the flight consonant with the standards of performance expected of him by the "check pilot," he is rescheduled for additional "check flights" until he is deemed proficient.


As a review of the existing knowledge of the psychology of aviation the present paper has a number of limitations. First, it does not include the significant research that has been done in this country in the last year and a half under the auspices of the Committee on Selection and
Training of Aircraft Pilots of the National Research Council and the Civil Aeronautics Authority. Secondly, there is good reason to believe that some important work, particularly in Germany and Russia, is not being published. Thirdly, much of the Russian published material, a considerable amount of the Italian, and even some in the more accessible languages could not be examined in time for the present paper. Nevertheless, the writers have in their possession 626 abstracted references that bear upon the psychology of aviation, and 92 of these titles have been selected for mention here.


A paper-and-pencil multiple-aptitude test battery and a computer-based psychomotor test battery were administered to a sample of 354 Air Force recruits. The tests of the multiple-aptitude battery were used estimate psychometric g and to predict the psychomotor tests. The multiple correlation of the multiple-aptitude tests and each psychomotor test as a criterion was .80, corrected for range restriction. The average correlation of a psychomotor tests and psychometric g, corrected for range restriction and unreliability, was .73. The multiple-aptitude tests and the psychomotor tests were correlated and subjected to a principal components analysis. The average saturations of the psychomotor and multiple-aptitude tests were .76 and .87, respectively. Confirmatory factor analyses disclosed hierarchical general cognitive and general psychomotor factors, two lower order multiple-aptitude test factors and three lower order psychomotor test factors.


Examines the role of general cognitive ability (g) in the selection of military pilots. Presentation of the brief history of the use of G in pilot selection; Measurement of g; Awareness of the prominence of g in job performance. Examines the role of general cognitive ability (g) in the selection of military pilots. Presentation of the brief history of the use of G in pilot selection; Measurement of g; Awareness of the prominence of g in job performance.


A causal model of the role of general cognitive ability and prior job knowledge in subsequent job-knowledge acquisition and work-sample performance during training as developed. Participants were 3,428 U.S. Air Force officers in pilot training. The measures of ability and prior job knowledge came from the Air Force Officer Qualifying Test. The measures of job knowledge acquired during training were derived from classroom grades. Work-sample measures came from check flight ratings. The causal model showed that ability directly influenced the acquisition of job knowledge. General cognitive ability influenced work samples through job knowledge. Prior job knowledge had almost no influence on subsequent job knowledge but directly influenced the early work sample. Early training job knowledge influenced subsequent job knowledge and work-sample performance. Finally, early work-sample performance strongly influenced subsequent work-sample performance.


Two studies were conducted to examine the role of general and specific ability in predicting performance in military technical training. The first was a principal components analysis of the Armed Services Vocational Aptitude Battery (ASVAB); the second was a series of regression analyses using principal component scores derived from test scores as predictors and final school grades from Air Force technical training as the criterion. In the first study, 10 principal components were derived using a nation-wide representative sample of American youth. Weights derived from this analysis were used to compute principal component scores for over 78,000 subjects in Air Force technical training in 89 jobs. The first principal component was a general ability factor (g). Some specific ability components were also interpreted. The
subjects for the second study were approximately 78,000 airmen who had taken parallel forms of the ASVAB and completed technical training. Using Final School Grade as the criterion, multiple regressions were computed to determine if g was a potent predictor for all jobs and if predictive accuracy would increase if other principal components, measures of specific abilities, were added to the prediction. The regressions were computed from both uncorrected and corrected correlation matrices to properly estimate the R2 values. For each of the 89 jobs, the first principal component, g, was the most potent predictor, and for 19 of the jobs, additional principal components increased the coefficient of multiple correlation. The magnitude of the increase in R2 was estimated to be about .022 on average. Although this may seem small, practical benefits could be realized when applied to large groups of individuals such as applicants for military service.


The concept of short-term memory as a limited capacity structure is embedded within information processing approaches in psychology. As a consequence, short-term memory tasks have been employed as attention loading devices in studies of dual task performance (Logan, 1979). The assumption behind such studies is that the memory task will load the information processing system and will usurp processing capacity and/or space in the central structure. Performance on the non-memory task is examined to see if processing stages of that task draw on the same capacity or space as the memory task. This paper considers two hypotheses about how short-term memory might interact with another task in a dual task situation. When two tasks are combined, the activity of controlling and organizing performance on both tasks simultaneously may compete with either task for a resource, be that resource space in a central mechanism, general processing capacity, or some task specific resource. If there is some special relationship between short-term memory and control, especially if there is an identity relationship between short term and a central controlling mechanism, then short-term memory performance should show a decrement in a dual task situation. Even if short-term memory does not have any particular identity with a controlling mechanism, but both tasks draw on some common resource(s), then a tradeoff between the two tasks in allocation of resources is possible, and could be reflected in performance.


The authors reviewed the literature pertaining to the use of psychological testing with Air Force pilots and conclude that many of the instruments that have been used are not psychometrically sound and were used in an inappropriate manner. For example, they point out that the often used Minnesota Multiphasic Personality Inventory (MMPI) was developed over 40 years ago and normed on psychiatric patients. They advocate the use of three tests that they believe adequately measure a broad range of cognitive, personality and psychopathological domains. The first of these tests is the Multidimensional Aptitude Battery (MAB) which is a test of intellectual ability based largely on the WAIS-R. Verbal components include information comprehension, arithmetic, similarities and vocabulary. Performance measures include digit symbol coding, picture completion, spatial thinking, picture arrangement and object assembly. The Personality Research Form (PRF; Form E) measures normal personality characteristics and has 352 true-false items, organized into 21 scales. The following are some of the traits measured: achievement, aggression, autonomy, dominance, harm avoidance, impulsivity and social recognition. The Millon Clinical Multiaxial Inventory (MCMI) is a 20-scale instrument designed to measure dimensions associated with psychiatric diagnoses. The MCMI assesses eight basic personality patterns, three pathological personality patterns and nine clinical syndromes. The sample used in this study consisted of 350 white males entering Undergraduate Pilot Training (UPT). For the MAB, the subjects had an average full scale IQ of 120, with an average verbal IQ of 117 and an average performance score of 121. Most subjects scored above the MAB normative sample. For the PRF, the pilots scored higher than college students on affiliation, cognitive structure, dominance and social desirability. They scored lower than college students on abasement, autonomy, harm avoidance and understanding. The authors pointed out that this could be a result of differences in age, education, or other moderators between student pilots and college students. The MCMI pointed to histrionic and narcissistic patterns personality in the student pilots. The authors note that this is in line with the lay perception of the pilot as highly sociable and having strong self esteem.
Laboratory Aviation Personality Survey: Development, Norming, and validation. Military Medicine, 167(12), 1026-1032.

This work describes the development of a new psychological test for aviators. The Armstrong Laboratory Aviation Personality Survey was developed through the integration of clinical theory, psychometric methods, and empirical testing. It is currently given to all incoming U.S. Air Force pilot candidates. Using a sample of 6,047 student pilots, a thorough test development plan was accomplished. The 15 final test scales assess personality, psychopathology, and crew interaction styles. The scales have normative data and are demonstrated to be reliable and valid. The Armstrong Laboratory Aviation Personality Survey is recommended for use in the aviation community for both clinical and research purposes. Future research is recommended and needed in the areas of training, airframe, and special duty selection. Additional clinical work is indicated in the areas of psychiatric, psychological, and aeromedical evaluations.


Personality Survey (ALAPS). Brooks AFB, TX, Armstrong Laboratory. This work describes the development of a new psychological test for aviators. The Armstrong Laboratory Aviation Personality Survey (ALPS) was developed through the integration of clinical theory, psychometric methods, and empirical testing. Using a sample of 200 student pilots, a thorough test development plan was accomplished. The 15 final test scales cover personality, psychopathology, and crew interaction styles. The scales have normative data and are demonstrated to be reliable and valid. Additional validity work is suggested to further improve the test.


Paper-and-pencil tests have been used by the Air Force to aid in the selection of candidates for pilot training for more than forty years. The Aircrew Classification Battery (ACB) was first developed for this purpose by the Aviation Psychology Program during World War II. During the early stages of the development of the ACB, selection for pilot training was not based on predictor scores and most early classes had attrition rates of 35 to 75 percent. Thus, initial biserial correlations in the .40's and .50's were reported using success in flight training as the criterion (Flanagan, 1948). Selection into Undergraduate Pilot Training (UPT) is now based on Pilot composite scores on the Air Force Officer Qualifying Test (AFOQT). Only applicants who have successfully passed a rigorous physical examination are considered. Additionally, all UPT cadets must be graduates of a four-year college program prior to entry into UPT. Since highly restricted applicants are rank-ordered for selection based on AFOQT performance and since current UPT attrition rates are only about 15 percent, the biserial correlation has dropped to the present value of .16.


Using Cleary's (1968) model of test bias, relations between aptitude scores and training performance were evaluated for race and gender subgroups of military officer candidates. Regression analyses revealed level bias with minority performance being overpredicted by a small and constant amount at all aptitude levels, suggesting that test usage results in higher selection rates for female and Black cadets. These results are consistent with the literature in education, industry, and prior studies conducted in the military.


Joint interdependence grows out of the growing reliance of the Army on the Air Force as it becomes more agile and sheds some of its organic fires. Therefore, this research paper only addresses those areas where air and ground operations merge, on the battlefield. There are four questions addressed herein: What are the implications of joint interdependence? What are the doctrinal friction points? Where is the potential for operational seams? What might be the options for a way ahead? Organizing the services to become more interdependent makes sense operationally and strategically. Yet, experiences in Afghanistan and Iraq demonstrate that the services have much to accomplish to institutionalize joint interdependence despite the spirit of cooperation that now exists between the air and land services in both areas of operations. The simultaneous ground operations of the US Army's V Corps and the US Marine Corps' I MEF during Operation Iraqi Freedom provides a unique opportunity to evaluate the issues of joint
interdependence and propose potential solutions towards creating mutually enabling air and ground operations. This recent experience combined with the historical accounts of past air-ground cooperation provides some of the answers to the questions posed above and is also indicative of the difficulty in actually institutionalizing the organizational, training, and doctrinal changes necessary to make an interdependent land and air force. This will be hard work. Understanding the implications of creating a truly interdependent force capable of withstanding the pressures of the next inter-war period is the first step.


The main elements in the design and validation of personnel selection procedures have been in place for many years. The role of job analysis, contemporary models of work performance and criteria are reviewed critically. After identifying some important issues and reviewing research work on attracting applicants, including applicant perceptions of personnel selection processes, the research on major personnel selection methods is reviewed. Recent work on cognitive ability has confirmed the good criterion-related validity, but problems of adverse impact remain. Work on personality is progressing beyond studies designed simply to explore the criterion-related validity of personality. Interview and assessment centre research is reviewed, and recent studies indicating the key constructs measured by both are discussed. In both cases, one of the key constructs measured seems to be generally cognitive ability. Biodata validity and the processes used to develop biodata instruments are also critically reviewed. The article concludes with a critical evaluation of the processes for obtaining validity evidence (primarily from meta-analyses) and the limitations of the current state of the art. Speculative future prospects are briefly reviewed.


This paper addresses the use of officer selection and classification tests from the early 1940s to 1986, with emphasis on the evolution of the Air Force Officer Qualifying Test (AFOQT). It is intended as a readable historical overview of officer testing, not as a detailed technical document. The paper emphasizes AFOQT test forms and content as they relate to each other, up to and including the present operational version, Form 0.


This retrospective account of the emergence of engineering psychologists – in the military, in academia, in the aviation industry, in troubleshooting system problems, in consulting, and in course setting for civil and military agencies – is based largely on my recollections and many years of correspondence with others of similar vintage or older.


The Army Fixed-Wing Aptitude Battery (AFWAB) has been used operationally since 1956 to select trainees for the Army Fixed-Wing Flight Training Program. The present study was undertaken to evaluate the battery for use in selecting trainees for the ROTC Flight Training Program. Additional purposes were to provide information which could be used to establish cutting scores appropriate to the Army's ROTC flight training requirements for a given year and to study the effect of weighting the tests by a multiple correlation procedure. The battery was administered to samples of ROTC Flight Training Program applicants representative of ROTC classes of academic years 1956-57, 1957-58, and 1958-59 (total N = 1245). The AFWAB was found to have useful validity against a criterion of successful completion of the Flight Training Program (r = .33). All component tests were found to contribute to the selective efficiency of the battery (correlation coefficients ranged from .20 to .24, intercorrelation coefficients from .11 to .52). The unit-weighted composite proved to be as effective as the administratively more cumbersome optimally weighted score.

Association Conference, San Antonio, TX.


Downsizing, declining military budgets, and changing world conditions have combined to create a growing need for the military services to coordinate their research efforts. Responding to this need, coordinate manpower, they reached an agreement in 1990 to personnel and training research through a committee known as TAPSTEM. Selection and classification was identified as a major component of manpower and personnel research, and a Selection and Classification Subcommittee was established to assist the Manpower and Personnel working group in addressing this area. This year, the subcommittee was charged with developing a long-range joint service selection and classification research plan. This paper describes this plan.


The purpose of the work was to develop a highly flexible psychomotor testing system capable of reproducing the psychological task structure of two electromechanical tests used earlier in Air Force pilot selection programs. These were the SAM Complex Coordination Test and the SAM Two-Hand Coordination Test. The work was conducted in two phases, the first of which resulted in the definition, design, assembly, and testing of the psychomotor testing system. The second phase involved the testing of 120 Air Force pilot candidates and analysis of the data. The system developed to implement these tests consists of two test stations (expandable to eight) and a test control unit. Test control station functions are performed with a PDP-8/L digital computer which can generate graphical, alphanumeric, or point displays on a direct-view storage tube. The feasibility of this psychomotor testing system was demonstrated and found to be highly flexible and efficient, with a capability for conducting test sessions under automated conditions.

The authors examined the utility of a decision-making task and various personality variables for discriminating between pilots who had been involved in an accident and those who had not. Subjects were 51 volunteer military aviators with ranks ranging from Chief Warrant Officer-2 to Lieutenant Colonel. Prior accident involvement was determined through an investigation of the United States Army Agency for Aviation Safety (USAAAVS) accident records. Each aviator listed as a causal factor in at least one aviation accident (either major, minor, or incident) was classified as pilot-error accident involved (PEAI), otherwise, they were classified as pilot-error accident free (PEAF). Several measures were administered to all of these aviators. The first was the 16 PF (Form A), a personality inventory, which consists of 16 primary factors and four secondary factors. The second inventory was the Mehrabian Achievement Scale which provides an indication of need for achievement or desire to attain success. A decision-making task was also administered which involved having the subject decide when to leave a light on or turn it off, based on a set of rules. Subjects' scores were the means of each of the various instruments taken over several different trials. The first analysis included the measures from the 16 PF and the N-Ach score from the Mehrabian. Three of the 21 scores discriminated between the PEM and the PEAF aviators: (1) Group Dependent vs. Self Sufficient; (2) Practical vs. Imaginative; and (3) Forthright vs. Shrewd. These three scores correctly classify 86 percent of the sample into either the PEAT or the PEAF groups. The second stepwise discriminant analysis included the scores from the decision-making task. Again, none of these scores successfully discriminated between the two groups of aviators. This study was cross-validated in a later paper (Sanders, Hoffmann, & Neese, 1976) and the results did not replicate.


The authors conducted a cross-validation study in an attempt to replicate the results obtained by Sanders and Hoffman (1975). The sample consisted of 66 military aviators. Once again the 16 PF (Form A) was administered. Pilots were classified into one of two groups, either pilot-error accident involved (PEAI) or pilot-error accident free (PEAF) based on United States Army Agency for Aviation Safety (USAAAVS) accident records. The stepwise discriminant analysis did not discriminate between PEAI and PEAF aviators. A second stepwise discriminant analysis was performed using age, total military flight hours and years of flight status. None of these variables discriminated between PEAF and PEN pilots. The authors conclude that individual differences in personality characteristics of aviators prevented the identification of personality traits associated with the PEAI and the PEAF groups.


Unmanned/uninhabited aerial vehicles (UAVs) are an increasingly important part of military operations throughout the world. However there is no consensus about who should fly these aircraft. The United States Air Force (USAF) Corona South four-star general officer Summit in 1997 resulted in tasking the Air Force Research Laboratory to conduct a study to compare the speed and accuracy with which various groups of pilots could learn to fly the RQ-1A Predator UAV. This study primarily addressed stick-and-rudder skills; we did not measure such operationally relevant factors as communication skills, command experience, or knowledge of combat operations. Seven groups of military and civilian pilots, varying in amount and kind of flying experience, completed a series of multimedia tutorials on principles of flight and procedures for operating the Predator, then flew a high-fidelity RQ-IA simulator. Each participant flew basic maneuvers and landings (including difficult crosswind landings) until a very high standard of aircraft control performance was achieved, then flew 30 reconnaissance scenarios. During this time, detailed measures of performance were continuously and automatically recorded. The results show that, though Predator pilots performed best (and nonpilots performed worst), USAF T-38 graduates and civilian pilots with single-engine instrument training performed nearly as well as a group of highly experienced military pilots assigned (but not yet trained) to fly Predator. A possible explanation for the relatively good performance of the T-38 and civilian instrument pilots is that there may be advantages to recent experience flying aircraft that have handling characteristics that are similar to the Predator.


An analysis was performed to identify specific skills required to successfully perform operations and unit maintenance tasks for the future tactical unmanned vehicle (TUV) and to determine if U.S. Army soldiers and U.S. marines with a skill level of IO have those skills. This analysis was performed by the Human Research and Engineering Directorate of the U.S. Army Research Laboratory at the behest of the Program Manager Unmanned Ground Vehicles/Systems. Military occupational specialties examined included U.S. Army infantryman (11B), cavalry scout (19D), and the Marine Corps rifleman (0300). System-required operations and unit maintenance functions and tasks were identified. Soldier-marine operations and unit maintenance skills were compared to these tasks. Results of the analysis show that of 209 operations skills required by the TUV system, 82 were mismatched because of a higher skills requirement, untrained system-specific skills, or a combination of both. Additionally, all 25 unit maintenance tasks were identified as requiring system-specific training.


It has long been the belief of the writer of these notes that the functions of certain branches of the service auxiliary to the line in the United States army are not in general well understood even by people who are interested in military affairs. The reasons for this are, no doubt, in part due to the fact that the functions of these auxiliaries are many and varied, and are not outlined in available form or even clearly defined except in the brief and general statements of laws, regulations, or orders affecting the service. This lack of knowledge seems to be especially true of the corps with which the writer has long had the honor to serve, the name of which gives no indication what-ever of its functions, scope, and value to the army. It is thought, therefore, that an outline of the duties and field of usefulness of the signal corps of the army may be of value to those interested in military affairs, to others upon whom the army must depend for its maintenance, and to officers and men of the national guard who will be called upon to perform the duties of signalmen. To instructors and students at military schools, to officers of the regular army who may be called by detail or by accident of service to construct and maintain lines of information, and to the great mass of the volunteers if called to the defense of the country, these notes may also be of interest and use. For these reasons they have been prepared. The writer begs to acknowledge his indebtedness to Lieut. Col. Samuel Reber, Maj. Edgar Russel, Capts. Charles S. Wallace, George S. Gibbs, and G. Soulard Turner, of the signal corps, and Mr. William M. Reading, of the signal office, for their assistance in the preparation of these notes.


The Army Research Institute (ARI) was directed by the Deputy Chief of Staff for
Personnel to develop and validate motivational measures to aid in selection of enlisted personnel. The development of the Military Aptitude Predictor (MAP) was begun in 1973 for this purpose. The initial MAP included the trainee's age, education, aptitudes, and civil court convictions (moral waiver) as predictors of military performance, and was intended for use at the Armed Forces Examining and Entrance Stations (AFEES). The 1975 version (MAP-75) was intended for use by recruiters to provide a prediction of Basic Combat Training (BCT) behavior and performance at an earlier point in time, thus saving the Army AFEES transportation and processing costs. In April 1975, the Secretary of the Army directed implementation of the MAP to begin 1 August 1975. This initial testing was conducted in one District Recruiting Command (DRC) in each of the five Recruiting Regions during the period 1 August - 30 September 1975. Its operational use was suspended on 1 October 1975 and, immediately following suspension of testing, ARI conducted interviews in the five DRC's to determine the nature and extent of problems associated with its use.


In 1991, a new psychological selection procedure was employed during the selection of Japanese Space Station astronauts. It was based on international selection criteria developed by an international psychological/psychiatric working group. A total of 372 individuals--fewer than expected--submitted applications for Space Station astronaut in Japan. Of the applicants, 233 were given several psychological written tests [Anxiety Scale, Performance Test, General Aptitude Test Battery, and Environmental Adjustment Test Battery (EATB)] in Phase I of the selection. Forty-five applicants went on to take the General Aptitude Test Battery (GATB), Human Assessment Method (a group test), a semi-structured psychological interview and an intelligence test in Phase II of the selection. All applicants were found to be highly intelligent. Interestingly, an unexpectedly large number of candidates were disqualified by the newly developed EATB. Assessment of individual functioning in a group (the Human Assessment Method) resulted in no applicant being ranked in either the "qualified with reservation" (QR) or "disqualified" (DQ) category. Much has been learned from this initial application of psychological "select-in" testing, but further efforts are needed to improve both psychological criteria and evaluation methods and to determine their reliability and validity.


This was the third progress report on the development of a personality battery to supplement present aircrew selection techniques. The goal of the program was to identify prospective aircrew trainees who were predisposed to difficulties in adapting to the rigors of military flying. Predictor measures included a wide variety of aptitude, personality, perceptual,
attitude and psychomotor variables. Personnel information from Personnel File Form 66 and the 201 File concerning rate and extent of promotion, type of assignment, extent of command responsibility and flying duties involved were gathered to use as criteria. Psychologists made adjustment ratings, one of which was a purified pass/fail criterion. The high group included those who graduated from training and were well-adjusted. The low group consisted of pilots who failed training for reasons of poor motivation, excessive emotional reaction, or overt symptoms attributable to stress in the program. The following are the results obtained when correlating the different predictors with the various criteria. The Aviation Interest Key is comprised of 25 items covering the attitudes of parents toward the trainee's participation in hazardous sports, a history of motion sickness, prior military flying experience, etc. It yielded correlations of .37 to .41 with the purified pass/fail criterion controlling for pilot stanine. Seven Minnesota Multiphasic Personality Inventory (MMPI) scales (Hs, Pd, Winne, Taylor, Seaquist, D and Hy) correlated from .10 to .40 (mean = .24) with the purified pass/fail criterion measure. The Pilot Opinionaire used an indirect polling approach to assess attitudes toward various aspects of military aviation (and also included an authoritarianism scale). Correlations were found between it and the pass/fail criterion that ranged from .28 for cadets to .11 for officers. The author also demonstrated that even though aptitude measures were much more highly correlated with the pass/fail criterion (r = .53), they correlated only .13 with information from Form 66 (which involves job performance). Whereas, the correlations of the personality variables remained relatively constant, albeit low (r = -1.1). At this point in time, only four screening tests have been validated against post-training criteria, the Personal Inventory (PI), the Cornell Index (CI), the Cornell Word Form (CWF) and the Sentence Completion (SC) Factor scores. The correlations of these tests with various personnel file form (Form 66) information ranged from -.07 (for Interpersonal Attitudes) to .23 (for Self-Enhancement).


The author characterizes the qualifications for jobs in combat aircrews as encompassing three areas: (1) physical qualifications; (2) abilities and aptitudes; and (3) personality factors. He notes that the USAF has developed thorough and effective standards for selection procedures for the first two areas, but not for the third area. This report summarized the progress to date on a large scale developmental project (started in 1949) to update the USAFs use of personality factors in selection. The present research included only pilots. Subjects were administered a wide range of paper and pencil, projective, performance and apparatus tests. Students entering flying school were not told that they could be eliminated based on the results of the test, but they were led to believe that the tests would be used for administrative decisions (i.e., selection). Validation of the tests were in progress (or planned) at the time of this article was written for three stages in the students careers: (1) training; (2) post-training and operational experience; and (3) combat performance (based on data collected in Korea). The results indicated that ratings made by classmates and by instructors during training were superior to psychologists' evaluations in predicting combat performance. However, a significant positive correlation was found between performance ratings in combat and an absence of pathological behavior symptoms in the clinical reports made by a field survey team. Also, pilots with superior adjustment ratings in training and higher pilot stanines tended to have more accidents. At the time this article was written, all studies had been conducted using training level criteria, which relies on a pass/fail criterion. The author advocates using "purified" pass/fail measures. For example, one approach is to classify
the population into more refined administrative categories according to their performance in training. Finally, the author stated that there are three broad, interrelated areas in which efforts should be concentrated: (1) specific motivational structure; (2) character integration; and (3) tolerance of frustration and anxiety.


A field tryout of an experimental test battery for adaptability screening of flying personnel was begun in September, 1956, by the U.S. Air Force School of Aviation Medicine at Lackland Air Force Base, Texas. This battery consisted of seven paper and pencil, machine scored, personality and motivational tests which had been developed and validated in previous research.1-6.10 These tests have been administered, under operational testing conditions to 9,500 students entering pre-flight schools for aircrew training as part of their entrance physical-examination. The Field Testing Laboratory of the Department of Medical Psychology of this school, carried out this project, under the direction of one of us (R.C.,T.).


There are a variety of ongoing attempts to generate unmanned aerial vehicle (UAV) technologies to exploit the advantages that these semiautomated and automated airborne platforms promise to render. (Although we refer specifically to UAVs here, our arguments apply, in principle, to all remote vehicles whatever their medium of operation. The principles themselves also extend to other forms of nontransport-based entities.) With regard to such operations, the collective community is searching for the ratio between operator(s) and vehicle(s) that will prove most efficient and effective.


The reduction of military pilot training attrition rates is a central issue for any successful and cost-effective selection procedure. With this object in mind, a validation study was designed using the "Pilot Aptitude Tester" (PILAPT) together with other assessments used by the Italian
Air Force (IAF). The main hypothesis was that the combined use of traditional and computer-based techniques would reduce false positive rates in selecting candidates admitted to flight training.


A computer-based Empirically Validated Task Analysis (EVTA) of Canadian Forces light observation helicopter operations was conducted from video records of cockpit activity gathered during flight. The task analysis was performed in order to provide data for function analysis and work-load prediction studies in support of the Canadian Forces Light Helicopter replacement project. Observable behaviors were categorized according to the type of activity involved and communications were analysed for content, agencies involved, and relevance to the crew's task. The results of this study indicate that data gathered from a controlled test environment can differ considerably from those obtained in operational settings and that miniature video cameras can be useful in obtaining information from environments which hitherto may have been inaccessible to all but operational personnel.


The purpose of this document is to provide a critical review of the literature on situation awareness and to determine whether current situation awareness measurement techniques are adequate to be used as a part of personnel selection tests. In order to determine if this is the case, the discussions in this document are centered on (a) providing a working definition of situation awareness; (b) understanding the various situation awareness models; (c) evaluating current situation awareness measurements; (d) identifying key domains where individual differences in situation awareness matter; and (e) identifying correlates of situation awareness levels. It is concluded that the current measurement tools may not be sensitive enough to be used for selection purposes; therefore development of a new, job task based, selection battery is recommended.


Laboratory.


Shephard, A. H. (1950). Losses of skill in performing the standard Mashburn task arising from different levels of learning on the reversed task: Office of Naval Research.


Shipley, B. D. (1984). Productivity and difficulty as new criteria for validating aviator selection tests. Paper presented at the Human Factors and Ergonomics Society Annual Meeting. The US Army Research Institute is conducting research to improve the quality of the Army's aviator selection testing program. The research is motivated by increasing costs of training and by changing aviator ability requirements due to advanced aircraft and modern tactics. This paper describes the development of a new criterion variable to support the testing improvement research.


presented at the 33rd Annual Meeting of the Human Factors Society.

In this paper, the relationships between a Risk Test, the Pilot Personality Questionnaire (PPQ), and a pass/fail training criterion and academic grades were examined. The authors also examined the relationships between the two predictor measures and the current U.S. Navy/Marine Corps aviation selection battery, the Academic Qualifications Test flight Aptitude Rating (AQTIFAR). Subjects were 407 student naval aviators (SNAs) and 182 student naval flight officers (SNFOs) who took the Risk Test and 110 SNAs and 114 SNFOs who took the PPQ. In the Risk Test, subjects are asked to choose which of ten squares is the reward square accumulating points as they go. The average number of responses (NR) and the corresponding average reaction time (RT) are measured for each trial. The PPQ is a self-administered personality inventory containing 112 multiple-choice items, which is a combination of 4 different personality tests: Locus of Control, Work and Family Orientation, Personality Attributes Questionnaire and a Social Desirability Scale. The AQT is a general cognitive ability measure and the FAR is comprised of the Mechanical Comprehension Test, the Spatial Apperception Test and the Biographical Inventory. Results showed that for the SNAs, the number of responses on the Risk Test correlated .13 (p < .05) with Undergraduate Pilot Training (UPT) pass/fail criterion and -.13 (p < .05) with academic grades. Reaction time correlated -.18 (p < .05) with UPT pass/fail. For the PPQ, the Social Desirability Scale correlated -.30 (p < .05) with UPT pass/fail and -.45 (p < .05) with academic grades. For the SNFOs, the number of responses on the gambling risk test correlated .28 (p < .05) with UPT pass/fail, while reaction time correlated -.45 (p < .05) with UPT pass/fail. Neither the number of responses nor the reaction time variables from the Risk Test were correlated with academic grades. For the PPQ, high self control correlated .40 (p < .05) with UPT pass/fail, but no other scales were correlated with this criterion. For the academic grades criterion, PPQ aggressiveness correlated .54 (p < .05), high competitiveness correlated -.52 (p < .05) and submissiveness correlated -.55 (p < .05).

The authors concluded that those trainees who exhibited more risk-taking behavior were more likely to complete flight training. However, the experienced pilots showed less risk-taking behavior than the SNAs. This may mean that although in training those that are successful are more likely to take risks, once they become pilots they demonstrate more cautious behavior (i.e., take more calculated risks where they can assess the likelihood of failing).


Current aircrew selection research at the Naval Aerospace Medical Research Laboratory has focused primarily on psychomotor and cognitive abilities. Evidence from studies on flight training attrition suggests that a number of failures may be attributed to personality or motivational factors rather than a lack of abilities. Because flight training success is a dynamic interaction of abilities, motivation, and personality factors, all three areas should be included to optimize the predictive validity of aircrew selection batteries. Two sets of data are presented; one set is from a computer based risk assessment task, and the other is from the Jenkins Activity Survey. The data indicated few relationships between risk assessment measures and flight training criteria. We found only one indication that increased risk taking was associated with successfully completing primary flight training. The Jenkins Activity Survey results indicated contradictory relationships between the scale measures and flight training criteria in the few significant findings observed.

Research is currently being conducted to develop reliable predictor tests which might aid in the making of decisions concerning aircrew assignment. The current approach involves comparing the performance of several different aviator communities on a test battery measuring various aspects of cognitive and psychomotor functioning. Four groups of pilots (two experienced jet groups, one experienced helicopter (helo) group, and one untrained student group) were tested on this battery. Overall, the jet groups performed in an equivalent manner, while the helo group showed a lower test performance level in comparison. The student group performed at a lower level than the experienced groups in general. Within this student group, pilot trainees who were assigned to the jet pipeline did significantly better on several of these tests than those trainees who were assigned to either helicopter or land-based fixed-wing pipelines. Many of the test performance differences seen between these jet and helo student pipeline groups were also seen between the experienced jet and helo pilots tested. Shull, R. N. and G. R. Griffin (1990). Predicting F-14 air combat maneuvering (ACM) performance using an automated battery of cognitive/psychomotor tests. Naval Air Station Pensacola, FL, Naval Aerospace Medical Research Laboratory. Some studies have suggested the possibility of predicting operational performance in fleet aviation environments. The current report concerns the use of an automated performance-based test battery, involving cognitive and psychomotor functioning, to predict the operational performance of fighter pilots. A group of jet pilots completing Air Combat Maneuvering (ACH) training in the F-14 were tested on this battery. The few significant correlations found between the test measures and ACM performance measures were illogically patterned and of insufficient quantity or strength to establish that such a battery would reliably predict ACM performance. This could have been due to the homogeneous nature of the subject group in terms of pilot skills and abilities. Given these results, this particular test battery would probably not be useful in the prediction of flight performance at such a late stage of training as ACM, but it might predict flight performance in earlier training.

Shull, R. N., & Griffin, G. R. (1990). Predicting F-14 air combat maneuvering (ACM) performance using an automated battery of cognitive/psychomotor tests. Naval Air Station Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Some studies have suggested the possibility of predicting operational performance in fleet aviation environments. The current report concerns the use of an automated performance-based test battery, involving cognitive and psychomotor functioning, to predict the operational performance of fighter pilots. A group of jet pilots completing Air Combat Maneuvering (ACH) training in the F-14 were tested on this battery. The few significant correlations found between the test measures and ACM performance measures were illogically patterned and of insufficient quantity or strength to establish that such a battery would reliably predict ACM performance. This could have been due to the homogeneous nature of the subject group in terms of pilot skills and abilities. Given these results, this particular test battery would probably not be useful in the prediction of flight performance at such a late stage of training as ACM, but it might predict flight performance in earlier training.

Abstracts International, 36, 4942A.


The authors state that until recently, USAF pilots had been assigned to either fighter aircraft or tanker-transport aircraft based upon the recommendation of the advanced training recommendation board (ATRB). The purpose of this study was to develop a model to predict which USAF pilot candidates would be considered by the ATRB to be suitable for assignment to fighter aircraft using scores from the Air Force Officer Qualifying Test (AFOQT), the Basic Attributes Tests (BAT), college GPA and an index of candidates’ preference for tanker-transport aircraft. A sample of 426 Officer Training School candidates were randomly assigned to one of two subsamples. For each subsample, multiple correlations were computed between the 42 variables and the ATRB recommendation outcome. These regression models were then each cross-validated in the other subsample. The full models for sample A and sample B were both significant (R = .61, p < .001; R = .50, p < .05, respectively). For each subsample, multiple regression was used to eliminate variables that did not contribute significantly to the prediction of ATRB outcome. For sample A, the final reduced model contained two predictors: AFOQT Instrument Comprehension and Response Time (RT) from the BAT Mental Rotation task (R = .31, p < .01). For sample B, the reduced model contained seven predictors: AFOQT Instrument Comprehension and Block Counting; the BAT percent correct and RT variability from the Mental Rotation test and average tracking difficulty from the Time Sharing test; college GPA; and the desire to fly tanker-transport rating. For the cross-validity results, the model developed
for sample A, had a multiple correlation of -22 with ATRB recommendation outcome in subsample B. The model developed in sample B, had a multiple correlation of .26 with ATRB recommendation outcome in sample A. Based on these results a final model was developed that produced a multiple correlation of .40 @ < .001).


To examine the utility of personality testing for enhancing current Air Force pilot selection procedures, a sample of 509 USAF officers was given a computer-administered personality inventory, the Automated Aircrew Personality Profiler (AAPP) prior to entry into Undergraduate Pilot Training (UPT). Factor analysis of 16 scale scores indicated that the inventory comprised measures of five personality characteristics, of which three were directly associated with UPT training outcome (pass or fail): Self-confidence, Values Flexibility, and Hostility. UPT graduates scored higher on both positive dimensions and lower on hostility than did those individuals eliminated for flying training deficiency. The AAPP failed to add predictive utility to a selection model that combined test scores from the Air Force Officer Qualifying Test (AFOQT) and the Basic Attributes Tests (BAT) battery.


To examine the utility of personality testing for enhancing current Air Force pilot selection procedures, a sample of 509 USAF officers was given a computer-administered personality inventory, the Automated Aircrew Personality Profiler (AAPP) prior to entry into Undergraduate Pilot Training (UPT). Factor analysis of 16 scale scores indicated that the inventory comprised measures of five personality characteristics, of which three were directly associated with UPT training outcome (pass or fail): Self-confidence, Values Flexibility, and Hostility. UPT graduates scored higher on both positive dimensions and lower on hostility than did those individuals eliminated for flying training deficiency. The AAPP failed to add predictive utility to a selection model that combined test scores from the Air Force Officer Qualifying Test (AFOQT) and the Basic Attributes Tests (BAT) battery.


In this paper, past research (e.g., Holden & Fekken, 1987) is discussed showing that individuals who scored high on a personality trait manifested shorter response latencies than individuals who scored low on that trait on items that were "endorsed." The opposite was found for items that were "rejected." Some interpret these results in terms of a "self-schema," which is an integrated network of self knowledge. Five hundred and nine Air Force student pilots entering Undergraduate Pilot Training (UPT; 332 for which training outcome was available) participated in this study. The Automated Aircrew Personality Profiler (AAPP) was administered and includes 94 items from the Minnesota Multiphasic Personality Inventory (MMPI). Scores were computed on the following scales: Sociability (Psychoticism), Emotional Stability (Neuroticism), Extraversion, Competency (Inadequacy) and Cynicism. The criterion used was UPT training outcome (pass/fail). Response latencies were standardized within subjects to control for reading speed and by item to control for item length and vocabulary level. For each subject and each
scale, a mean response time for endorsed items and one for rejected items was computed. Correlations between personality scores and UPT outcome were fairly weak. The only significant correlations with UPT were for Sociability \((r = .13, p < .01)\) and Cynicism \((r = -.14, p < .01)\). Correlations between response time scores with the criterion were only significant for endorsed items and then only for Extraversion and Cynicism. The author noted that the internal consistency for the sociability scale was only .34. Correlations between scale and response time variables for each trait were relatively small (ranging from .02 to .35). Correlations between the two response time measures (i.e., endorsed and rejected) for each trait were also small (ranging from -.01 to -.26).


Past research (e.g., Siem, 1991) has indicated that personality measures based on inventory response latencies, as well as scale scores, have validity for predicting pilot training performance. The purpose of this study was to examine the incremental validities of these measures in predicting training performance. Specifically, two issues were addressed: (1) the factor structure of the scale scores and response latency scores were examined, with the expectation that the measures of the same trait would load on the same factor; and (2) the extent to which these personality measures would contribute unique variance to the prediction of training outcome was examined. Three hundred thirty-two college graduates who had been selected for Air Force pilot training were included in the factor analyses, 277 of these students were subsequently included in the multiple regression analyses. Scores from three predictors were used in this study: (1) the Basic Attributes Tests (BAT) - accuracy and response time scores were collected for the tests and psychomotor tasks; (2) the Automated Aircrew Personality Profiler (AAPP) - data on item endorsement and response time were collected for each item; and (3) the Air Force Officer Qualifying Tests (AFOQT). Undergraduate Pilot Training (UPT) was dichotomously scored pass/fail. Results of the factor analysis related to the first hypothesis showed that one factor was defined by the scale scores for the Emotional Stability, Sociability and Cynicism dimensions. Two factors were defined by loadings both of a scale score and a latency score (Extraversion & Cynicism). The remaining two factors were defined by loadings of latency variables only for Cynicism and Emotional Stability. Results related to hypothesis 2 showed that the full model correlated .36 \((p < .01)\) with UPT pass/fail, the model with the five AAPP variables removed resulted in a R-change of .06 \((p < .05)\). Thus, the personality variables did add to the prediction of UPT outcome beyond the AFOQT and the BAT. However, further analyses showed that all but one of the five AAPP variables, namely, Extraversion, could be removed from the model without a decrease in validity.


Assesses the predictive utility contribution of personality measures on the pilot selection procedure of the U.S. Air Force. Demonstration of a direct relation between hostility, self-confidence and values flexibility to training outcome; Negative enhancement of the predictive validity of the selection system; Correlation between personality and aviation performance. Assesses the predictive utility contribution of personality measures on the pilot selection procedure of the U.S. Air Force. Demonstration of a direct relation between hostility, self-
Developments in research concerning personality characteristics have led to a renewed interest in applications of individual differences measures for selection of pilot candidates. Recent research efforts have focused on selecting for positive characteristics, rather than screening out pathological traits. Another development is the use of tests in which the dimension of interest is not readily apparent to the examinee. In the present study, five personality and attitude measures were administered to 883 USAF pilot candidates as part of an experimental test battery under consideration for operational use in pilot selection and classification. These tests were designed to assess decisiveness, risk-taking, self-confidence, survival attitudes, and field dependence/independence. Scores from these tests were examined for their utility in predicting training outcome (graduation or elimination) and advanced training recommendation (fighter or non-fighter aircraft). Results indicated that as a group, the tests demonstrated weak relationships with the performance criteria. No test manifested a consistent pattern of validity for both performance measures. Only the test of self-confidence appeared to contribute to predicting completion of training. Future research efforts are discussed with regard to refining the current test of self-confidence and establishing its construct validity.


Employment interviews are widely used in industry and the military for personnel selection and classification, although the scientific evidence for their validity and reliability is somewhat weak (Arvey & Campion, 1982; Harkness, 1987). The U.S. Air Force does not currently use a routine interview process for pilot selection, other than that conducted for recruiting purposes, although some components of the Air Force do rely on their own procedures. Units of the Air National Guard, for instance, have used an interview process with
some success (Armour, personal communication, October 10, 1987). Based on the perceived success of an interview process used by the Air National Guard, a study was undertaken to assess the validity of an interview for selecting Air Force pilot candidates. In particular, the issue of interest was whether instructor pilots could use a structured interview technique to assess an applicant's potential for pilot training. The content of the interview was derived from a review of the instrument used by the Air National Guard and through discussions with subject matter experts of the Air Force Air Training Command (ATC), the organization responsible for training all Air Force, Air National Guard and Air Force Reserve pilot candidates.


Studies based on pilot performance in combat suggest that individual differences in personality characteristics are important to effective performance (Bair, 1952; Jenkins, Ewti, & &roll, 1950; Stanley, 1973; Youngling, Levine, Mochamuk, & Weston, 1977). Nonetheless, qualitative reviews of the literature have generally been pessimistic toward empirical evidence supporting the hypothesis that personality characteristics are important to effective pilot performance, despite the large number of characteristics examined using a variety of different instruments in a multitude of studies (Dolgin & Gibb, 1988). One possible explanation for the failure to find stronger relationships may be due to the lack of a suitable conceptual framework for evaluating results from different studies. Efforts to identify such a framework have led to a consensus model of personality based on the observation that five global factors adequately describe individual differences in personality traits (Digman, 1990; Tupes & Christal, 1961). These factors, known as the “Big Five,” are (I) Extroversion, (II) Agreeableness, (III) Conscientiousness, (IV) Emotional Stability, and (V) Culture, or Openness to Experience (Costa & McCrae, 1992).


This article presents validity evidence for the pilot selection battery employed by the Royal Canadian Air Force (RCAF). This battery includes a variety of predictor variables ranging from cognitive ability to biodata to psychomotor coordination. Several different criterion measures were also used ranging from a training pass/fail criterion to various flight training and ground school ranks. Results showed that the RCAF Visual Link Test was the most valid predictor variable (zero-order validities ranging from .18 for initial training school rank to -57 for elementary flying training school rank). The validities for the Aircrew Information Sheet (biodata) ranged from .01 for elementary and service flight training school to .20 for initial training school rank. The validities for the Aircrew Interview Report Form (motivation and
attitude appraisal) ranged from .06 for elementary flight training school pass/fail to .16 for initial training school rank. The authors conclude that the RCAF battery compares favorably with the 1943 U.S. Air Force selection battery.

Silva, J. M. (1997). Using Psychomotor Ability for Selecting TOW Gunners. (pp. 30). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. The situation awareness (SA) and mental workload of 56 subjects were evaluated as they monitored one or more attributes of six objects moving systematically over a rectangular grid. Subjects were assigned to one of seven groups depending upon whether they were to monitor object locations (location task), object colors (color task), whether the objects flashed (flash task), or some combination of these three. Both task performance and subjective ratings were used to assess subjects' awareness of the three object attributes. In addition, subjective ratings of mental workload were collected. All subjects performed the monitoring task under four different conditions formed from the factorial combination of 1) the probability that objects of a certain color would flash and 2) whether object colors remained consistent or changed during the course of a trial. The results pointed to the usefulness of both flash and color task performance as measures of SA. Subjects were very poor at the location task, suggesting either their location awareness was poor or the location task is not a good measure of that awareness. Subjective ratings proved useful but occasionally dissociated from task performance. One possibility is that subjective ratings reflect rational inferences by the subjects rather than the outcome of their introspections.

Simon, E., Watts, D., & Bohnker, B. K. (2008). Helicopter mishap attributed to single seizure. Military Medicine, 173(3), 322-323. A case report is presented of a 36-year-old U.S. Coast Guard aviator who had a single seizure while operating a helicopter on the ground. His seizure activity produced a loss of consciousness during which he pushed the cyclic to the left anterior quadrant that resulted in a ground mishap. No risk factors were identified in an extensive neurological workup. The current guidance for handling seizures in military aviation personnel is reviewed, along with considerations for treatment. Although the military aviation selection process carefully screens applicants for seizure history and potential, occasional seizures in the aviation population remain possible. Such events may result in military aircraft mishaps despite careful risk factor surveillance, as demonstrated by this case.


occupational question of whether this pilot is like other pilots and is safe to fly and able to complete the mission. The second is the clinical question of whether this pilot has a diagnostic and treatable psychiatric condition or disorder that immediately excludes the pilot from flying until treatment is successfully completed. The MMPI may be helpful in answering these two questions, but a normative sample of pilots has been lacking in the past. Given these selection and continuation standards, military pilots are expected to have the same general MMPI profile across time and military services. It was hypothesized that "normal" (not having received any psychiatric diagnosis) pilot groups from the US Air Force, Army and Navy have MMPI profiles not significantly different from one another regardless of the passage of time. A second hypothesis was that groups of military pilots would be significantly different from the civilian norms on at least eight scales: K-correction (K), Hypochondriasis (Hs), Hysteria (Hy), Psychopathic Deviate (Pd), Psychasthenia (Pt), Schizophrenia (SC), Hypomania (Ma), and Social Introversion (Si) as shown in a previous study (Fulkerson, Freud and Raynor, 1958).


Air Traffic & Air Weapons Control offer highly complex, critical entry-level jobs AEC relies upon the same generic system use to select General Service Officer classifications & NCM occupations Training serials filled on 'first-passed-the-post' basis Only difference for AEC selection is interim CFAT Total cut-off score at 60th percentile AEC has experienced significant training attrition rate since 2000 (e.g., 1999 - 60%, now averaging 35%) Current selection criteria unable to further decrease training attrition aim To determine predictive validity of RAFAAT & TSD-PI in Aerospace Control training & job performance.


Net-centric warfare and interoperability are fast becoming basic tenets of modern military strategic thought. The Canadian Forces and its NATO allies are currently conducting research into the effective use of current and emerging technologies such as airborne sensors and uninhabited aerospace vehicles (UAVs) to enhance their intelligence, surveillance, and reconnaissance (ISR) capabilities. Effective sensor operation is critical to the successful support of UAVs to Canada’s joint and combined net-centric warfare capability. The selection, training, and employment of Canadian Forces personnel as sensor operators will depend upon an accurate analysis of this position’s requirements and upon the determination of whom among us has the appropriate training and experience to competently fill this vital ISR position. Canadian Forces UAV experimentation is developing an understanding of the generic task and knowledge requirements of the Medium Long Endurance (MALE) UAV Sensor Operator position to that end. This paper discusses the methods and techniques used over the course of three major research events to determine the position and personnel selection criteria for MALE UAV Sensor
Operators and provide preliminary results from Canadian Forces research to date.


The 2010 Pilot Source Study, commissioned to research the success of pilots in initial training for Part 121 operations, analyzed the training performance of 2,156 new-hire pilots in the years 2005-2009. Six regional airlines provided data that was mined from human resource and pilot training files. Five university researchers independently analyzed the data and integrated their results. The study expressed success in terms of fewer extra training events and fewer non-completions in regional airline training. Statistically, the best performing pilots were those who had flight instructor certificates, graduated from collegiate accredited flight programs, received advanced (post-Private) pilot training in college, graduated with collegiate aviation degrees (any aviation discipline), and had between 500 and 1,000 pre-employment flight hours. Pilot source characteristics that had no significance in regional airline pilot training success were: having a non-aviation college degree and having prior corporate pilot or airline pilot experience.


Devices such as the 16PF and MMPI have been widely employed in the evaluation of personnel in aviation settings. The present study investigated the problem of item ambiguity (the degree to which an item elicits multiple interpretation) which may limit the utility of such devices when used in screening procedures. Subjects complete either the 16PF or the MMPI while concurrently rating each item on a five-point ambiguity scale. The ambiguity for each item was determined and the relationship between ambiguity and sex of the respondent, the individual factor scales, and the scores of subjects on the scales was considered. The implications of the findings for item construction and use of the test in various applications were discussed.


This document is a proposal for a publishable research paper. Ophthalmologists at the USAF School of Aerospace Medicine suspect one of the two depth perception tests used to
screen candidates for Undergraduate Pilot Training (UPT) inadequately evaluates depth perception abilities required to operate high performance jet aircraft. They hypothesize that candidates passing only the Verhoeff test for near stereopsis have higher attrition rates from UPT than candidates passing the VTA, a test of distant stereopsis. A historical prospective study will be used to test this hypothesis. A cohort of students attending USAF UPT during Fiscal Year 1990 will be compared. UPT attrition rates will be compared among type of depth perception test passed. Multiple logistic regression methods will be used to study attrition rate increases. Confounding variables evaluated include: commissioning source, previous flying experience, training base assigned, degree of phoria and gender of student. If this hypothesis is validated, the USAF may amend its pre-selection depth perception criterion for medical qualification of UPT candidates, thus resulting in significant financial savings to the US Air Force.


The main selection criteria for individual tests and test batteries used to select military pilot applicants are the construct and criterion validity, the overall cost of testing and the time requirements. Naturally, the derivation of decisions from a test battery requires a sufficiently high correlation between the tests and the criterion variable. However, recent metaanalysis (cf. Burke, Hobson & Linsky, 1997; Hunter & Burke, 1994; Martinussen, 1996) indicates that the correlation coefficients between a single test and the criterion measure do not exceed an absolute value of .30. There are a variety of causes for this, ranging from a lower reliability of the criterion or predictor variables (Lienert & Raatz, 1998; Goeters, 1998), an attenuation of the variance in the predictor variables due to selection (Lienert & Raatz, 1998; Goeters, 1998) to the lack of symmetry between the generality of the predictor variables and the generality of the criterion variable. With regard to the latter cause Wittmann and Süss (1997), Ajzen (1987) and Ree and Carretta (1996) pointed out that for more general and global criteria such as successful performance in a flight-simulator or an educational program, aggregate measures such as general ability (—g—) are better suited for prediction than more specific predictors. Thus one way to handle this problem is to combine the available information about an applicant to generate a prediction about his success. In general, one can resort to various methods of statistical judgment formation in order to do so. But classical methods of statistical judgment formation such as discriminant analysis or regression analysis are vulnerable to violations of their statistical assumptions and often lack stability in cross-validation in practical applications (cf. Bortz, 1999; Brown & Wickers, 2000). A promising alternative is the use of artificial neural networks. This statistical method has few requirements with respect to data characteristics and has proven to be a robust procedure for pattern recognition tasks (Bishop, 1995; Kinnebrock, 1992; Mielke, 2001; Rojas, 2000; Warner & Misra, 1996).

The main selection criteria for individual tests as well as test batteries used to select pilot applicants are the criterion validity, the overall cost of testing and time requirements. The selection of the respective tests can be based on recommendations of the Joint Aviation Requirements for Crew Licensing 3 (JAR-FCL3) and validation studies. Naturally, the derivation of decisions from a test battery requires a sufficiently high correlation between the tests and the criterion variable. However, recent metaanalysis (cf. Hunter & Burke, 1994; Burke, Hobson & Linsky, 1997) indicates, that the correlation coefficients between a single test and the criterion measure don’t exceed an absolute value of .30. There are a variety of causes for this, ranking from a lower reliability of the criterion- or predictor variables (Lienert & Raatz, 1998), an attenuation of the variance in the predictor variables due to selection (Lienert & Raatz, 1998) to the lack of symmetry between the generality of the predictor variables and the generality of the criterion variable. With regard to the later cause Wittmann and Süß (1997), Ajzen (1987) and Ree and Carretta (1996) pointed out, that for more general and global criteria such as successful performance in a flight-simulator or an aviation educational program, aggregate measures such as general ability (g) are better suited for prediction than more specific predictors. Thus, one way to handle this problem is to combine the available information about an applicant to generate a prediction about her or his success. In general, one can resort to various methods of statistical judgment formation in order to do so. But classical methods of statistical judgment formation, such as the discriminant analysis or the regression analysis, are vulnerable to violations of their statistical assumptions and often lack stability in cross-validation in practical applications (cf. Bortz, 1999; Brown & Wickers, 2000). A promising alternative is the use of artificial neural networks. This statistical method has few requirements with respect to data characteristics and has proven to be a robust procedure for pattern recognition tasks (Bishop, 1995; Kinnebrock, 1992; Mielke, 2001; Rojas, 2000; Warner & Misra, 1996).


Staff. (1943). Initial selection of candidates for pilot, bombardier, and navigator training. Washington, DC: Assistant Chief of Staff, Intelligence, Historical Division.


Staff. (1955). Psychological Research on Pilot Training in the AAF (Psychological Research Project [Pilot]). American Psychologist, 10, 7-16. This is the tenth of a series of articles (1, 2, 3, 4, 5, 6, 7, 8, and 9) dealing with the Aviation Psychology Program under the direction of the Office of the Air Surgeon, Headquarters Army Air Forces.

Stanley, M. D. (1973). *A Method for Developing a Criterion for Combat Performance of Naval Aviators*. Master of Science, Naval Postgraduate School, Monterey, CA. Current Naval aviator selection and screening procedures are based on the individual's statistical probability of completing flight training and do not determine the capability of the student to adapt to an operational environment. The resultant failure of some student aviators to complete the advanced stages of training and the ineffective performance of others in operational missions have caused a considerable financial loss and a lessening of combat readiness. A critical incident study, using 30 aviators who have combat experience, indicates that there are 10 categories of behavior which characterize effective and ineffective Naval aviators. Procedures to identify these categories early in flight training are discussed.

Stanny, R. R., Reeves, D. L., Blackburn, M. R., & Banta, G. R. (1988). Neuroelectric Selection of Naval Aviation Personnel: An Evaluation. Naval Air Station Pensacola, FL: Naval Aerospace Medical Research Laboratory. The problem of devising rational means to select candidates for flight has been with us for over 75 years now. It has become more important as aircraft have become more complicated and training more expensive. Selection for flight in the U.S. Navy is based on considerations of academic histories, written tests, physical examinations, and biographical information. The net effect is a substantial improvement over random selection. Nevertheless, perhaps 30% of those who enter flight training do not finish (1). Hence there is continuing interest in improving the selection process. Possibly the first attempt to use the EEG to forecast the performance of naval aviators was carried out by Alexander Forbes and Hallowell Davis as a part of the Pensacola Study of Naval Aviators of 1940-1941 (2,3).1 Forbes and Davis examined the electroencephalograms (EEGs) of several hundred Navy flight candidates for signs of epilepsy. They found no reliable association between the presence of minor EEG abnormalities and later flight performance.


Stead, G. (1994). *Qantas pilot selection procedures: past to present*
The causal role of general cognitive ability (g), personality, English proficiency, prior job knowledge, and training job knowledge on passing/failing flying training in the Norwegian Air Force was investigated. All the independent variables, with the exception of personality, were collected by paper and pencil measures. Personality was collected through the Defense Mechanism Test (DMT (Kragh, 1960); DMT-NPI (Neuman, 1978)), a tachistoscopic presentation of stimuli and ratings by a psychologist. The criterion consisted of a dichotomous pass/fail measure of pilot performance determined by the flight instructors. There was a direct influence of g on English proficiency and prior job knowledge, and an indirect path from g to pass/fail going through English proficiency, prior job knowledge, and job knowledge acquired in training. Personality showed a direct influence on the acquisition of prior job knowledge and on passing/failing pilot training only.

Stewart, J. E. (2006). Locus of control, attribution theory, and the "five deadly sins" of aviation. Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. The construct of Locus of Control (LOC) has been shown to predict a broad range of attitudes and behaviors, including risk taking and risk management, the performance of multiple tasks, distractibility, and the subjective perception of time. The above topics and many others have applicability to aviation settings. Over the past two decades, a few researchers have examined the relationship between LOC and hazardous attitudes, pilot errors, and other variables relating to safety and risk management. Most of this work has been correlational, and, in many instances, sample size has been quite small. The present paper reviews this work and other areas of research, which, though not specifically tied to aviation, have potential relevance to it. These include concepts from attribution theory, such as the optimism bias, in which people tend to attribute greater competency and lesser vulnerability to themselves than to similar others. Suggested applications of established and existing research in applied areas of social psychology are examined, with a focus on their relevance to aviation.


A number of procedures have been used by the U.S. military to improve the selection of potential military pilots. These procedures have usually focused on the assessment of psychomotor skills, intellectual abilities, and personal attitudes/interests commonly found in successful aviators. Since WWI, the U.S. Navy and Marine Corps have selected student aviators based on a battery of paper-and-pencil tests. Eligibility to take the selection test battery requires a college degree and a flight physical. The current selection test, the U.S. Navy/Marine Corps Aviation Selection Test Battery (ASTB), was developed to replace the Aviation Qualification Test/Flight Aptitude Rating (AQT/FAR). The AQT/FAR had been used unchanged since 1973. This test series was developed during WWII and includes four basic content areas: 1) general scholastic-related ability, 2) mechanical reasoning, 3) spatial reasoning, and 4) background experiences and interests. Extensive research has shown that the AQTIFAR was a valid predictor of aviation training success. The ASTB, like the AQT/FAR, is a paper-and-pencil test that takes approximately 2 hours to complete. The ASTB was developed jointly by the U.S. Navy and the Educational Testing Service (ETS) in Princeton, New Jersey. The development of questions for the ASTB was based on an extensive job and skill analysis conducted by ETS and the Navy. It is broadly useful for aviation and naval officer training and is comprised of 6 subtests that cover the same content areas as the AQT/FAR. Unlike earlier versions that targeted only attrition, the item selection and scoring of the ASTB is based on the ability to predict primary flight grades and training attrition.


Attrition in the training of U.S. naval aviation officer candidates represents a historic problem. The early identification of those likely to attrite during training would significantly reduce overall training expenditures. In this study, we assessed the value of biographical information for predicting early attrition at the indoctrination level of naval aviation officer training. We selected a random sample of 1551 aviation officer candidates and naval aviation cadets for analysis. The subjects selected had taken the Aviation Selection Test Battery (ASTB) between 1987 and 1990 and had completed the aviation indoctrination program operated by the Naval Aviation Schools Command in Pensacola, Florida. A principal component factor analysis of Biographical Inventory items was conducted with those who passed (N = 1176) and also with those who attrited (N = 375) basic aviation indoctrination. The resultant factors were then forced into a discriminant function analysis to determine if the factors obtained were different for the
two groups. We found that the factors were significantly different for the two groups. The results indicate that biographical data may be useful in identifying candidates who are most likely to attrite early from naval aviation training.


The purpose of our investigation was to determine if computer-based selection tests could predict training track assignment for student naval aviators. This study evaluated the predictive efficacy of an experimental battery of computer-based pilot selection tests for training classification. Student naval aviators are currently assigned to an aircraft training track based primarily on performance in primary training. Students were tested on the experimental test battery and classified into one of three aircraft training tracks based on their test scores. The resulting classifications were compared to actual selections made as the students progressed through naval aviation training. Using a sample of 237 students, linear analyses were conducted to evaluate the efficacy of predicted decisions. The unique contribution of the experimental battery was determined by comparing scores on the experimental battery to scores on the Navy/Marine Corps Aviation Selection Test Battery, a paper-and-pencil pilot selection test used by the United States Navy and Marine Corps, and student primary flight training grades. A significant classification model including one of the experimental selection tests was derived. The model was able to significantly predict fast attack pipeline selections before flight training.


The increasing cost of training aircrew to operate modern naval aircraft and the simultaneous decline in retention rate for these same trained aircrew increase the importance of utilizing the best selection methods available. This importance is underscored by the fact that every aircrew selectee who fails to complete training contributes to a potential operational personnel shortage if expected replacements necessary to maintain military readiness do not materialize as planned. Pilot selection research to date has generally focused on the testing of various psychomotor and cognitive abilities (Carretta, 1986; Davis, 1989; Dolgin & Gibb, 1989; Hilton & Dolgin, 1990). While these abilities would seem logically necessary for successful performance in flight training, some failures may be due, at least in part, to personality and/or motivational factors (Helmreich, 1982). Historically, researchers have tried to find the ideal aviator personality profile among numerous personality measures. This ideal aviator personality profile has often been anecdotally called “the right stuff.” Promising results have been found in identifying characteristics that improve the likelihood of later success in aviation such as
persistence, motivation, coolness under pressure (clear thinking), and novel problem solving (e.g., Retzlaff & Gibertini, 1987). Other researchers have considered personality factors with varying degrees of success (Dolgin & Gibb, 1989; Hunter & Burke, 1991). Certain personality characteristics or traits may correlate highly with success in initial/primary flight training. For example, interpersonal orientation, self-assertiveness, and achievement motivation are associated with pilot attitude and performance (Helmreich, Sawin, de Carsrud, 1986). Important developments in personality assessment have included attempts to avoid response bias by masking the personality dimension of interest and to screen for positive attributes, in contrast to a past emphasis on psychopathology (Picano, 1991).

Street, D. R., Helton, K. T., & Dolgin, D. L. (1992). The unique contribution of selected personality tests to the prediction of success in naval pilot training. Naval Air Station Pensacola, FL: Naval Aerospace Medical Research Laboratory.

This study concerns the relationship of naval flight training performance to scores on the Aviation Qualification Test/Flight Aptitude Rating (AQT/FAR) and the automated Pilot Personality Questionnaire (PPQ). We analyzed a sample of 211 pilot candidates who had taken the AQT/FAR and PPQ. We found that the PPQ competitiveness scale and three of the AQT/FAR subtest score means were significantly different (p < .05) for those who passed (N = 168) and those who attrited (N = 43) flight training. Discriminant analysis yielded a linear composite of the AQT/FAR and PPQ subtest variables that could be used to classify the students according to the likelihood of passing or attriting during flight training. The resulting discriminant function explained 9% of the variance in the pass/attrite criterion (r = .30). We found that a 50% reduction in attritions could be attained with a 23% increase in false rejections. The regression analysis as significant (p < .01) and indicated that three scales of the PPQ and the AQT and FAR scores accounted for unique variance in a linear prediction equation. The FAR and PPQ competitiveness scale were the most powerful predictors of overall flight training success.


The purpose of our investigation was to determine if personality testing and a five-factor model could improve the selection of Landing Craft Air Cushion (LCAC) vehicle crew members. Vehicle crew members for the LCAC are currently selected by their performance on a computer-based psychomotor selection system. The various psychomotor tests in the selection system have demonstrated predictive validity in LCAC crew training. Certain personality characteristics may also be involved in the LCAC vehicle crew training success. In fact, various researchers have found that personality testing may improve the selection of Navy/Marine Corps aviators. Increasing evidence indicates that a five-factor model may be useful in describing the personality characteristics involved in training success. We believe that a five-factor model may improve the selection system used for LCAC vehicle crew members. A principal component analysis with varimax rotation was conducted to determine the underlying structure of the Adult Personality Inventory with 168 LCAC crew candidates. The resulting factor scores were then statistically analyzed to determine the relation of the personality factor scores and the performance-based test to an underway grade in training criterion. The results indicated that one personality factor, openness, significantly improved predictions of the criterion (p < 0.05). Based
on these results, we believe that personality testing would improve the selection of LCAC vehicle crew members.


The Biographical Inventory in the Navy's aviation selection battery is interesting and important from both historical and psychometric perspectives. This traditional inventory, composed of heterogeneous items selected and keyed to predict retention vs. attrition of student naval aviators in flight training, has evolved over the years but can be traced back to the one used in World War II. This device has consistently been one of the most valid predictors of retention vs. attrition in the battery, overshadowing tests of general ability, mechanical comprehension, spatial ability, and aviation information (e.g., Fiske, 1947). It has been speculated that the inventory taps maturity and risk taking (Petho, 1980), but the reasons for its success have so far not been explored. Accordingly, the purpose of this study was to assess the dimensions underlying the inventory and their relations with retention vs. attrition of aviators in naval aviation training. (A parallel investigation of flight officers is underway.)


A biographical inventory has been used in the selection of students for naval aviation training since World War II, and its validity in predicting their retention in this training has been well established. This study investigated the constructs underlying the inventory and their relations to student retention criteria. A factor analysis of the items on the inventory for student pilots identified five factors. One factor, being a commissioned officer, appeared to account for the inventory's validity.


Discusses various functions of the Committee on Classification of Personnel in the Army: (1) classifying personnel according to their military qualifications (2) establishing the Trade-Tests division (3) enlisting the occupational needs of units in a division (4) extending the personnel work to staff corps troops (5) establishing the Central Personnel Bureau (6) appointing a committee on education and special training (7) organizing the War Service Exchange (8) rating the officers and candidates for commissions in the Officers Training Camps (10) cooperating with the Provost Marshall General (11) reducing the army paper work (12) enlisting the intelligence ratings of army men and (13) selecting aviators and navy men.


'The Army and Its Air Corps' was James P. Tate's doctoral dissertation at Indiana University in 1976. During the past 22 years, Tate's remarkable work has gained wide acceptance among scholars for its authoritative and well-documented treatment of the formative years of what eventually became the United States Air Force. Thoroughly researched but bearing its scholarship lightly, Tate's narrative moves swiftly as it describes the ambitions, the frustrations, and the excruciatingly slow march to final success that never deterred the early airmen. 'The Army and Its Air Corps' is one in a series of airpower history classics that the Air University Press is pleased to bring before a wider audience.


The authors attempted to develop a motivational screening device that would reduce the number of voluntary eliminees from undergraduate pilot training (UPT) and to develop an instrument which could be used to measure changes in motivation as subjects progress through pilot training (and later on the job). Both an a priori and an empirical keying approach were used to develop measures intended to predict several different types of attrition. Two hundred twenty-four self initiated eliminees (SIE) from UPT identified their reasons for attriting and these were subsequently examined to develop subcategories (i.e., criterion measures) that could be used to classify eliminees (e.g., SIE, flying deficiency, medical elimination, etc.). The initial item pool for predictor battery development contained 2500 items, 899 of which were included in the first version of the battery. The predictor battery consisted of two biographical inventories (BIs) with
approximately 300 items each and an Activities Index, which was a 300 item personality test. The battery was administered to all active students in the Officer training School (OTS), which was operated at Lackland AFB between January and November 1970. Of the a priori scales, only the BI creativity key demonstrated predictive validity with total attrition (r = -.10, p < .05). All other validities for a priori BI keys and for all of the other criteria were non-significant. Of the a priori personality scales, Abasement, Change, Energy, Science, Sensuality, Audacity and Motivation all demonstrated marginal negative validity for several of the attrition measures (generally on the order of -.10 to -.16). The validities for the Harm Avoidance scale ranged from .09 (ns) for elimination due to flying deficiency to .18 (p < .01) for UPT/OTS-SIE. Supplication also correlated positively with attrition, although the validity coefficients were generally small. The empirically derived scales resulted in moderately large validity coefficients (generally in the .20s and -30s) for almost all the attrition criteria, except for elimination due to flying deficiency. They also correlated significantly with self-rated lack of dedication to complete UPT (correlations ranged from .38 @ < .01) for flying deficiency to .72 @ < .01) for UPT/OTS-SIE). Generally, the a priori scales only accounted for a small proportion of attrition variance. Significantly more positive results were obtained using the empirically developed scales.


Frasca simulator, age 50-69, correlated factors with performance, speed/working memory correlates most with overall flight performance, age improved prediction, motor performance not age dependent.


Good visual contrast sensitivity (CS) is often described as a visual capability important for success as a military aviator and so has been suggested as a physical standard for personnel selection and retention. To evaluate this idea, we measured the CS of 135 U.S. Navy fighter pilots ranging in age from 24 to 44 years (mean = 30.20, S.D. = 4.06) and compared these to the CS of non-aviators. We obtained the non-aviator data from published studies of other investigators who used similar procedures with the same widely used, commercially available
apparatus (Nicolet CST 2000). In addition to this comparison, we correlated the pilots’ CS with their air-to-air target detection distances measured during air combat maneuver training and to their night carrier landing performance scores. The major findings were: 1) The mean CS of the aviators and the non-aviators were within +/- 1.0 S.D. of each other in most instances, and those few instances where a greater difference was found were parsimoniously explained by methodological and procedural factors; 2) sensitivities to different spatial frequencies were highly correlated among themselves, indicating much redundancy among the measurements; 3) there was no evidence of a relationship between CS and air-to-air target detection distances or night carrier landing performance.


Jet pilots (JP) (N = 44), helicopter pilots (HP) (N = 29), and college students (CS) (N = 41) were tested with a battery of vision tests designed to assess vision skills important for success as a naval aviator. Tests included measures of reaction time, high-contrast acuity, low-contrast acuity, spot detection, far-to-near gaze shift, near-to-far gaze shift, low-contrast acuity with glare, and dark focus. A Multivariate Analysis of Variance (MANOVA) compared the vision test performance of the three subject groups (JP, HP, and CS). Only with the Far-to-Near test was there no difference among the three groups. On all other tests, JP outperformed CS. The difference between HP and CS was less consistent and less dramatic than the difference between JP and CS. Only with the glare test were CS significantly better than HP. The results were interpreted as reflecting the influence of various selection factors, operational requirements, differential attrition, and age.


To determine medical qualification to fly, aviators who sustain head trauma or acquire other conditions that effect mental skills must undergo neuropsychological evaluations. A challenge for psychologists tasked with performing these evaluations is that conventional neuropsychological tests are characteristically normed using samples reflecting the general population. It would be more appropriate to compare aviators with a sample of their peers. The present paper demonstrates the importance and usefulness of aviator-specific psychometric norms. To do this, normative tables were developed from a large sample of United States Air Force pilot training candidates using the Multidimensional Aptitude Battery, a standardized intelligence test. Psychologists may find these tables useful when evaluating aviators.


Develops a non-verbal intelligence test similar to the Beta examination of the army. Four types of tests used in the Beta test were included, along with four additional tests. Describes the material, the instructions given by the examiner, and the administration and scoring procedures of the tests. Discusses the significance of the examination score for a candidate who takes it without any previous acquaintance as compared to one who is already familiar with the two tests. Emphasizes the practical advantages of such tests.


Reviews the Martin Scorse movie, The Aviator. Emphasizes fear as a motivator for Howard Hughes' acquisition of power.


Because of the similarity of the AVO console to a desktop flight simulator, a preliminary idea of what aptitude factors would predict success in AVO training might be obtained through a validation study using a desktop flight simulator as the criterion task. In this study, 171 young men and women learned to fly a simulated light aircraft. The participants also took a comprehensive g-hour battery of computer-based tests. Results indicated that learning was predicted well by general cognitive ability (g)/working memory (WM) and multilimb coordination. When structural equation modeling was used to distinguish declarative and procedural learning, declarative learning was found to be strongly dependent on g/WM, which is consistent with previous research. That g/WM did not correlate with procedural learning explains why correlations of this ability with learning diminished as instruction moved from a declarative to procedural emphasis. Only multilimb coordination ability was strongly related to procedural learning.


A collection of 25 computer-administered tests was given to a sample of 172 adults to examine the overlap between the cognitive and perceptual-motor abilities domains. The focus was on multilimb coordination ability as a criterion variable, as this ability has a long history of validity in predicting pilot training performance. Structural equation modeling was used to test two related models of human abilities. The first one being a causal model and the second being a nested hierarchical model. The results suggest that multilimb coordination ability is not simply another manifestation of general cognitive ability, but instead the result of several abilities such as dynamic visual processing, visuo-spatial processing and working memory. A second narrower perceptual-motor factor, target tracking, was found to be related to processing speed but no other cognitive factor.


Two 4-part chart reading tests and two 3-part dial-reading tests were studied for the efficiency with which they predicted scores obtained on the Dial and Table Reading test. Predictions from combination of 2 or 3 chart-reading subtests were better than those obtained from the dial-reading tests and were almost as efficient as were all 8 parts of the 2 chart-reading tests. Sets of 2 or 3 chart-reading subtests, requiring 2 minutes of testing time for each subtest are equivalent in predictive power to the Dial and Table Reading test.


Abstracts are assembled of 126 technical documentary reports issued by the Air Force's personnel research laboratory from January 1958 through December 1961. They cover research projects in selection, classification, and utilization of Air Force personnel. The reports are indexed by personal author, corporate author, and project number.

The research described in this report was performed under Project A, the U.S. Army's current, large-scale, manpower and personnel effort to improve the election, classification, and utilization of Army enlisted personnel. This report documents the development and field test of behaviorally anchored rating scales for nine Military Occupational Specialties (MOS). These include combat, combat support, and noncombat MOS. For each MOS, the behavioral analysis method was used to generate examples of performance. These examples were used to identify performance effectiveness dimensions and to develop behavioral definitions of performance for each dimension. Across the nine MOS, behavioral summary rating scales contained from 7 to 13 performance dimensions. The nine sets of MOS-specific behavioral summary rating scales were field tested in continental United States and overseas locations in two groupings (Batch A and Batch B). For each MOS, ratings scales were administered to 120 to 160 first-term soldiers and their supervisors. Within each MOS, interrater reliability estimates for individual performance dimension ratings were reasonably high and rating distributions were acceptable, indicating no leniency or severity effects. Results from the field tests, along with suggestions from proponent review committees and Project A staff, were used to modify and prepare the nine sets of rating scales for the Concurrent Validation study.


Determining the characteristics of successful military pilots requires two types of assessments of individuals. First, it is necessary to define and measure their personal characteristics and, second, it is necessary to array them on some continuum of success. These two operations can represent either a concurrent or a longitudinal undertaking. In the concurrent approach, both the degree of success and the personal characteristics are evaluated at more or less the same time. This has the advantage of representing the current interaction of a subject’s attributes and the extent of his success. It is also less expensive in terms of the number of subjects required to obtain samples of a specified size. It suffers from the disadvantage that no information can be obtained concerning the association between success and personal characteristics over time. In other words, since measurement of an individual's personal characteristics may be directly influenced by the extent of his success at the time of measurement, no inferences concerning the predictability of success by personal characteristics are warranted.


Trites, D. R., Holtzman, W. H., Templeton, R. C., & Sells, S. B. (1953). Psychiatric screening of flying personnel: Research on the SAM Sentence Completion Test. Randolph Field, TX: USAF School of Aviation Medicine. This report describes research on the development of the SAM Sentence Completion Test, "What Is He Saying?" which was part of the Randolph Field Battery of experimental psychiatric selection tests developed and for this project. The test form a given in Repair No. 1 (34), with instructions for its administration. The experimental design of the program for the development of psychiatric screening devices for flying personnel is described in Reports Nos. 1 and 2 (3, 35). The Sentence Completion Test is a projective technique of personality study. A projective device is an ambiguous content situation which is designed to encourage the subject to invest the stimulus materials with his own wishes, impulses, fantasies, and values -in other words, to "project" certain aspects of his personality into the test response. Somewhat like the word-association method, the subjects are presented with stimuli consisting of one or more words representing incomplete sentences. The task of the subject is to complete each sentence. The SAM Sentence Completion Test was structured to evoke responses indicative of attitudes and feelings related to motivation for military flying, self-esteem, interpersonal relations, and conformity to social customs. These factors are described in report No. 1 as important factors in predisposition to psychiatric failure.


Selection procedures for OCS were changed early in 1952, and this report compares results of selection by the revised procedures with selection by those previously used. The existence of a backlog of applicants who had been accepted under the old procedures made it possible to study and OCS class containing reasonably large samples of candidates selected by both methods. The two crucial changes were elimination of college credit requirements and removal of the subjective ratings by past military supervisors. The general effect was to render the bases of officer candidate selection more objective, placing a floor under general aptitude requirements and improving the estimation of past military accomplishments. A greater percentage of the group selected under the new procedures graduated from OCS, and a significantly higher percentage ranked high in the class. Failures for military deficiency were in the same proportion for both groups, but the new selection procedures resulted in higher academic success.


and Training Research Center.

Procedures for screening and selecting applicants for USAF Officer Candidate School (OCS) were modified early in 1952. The new composite score consisted of an officer quality score derived from the Aviation Cadet Qualifying Test (ACQT), a score based on board interview, and other scores awarded for years of college completed, airman technical schools completed, airman rank, and length of service, weighted according to their presumed importance to success in OCS and to later officer performance. Scores on the composite and its components were obtained for OCS Class 53-B. Biserial correlations between these scores and the criterion of graduation did not differ significantly from zero. Only the ACQT officer quality score was significantly related to academic grades. 4 variables were significantly related to military grades (service, rank, and interview positively, and college negatively). Further analysis suggested that modified weighting of the components could improve the prediction of both criteria.


This report is a nontechnical review of the research programs with a listing of 75 scientific publications produced by Air Force personnel research in this area. Contributions of psychological tests to officer selection for flying training and technical training are summarized. 5 graphic figures illustrate the efficiency of tests in selection. The directions in which current research is moving are indicated. Aptitude tests are generally used in officer selection programs but are not yet fully exploited in officer classification.


The author reviews the various methods that have been used to select pilots over the years. This comprehensive review begins with a description of the "selection" method used by the Wright brothers in 1913 (i.e., they flipped a coin) to decide who would fly their plane and includes the present-day multivariate approach to selecting pilots. He describes the historical depiction of pilots as sportsmen with initiative and humor, as well as high-spirited, gregarious and lacking in imagination. The RAF developed the first basic aptitude tests used to select pilots during WWI, but these efforts receded after the war ended. However, by 1941 the high incidence of failure in RAF pilot training became a major problem and the focus on pilot selection once again increased. This cycle is contrasted against the remarkable success of the German Luftwaffe selection procedures. The Germans used a combination of psychological and aptitude testing and were very successful in selecting pilots. The author reviews the use of personality tests in general and concludes that they have had some success in predicting pilot training performance, but that their susceptibility to faking is a major failing. The use of "objective" tests (e.g., the Rod and Frame test) is also reviewed. He concludes that while there is no one reliable way to select pilots, there is considerable evidence suggesting that the measurement of the "right stuff" is feasible.


The report presents an abbreviated account of developments in military psychology in the Army, beginning with the general ability tests developed and used in World War I. A chronological listing of "Events and Organizational Change" is accompanied by interpretive comment in terms or "Research Achievements and Directions." The following periods are treated: Beginning, 1917 to 1921, Interim Period, 1921 to 1939, World War II, 1939-1945, Early Post-War Years, 1945 to 1951, Period of Transition, 1951 to 1960, Developments, 1960 to 1969, and 1969 to the Present.


An interview with Dr. Dora Strawder and a discussion of aviation psychology.

This report presents results of an evaluation of the Army's 17 5/40 Initial Entry Rotary Wing (IERW) training program. The program consists of 175 flight and 40 simulator hours. In the last training phase of the program, students are divided into two training tracks, i.e., Utility Helicopter and Aeroscout. Data were gathered from IERW training files and through questionnaires administered to instructors at the Army Aviation Center and to unit instructors, supervisors, and program graduates at aviation field units world-wide. Data were gathered on graduates of both the 175/40 program and the 180/20 program that preceded it, and comparisons between the two programs are made. Results and conclusions are presented with reference to nine specific evaluation objectives. Major conclusions are: (1) the 175/40 IERW course is accomplishing its objectives; (2) the 175/40 course is an improvement over the 180/20 course; and (3) proficiency progression and individualized training can play an effective role in IERW training.

This report presents results of an evaluation of the Army's 17 5/40 Initial Entry Rotary Wing (IERW) training program. The program consists of 175 flight and 40 simulator hours. In the last training phase of the program, students are divided into two training tracks, i.e., Utility Helicopter and Aeroscout. Data were gathered from IERW training files and through questionnaires administered to instructors at the Army Aviation Center and to unit instructors, supervisors, and program graduates at aviation field units world-wide. Data were gathered on graduates of both the 175/40 program and the 180/20 program that preceded it, and comparisons between the two programs are made. Results and conclusions are presented with reference to nine specific evaluation objectives. Major conclusions are: (1) the 175/40 IERW course is accomplishing its objectives; (2) the 175/40 course is an improvement over the 180/20 course; and (3) proficiency progression and individualized training can play an effective role in IERW training.


The purpose of this paper is to present initial findings on utility of two psychomotor tests for improved selection of Air Force pilots. During, and for a few years following World War II, the predecessor organization of the Air Force Human Resources Laboratory conducted an extensive program of psychomotor research, and selected psychomotor tests developed under this program were an integral part of the World War II Aircrew Classification Batteries. Much of this early effort has been described by Melton (1947). Generally, it was found that such assessments had validity for predicting eliminations from pilot training beyond that achieved with a battery of paper and pencil tests. Use of psychomotor assessments in the Air Force pilot selection testing program was discontinue in the early 1950’s because of the expense and difficulty of maintaining and calibrating the required equipment under decentralized testing conditions.

Van Wagenen, M. J. (1919). Some Results and Inferences Derived From the Use of the Army Tests at the University of Minnesota. Paper presented at the Annual Meeting of the American Association for the Advancement of Science, Saint Louis, MO.

During the two academic years 1917-18 and 1918-19 the Army Test Alpha or its equivalent, Form E, was given to the freshman classes of nearly all the colleges at the University of Minnesota In some of the colleges the results were used as an aid in diagnosing the causes of student failures. In other colleges the purpose was purely an experimental one. The use of the tests has revealed some significant information regarding the students of the various colleges. For instance, over eighty per cent of the student body of the University as a whole were found to come from the upper fifteen to twenty per cent of the population in general. With the exception of two of the individual tests the women of the several colleges did just as well or even better than the men of the same colleges, but in these two tests—the range of information test and the arithmetic problems test—from seventy to seventy-five per cent of the men did as well or better than the median woman. Making correction for the excess of overlapping due to the use of a single test, from sixty-five to seventy per cent of the men may be expected to do as well or better in solving arithmetic problems and to have as wide or a wider range of the kind of information called for in the Army Tests than has the median woman of the same college.


This study examined the utility of Jackson’s Personality Research Form (PRF) in the selection of aircrew in the Canadian Forces. A total of 1962 male candidates completed either the English or French version of Form E. Major findings focused on the validity scales, anglophone versus francophone differences on the 22 scales and the predictive validity of the scales against performance in flying training. Results show that while Infrequency scores are within the range reported by Jackson, the Desirability scores obtained were very high. There are also some significant correlations between Desirability scores and scores on other trait scales. There were a number of differences between performance of anglophone and francophone subjects and these will be discussed, together with other early stage psychometric evaluations of the Form. Finally, the PRF may prove useful in the counseling of candidates who are interested in becoming military pilots.


Current diagnostic electroencephalogram (EEG) investigations in aircrew selection and certification lack both standardization and reference to universally applicable criteria for their effective use. Extrapolation from clinical EEG studies may not be appropriate. Recent studies on serial EEGs in aircrew are lacking, whereas follow-up of individuals who failed certification is nonexistent. Population-based EEG studies in healthy subjects are generally underpowered to establish the significance of pathological EEG findings. Advanced digital video/EEG recording, in combination with standardization of data exchange formats and automated detection of pathological grapho-elements, is cost effective when carried out for extended periods, e.g., during flight simulator sessions. Extensive databases of serial video/EEG records in aircrew may thus be easily obtained and validated over time. Prognostic inferences on the significance of pathological EEG discharges may subsequently be derived from these databases.


The Adaptability Rating for Military Aviation (ARMA) is that portion of the initial flight physical that assesses an aviator candidate's motivation for and potential adaptability toward an aviation career. A survey was mailed to all USAF operational flight surgeons in the continental
U.S. to describe the frequency and distribution of ARMA usage and attitudes. Descriptive statistics suggest that the ARMA is used suboptimally in accordance with current USAF regulation. ARMA training, flight surgeon satisfaction and lack of regulation clarity are described and discussed. More flight surgeons are dissatisfied with the ARMA than are satisfied, and the regulation is perceived as unclear in the area of final disposition for candidates with equivocal ARMA's. A post-hoc analysis to rule out the influences of rank, gender, experience and residency training was performed. Residency training in Aerospace Medicine is beneficial in terms of doing an ARMA, when required, and covering recommended areas. Females and those with less than 1 year experience perform an ARMA more frequently than males and experienced flight surgeons. Despite the limitations of the current ARMA, it should not be abandoned. Recommendations to improve it are provided. Doing better ARMA's can lead to decreased illness, injury, accidents, and attrition.


Viteles, M. S. (1942). An historical introduction to aviation psychology: Civil Aeronautics Administration, Division of Research.


Voas, R. B. (1957). Inventory testing of vocational interests of Naval aviation cadets: Final results. Pensacola, FL: Naval Aerospace Medical Institute, Naval Aerospace Medical Research Laboratory.

This study attempts to determine whether interest patterns, as measured by the Kuder Preference Record (KPR), differ between students who successfully complete flight training and those who do not. The KPR is a standard interest inventory that consists of nine scales: Mechanical, Computational, Scientific, Persuasive, Artistic, Literacy, Musical, Social Service and Clerical. This was a follow-up study for work completed by Rosenberg and Izard (1954), which found that students who failed flight training scored lower on the Mechanical and Scientific scales and higher on the Persuasive, Literacy and Musical scales than did those students who passed flight training. Six hundred and five aviation cadets participated: 465 who successfully completed training, 74 who withdrew at their own request and 66 who failed at some point during flight training or were eliminated for medical reasons. The author constructed an empirically keyed Voluntary Withdrawal (VW) scale based on results from Rosenberg and Izard. For those scales that differentiated successful cadets from unsuccessful cadets, a score of "1" was assigned to those items that unsuccessful cadets endorsed and a score of "0" was assigned to those items that successful cadets endorsed. A high score on the VW scale indicated a pattern of interests similar to cadets who had failed flight training. Results of this study showed that of the nine KPR scales, significant differences between successful and unsuccessful cadets were only found for the Mechanical and Scientific scales. In addition, the cadets who voluntarily withdrew from training scored significantly higher on the VW scale than cadets who succeeded, although there was quite a bit of shrinkage in the cross validity sample (r = .56 compared to r = .17). When scores on the Mechanical Comprehension Test were held constant, the differences between the successful and unsuccessful cadets on the VW scale became nonsignificant. Thus, the validity of the VW scale was apparently based mostly on its relationship to mechanical ability.


This study examines the effects of stress tolerance on a purified success in training criterion. Subjects were 1540 flight trainees who were participating in decompression chamber classes. Stress tolerance was operationalized as whether or not subjects replaced their air mask during a chamber ride or experienced ear blocks. Flight training criteria included poor performance leading to failure in training or voluntary withdrawal, and an indication of "anxiety toward flying" (assessed in an unstructured exit interview). Subjects were classified into one of three groups at the end of training: (1) "S" group were those who successfully completed training; (2) "P" group were those who failed (this group was subsequently divided into "PA" – poor performance with anxiety and "PN" - poor performance without anxiety); and (3) "M"
group made up of all subjects who did not complete training for various other reasons. Results indicated that those subjects who demonstrated anxiety in the chamber ride were more likely to demonstrate anxiety during flight training, that is, the S group demonstrated considerably less anxiety than the P group during the chamber ride. In addition, the PN group demonstrated less anxiety than did the PA group. Also, the PN group did not differ from either the S or M groups in level of anxiety.


In January 1999 Australian Defence Force Basic Flying Training School (ADF BFTS) was formed at Tamworth, NSW. The unit’s role is to conduct Ab Initio flying training for ADF pilots. This study examined preliminary results of ADF BFTS courses to validate selection procedures for the new ADF flying training system. Predictor variables included flight screening scores, previous flying experience (PFE), student age and aptitude test scores. Full data was available for 80 students from courses ADF1-ADF5 who underwent the Flight Screening Programme (FSP). Data was collected from the pilot selection database maintained by Central Pilot Selection Agency (CPSA). Chi Square analysis was used to examine the association between predictor variables and Failure for Air Work Rates (FAW) in Basic Flying Training. In accordance with earlier research, this study showed that the FSP is a valid predictor of performance in military flying training. Similarly there was a significant relationship between PFE and performance at BFTS. Aptitude test scores and student age were not related to performance at BFTS. Range restriction may have affected the analysis of aptitude test scores. Recommendations from the study include: that the FSP be maintained as one of a number of techniques used to select pilots for the ADF; and that aptitude test scores (via Pilot Index) continue to be used as a method of selecting prospective flight screening candidates. Future research will examine the validity of current selection procedures for the full continuum of military flying training.


Wagner, R. F. (1951). Development of standardized procedures for defining the requirements of aircrew jobs in terms of testable traits: USAF School of Aviation Medicine.


Three experimental peer rating forms were developed for use in research in prediction of the aviation training performance criterion--completion/attrition--from the training program for Aviation Warrant Officer Candidates at the U.S. Army Helicopter School. This paper describes the construction of the ratings, the "Potential Aviator Mating" forms, and compares the validity of these forms with the Contemporary Evaluation Form (CEF) used by the U.S. Army Helicopter School. The basic comparison involved validity between absolute scale and ranks. The original validity coefficients were sufficiently high to anticipate that the use of peer ratings may increase predictive accuracy in a multivariate system.


Walters, B. A., Huber, S., French, J., & Barnes, M. J. (2002). Using simulation models to analyze the effects of crew size and crew fatigue on the control of tactical unmanned aerial vehicles (TUAVs) (pp. 35). Aberdeen Proving Ground, MD: U.A. Army Research Laboratory.

This report describes a study conducted by Micro Analysis and Design, Inc., for the U.S. Army Research Laboratory (ARL). One area of research examined by ARL was the staffing required to operate tactical unmanned aerial vehicles (TUAVs). The primary objective of the study was to use simulation modeling to analyze how fatigue, crew size, and rotation schedule affect operator workload and performance during the control of a TUAV. Computer simulation models were developed with the Micro Saint Discrete Event Simulation software to simulate the tasks that operators perform when controlling a TUAV. These models, which contain system-specific attributes of the Shadow 2001 TUAV, included a fatigue function to predict performance effects for day and night missions. Subject matter experts (SMEs) provided the list of tasks involved in controlling a TUAV (during normal operations and emergencies), the order of these tasks, and the visual, auditory, cognitive, and psychomotor workload values associated with each
task. Twelve different crew configurations were examined for the tactical operations center (TOC) and the launch and recovery station (LRS), which ranged in size from 8 to 15 crew members. The conclusions from executing the models and interviewing SMEs (during 12- and 18-hour missions) indicate that reducing the number of aerial vehicle operators (AVOs) and mission payload operators (MPOs) in the TOC can result in more aerial vehicle mishaps during emergencies, increased search time, and a decreased number of targets detected. For example, compared to six AVOs or MPOs in the TOC, the addition of two crew members resulted in only slight performance gains of a 6% increase in target detection and a 4% decrease in target search time. However, when the members of the crew were reduced to four AVOs or MPOs in the TOC, there was substantial performance loss (20% decrease in target detection and a 15% increase in target search time). The general conclusion is that a crew of 12 (TOC [two MCs and six AVOs or MPOs]; LRS [two MCs and two AVOs]) is the most efficient trade-off between performance and crew size. The implications of these findings for other possible crew configurations are discussed, along with plans for further analyses.


Investigates the structured interviews for U.S. Air Force pilots. List of subject attributes for the interview; Types of predictors in the interview; Validity of the interviews.


Warner, J. D., & Knapp, B. G. (2000). Crew characteristics for common ground station applications (pp. 40). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory. The purpose of this study was to determine the required skill set for common ground station of the future (CGS-future) and compare it to the present military occupational specialty (MOS) 96H skill requirements in order to determine the appropriateness of the 96H for operating the CGS-future. We approached the objectives by conducting subject matter expert and documentation reviews of presently accepted training for MOSs 96H, 96D, and 96B; by employing the Job Assessment Software System (JASS) in order to assess what skills and abilities are needed for what duty and at what demand; and by creating a dynamic task-network performance model to simulate work flow and error rate during different operating conditions. Results indicated that there is skill shift (higher levels of analytical skills required) for the operator of the CGS-future. These skill demands are at a level similar to or higher than those required by MOSs 96D and 96B. However, a simple substitution of the more analytically trained MOSs (96B and 96D) is not the solution, since the training cost to learn CGS skills exceeds the cost to enhance current training. The cost-effective approach would be to determine what 96B and 96D skills and how much of them must be integrated into the 96H training. Furthermore, the complex relationship of the training to successful performance, especially under different mission demands, is not validated. Further use of the dynamic model to develop a body of data derived from careful manipulation of personnel mixes and mission requirements could provide valuable advice to decision makers who track these complex issues.


Waters, L. K. (1964). The relationships among the needs and values of flight candidates. Pensacola, FL: Naval School of Aviation Medicine.


Waters, L. K., & Wherry, R. J. (1959). Factor analysis of selection Lists and performance measures in the U.S. Naval School, pre-flight. Pensacola, FL: Naval Aerospace Medical Institute, Naval Aerospace Medical Research Laboratory.


Watson, S. E., & Walker, R. F. (1998). The basic research UAV task environment (BRUTE). Paper presented at the 42nd Annual Meeting of the Human Factors and Ergonomics Society. This demonstration features a PC-based Synthetic Task Environment (STE) developed by the United States Air Force (USAF) to investigate Uninhabited Air Vehicle (UAV) operations. UAVs are remotely operated airplanes, and are typically used for aerial reconnaissance tasks. The goals for the development of this UAV-STE were that it should: 1. Capture key aspects of UAV operations 2. Allow rapid configuration into existing and conceptualized UAV systems 3. Flexibly incorporate experimental manipulation and data collection 4. Avoid exotic hardware/software reconfiguration requirements.


In support of our Air Staff initiatives, we in the Military Testing Section (Ken Schwartz, Chief, USAF Military Testing and I) have spent the last two months working to revitalize the USAF personnel research program for strategic force management issues. This has consumed a large portion of our attention up through 4:00 PM on Saturday, September 30th, the last day of our fiscal year. I am happy to report that we were successfully in launching four major “general personnel research” contracts totally nearly two million dollars from unexpected fallout funds. In one of these contracts there is a task to evaluate the effectiveness and potential for using the new Self Description Inventory (SDI+) which has become part of the Air Force Officer Qualifying Test (AFOQT) since August 2005. We now have nearly 10,000 AFOQT answer sheets which represent not only current USAF Officers, but also “applicants” in the sense that they entered the commissioning source, but may not have progressed through final commissioning. This dataset will be an invaluable asset in our first, near-term plans for evaluating personality measures in any future operational Air Force personnel program.


Military testing in the United States Air Force (USAF) falls into several categories. Testing associated with Education & Training is vital for preparing a team-oriented, a mission-ready force. The other forms of testing, collectively known as Personnel Tests, have a direct impact on defining an individual’s official job and level of responsibility. The goal of personnel
testing is to select, promote, or certify the most qualified individuals. The Air Force Personnel Center at Randolph AFB (HQ AFPC) is charged with managing this force throughout the world to ensure the right person is at the right place at the right time. For this reason, all military personnel tests are centrally managed by the Air Force Military Testing Section of AFPC. The Air Force Military Testing Section (AFPC/DPPPWT) is responsible for the integrity and application of personnel tests that are used across the Air Force. This section performs many functions related to testing including 1) recommending new research/validation needs, 2) evaluating emerging test technologies; 3) managing test development contracts, 4) validating proposed instruments against professional standards, 5) applying accepted psychometric procedures in response to non-standard testing situations, 6) interpreting, reviewing, and updating Air Force testing policy instructions, 7) responding to inquiries and requests for waivers, 8) coordinating the Test Control Officer network, 9) providing oversight and coordination for testing and associated study materials printing and distribution operations, 10) receiving and scanning completed test answer sheets, and 11) scoring, uploading, and answering routine inquiries on all test results. Because many of these tests will affect the quality of people in the Air Force (as well as a person’s pay), AFPC/DPPPWT is a key player in combating test compromise, including support for detection, statistical analysis, and prosecution, working in close concert with the Air Force Office of Special Investigations (AFOSI) and the Test Development Flight of the Air Force Occupational Measurement Squadron (AFOMS/TE) and the AFOMS Commander.


The Officer Interview Board (OIB) is the final stage in the selection process for applicants applying for Officer level positions in the Royal Australian Air Force (RAAF). Concerns were raised about the validity of the OIB in the selection of Officer applicants, including; limited standardisation between OIBs, high rejection rate of applicants, and the potential for the selection criteria assessed at the OIB to have become outdated. This study examines the OIB, in the context of the entire RAAF Officer selection system, to identify areas that could be refined and/or redeveloped. The study aimed to develop a selection system that was current, streamlined, standardised and a valid predictor of future training and job performance. A job analysis was conducted to determine the essential attributes of entry-level RAAF Officers, and subsequently identified nine attributes deemed essential to assess in the selection system. An examination of the RAAF Officer selection system revealed that a number of the essential attributes were not being assessed. In addition, it was maintained that some of the essential attributes could be assessed through assessment techniques other than an interview format. It was
proposed that the OIB be redeveloped into an assessment-centre style Selection Board to assess the essential attributes of RAAF Officer applicants. This paper describes the development and final design of the RAAF Officer Selection Board. Future directions for the validation of the Selection Board will also be discussed.


It is now widely agreed that current ability measures reflect a complex interaction of environment with genetic potential. This leads to a basic measurement problem since persons with the same measured ability may vary widely in potential due to non-equivalent learning opportunities. The purpose of this paper is to present a model which may hold some promise in psychometrically distinguishing ability (current status ) from aptitude (potential). Data on a simple ability are analyzed according to the model to illustrate how some of the practical problems may be solved.


This paper aims to provide a brief synopsis of the Vienna Test System based on a short demonstration tape containing some 50 or so exercises, including sub-divisions, described as computer aided psychological diagnosis. It is intended not as a specific test review, but as a helpful guide to those interested in accessing a particular test from this extensive battery. It should be noted that only a brief introduction to many of the tests is possible as the demonstration tape only gives shortened samples of content and minimal normative data in terms of comparison groups.


Historians have generally considered that the British army dismissed the potential value of aircraft prior to the First World War, only realizing the implications of military aviation after its outbreak. Before 1914, however, many senior British army officers had begun to understand that air power would soon change the nature of warfare. They had reached this understanding as a result of demonstrations by aircraft during many of the manoeuvres and other training exercises carried out in the years 1910–13.


The methodological issues involved in demonstrating the existence of a time-sharing
ability are outlined. A survey of relevant experimental literature indicates that, while there is some evidence for a task specific time-sharing ability, there appears to be little for a more general "A-factor" of attention of dual task performance ability. An experiment is then described in which 40 subjects performed 4 tasks singly and in various pairwise combinations. The tasks, tracking, spatial judgments, digit classification and auditory memory, were selected to systematically load different stages of information processing. The patterns of task interference observed, conformed to predictions of structure specific capacity theories of attention, with structures defined by processing stages, processing modalities and cerebral hemispheres. Confirming previous research, little evidence was provided for a "general" time-sharing ability. More specific abilities were however suggested by the data to relate to visual scanning, and automation of auditory memory store.


A model of the attention allocation process in a dynamic environment with changing task demands was presented. The model describes the manner in which optimal and sub-optimal allocation of resources can be revealed by linear coherence analysis of the dual task performance and difficulty measures, and partitions the allocation process into subprocesses of performance demand evaluation and subsequent resource allocation. Eight subjects time-shared two compensatory tracking tasks under conditions of constant task difficulty, and under conditions when the difficulty (percent acceleration dynamics) of one task was varied over the course of the trial. Subjects were instructed to maintain constant performance on the variable difficulty task, and augmented performance feedback was presented on half of the trials. The data were evaluated in terms of the model, and coherence and error analysis revealed that allocation was far from optimal. The failure of augmented feedback to improve the optimality of allocation suggested that the limitation lay within the allocation, rather than the demand evaluation process. Some reasons were proposed for these limitations, and for the contrast of the current results with optimum allocation observed in constant difficulty dual task studies.


The aim of this pilot study is to examine training enlistment standards utilizing Army Training Requirements and Resources System (ATRRS) training data which records individual-level training events for all MOS, but is limited to pass / fail outcomes. Specifically, we seek to investigate the tradeoffs between training eligibility and Advanced Individual Training (AIT) completion brought about through lowering / raising minimum enlistment training standards, and to develop methodologies which can be utilized to assist school proponents in assessing the appropriateness of their Aptitude Area (AA) cut scores. For the initial effort in this pilot, the 50 MOS investigated belonged to a handful of school proponents who expressed an interest in the objectives of this study, plus a few additional ones recommended by Army Accessions Command. Subsequently, an additional 30 MOS that promised sufficient numbers of (failure) observations were also included. The authors specify and estimate binary logistic regression models of pass / fail training outcomes over the 2001 – 2004 period. Training outcome is estimated as a function of AA governing composite, Soldier demographic, and component membership variables. The estimated models are then applied to the larger Army enlisted contract population to examine the policy tradeoffs. For select MOS, the policy analyses are examined more closely using risk analysis simulation methods.


This paper reviews the process of selecting officers for U.S. naval aviation training and describes one of the principal selection tools, the Aviation Selection Test Battery (ASTB). The 1992 version of the ASTB is a paper-and-pencil test administered to all applicants for naval aviation training. ASTB scores and ground school and flight training performance data were available for 2852 student naval aviators and student naval flight officers, and these data were used to re-assess the validity of the ASTB in predicting student performance. The results indicated that the ASTB remains a valid predictor of ground school and flight training grades, and to a lesser extent, attrition from training. For a small subset of the sample used in these analyses, data from a computer-based performance test (CBPT) were also available. The CBPT required subjects to engage in multi-axis tracking tasks concurrently with other cognitive tasks, such as dichotic listening and working memory tasks. Scores from the ASTB, the CBPT, and grades from ground school were entered into a linear regression upon primary flight training grades. The results showed that the combination of ground school and CBPT scores can be used as a good predictor of performance (R2 = .33, p <.0001). Although these results will require cross validation, the CBPT shows promise as a new selection tool. The importance of these results is discussed in the context of a recently developed computer-based version of the ASTB.


This research study was undertaken to create recommendations for unmanned aircraft pilot medical certification requirements. The effort consisted of the convening of a panel of subject matter experts and interactions with groups engaged in the process of establishing unmanned aircraft pilot guidelines. The results of this effort were a recommendation and justification for use of the second-class medical certification.


Examines accuracy of peripheral target recognition and visual field narrowing among aviators and nonaviators. Observation of larger effective functional visual fields among aviators; Susceptibility of nonaviators to visual field narrowing; Effect of load increases on aviator performance. Examines accuracy of peripheral target recognition and visual field narrowing among aviators and nonaviators. Observation of larger effective functional visual fields among aviators; Susceptibility of nonaviators to visual field narrowing; Effect of load increases on aviator performance.


From the enormous early task of classifying thousands of personnel into military assignments under the pressure of World War I, to today's state-of-the-art work in computerized testing, military research has been at the forefront of personnel testing. This paper will review some current research developments, including a recent cross-national comparison of testing methods, and reflect on future research directions.


This study focuses upon enhancing the U.S. record of air-to-air combat kills through more stringent and comprehensive selection procedures. Specifically, it presents empirical evidence which demonstrates that a program can be developed to select pilots who will be effective in air-to-air combat. Developing such a selection program involves a job analysis of the fighter pilot task, the generation of testable trait hypotheses, the development of predictor variables and peacetime and combat criteria, and validation and cross validation during peacetime and during combat. Forty-five factors were identified only 10 of which are adequately evaluated within a current U.S. military selection program upon initial entrance into pilot training. Assessment of the 35 untapped factors is within technological reach; in fact, many of these factors can be assessed by tests which are presently available.


History has demonstrated that there is a pressing need for improved selection and training of fighter pilots. In World War II, only one of twenty pilots became an ace. The U.S. Air Force kill ratio in Southeast Asia was approximately 2.5 to 1. In contrast, the Israelis claim to have had
a kill ratio of 60 to 1. Our feasibility study has focused upon enhancing our record of air-to-air
Combat kills through more stringent and comprehensive selection procedures. Evidence is
presented which demonstrates that a program can be developed to select pilots who will be
effective in air-to-air Combat. From reviews of U.S. and foreign selection research dating from
World War II to the present and the assessment of pilot opinion from hundreds of aces, 45
factors were identified as potential predictors of fighter pilot combat effectiveness. Of these 45
factors, only 10 are adequately evaluated within current military selection programs upon
entrance into pilot training. Assessment of the remaining 35 untapped factors is within our
technological reach. In fact, many of these factors can be assessed by tests which are presently
available. We developed an Air Combat Effectiveness Study (ACES) program which would
establish selection test measures for virtually all of the factors identified as underlying fighter
pilot combat effectiveness. As part of the ACES program, selection test measures would be
validated against performance in air combat maneuvering ranges, thereby providing a method for
selecting fighter pilots during peacetime. We have emphasized selection for success in the
operational environment rather than success in training. Armed with these selection test scores
and an effectively executed validation program, researchers should, for the first time in history,
be able to specify a definitive profile of the ace fighter pilot.

Youtz, R. P. (1946). Objective measures of flying skill for the first 20 hours of Army pilot

(February 1947) for assignment to basic pilot training. Lackland Air Force Base, TX:
U.S. Air Force Human Resources Research Center.

Zaccaria, L., & Cox, J. A. (1952). Comparison of aviation cadets and student officers in primary
pilot training. Lackland Air Force Base, TX: USAF Human Resources Research Center.

(1956). Development of an interview procedure for USAF officer applicants. Lackland

Zachert, V., and Friedman, G. (1953). The stability of the factorial pattern of aircrew

This study consists of four factor analyses of the Army Air Forces Aircrew Classification
Batteries. The first was an analysis of the 1945 wartime battery, while the other three were
analyses of the 1947 postwar battery, consisting of essentially the same variables, but using
different samples. Eleven factors were found which had been identified and reported in previous
analyses. An additional factor, possibly an artifact, was identified as an age-education doublet.
The only factor which differed significantly in the analyses was pilot or flying interest. These
factor analyses show that the factorial content of the tests remains quite similar in both wartime
and postwar populations. The data reported in this study were collected as part of the United
States Air Force Human Resources Research and Development Program and described in
Research Bulletin 52-16. The opinions or conclusions contained in this report are those of the
authors. They are not to be construed as reflecting the view or indorsement of the Department of
the Air Force.


responsibility (OPR) for Air Force personnel testing programs since 1970. As the OPR, we are responsible for the development of plans, policies, and procedures necessary to monitor and implement all Air Force personnel testing programs. Whenever a new personnel test is implemented, AFMPC must establish the policies and procedures necessary for proper test administration and control of the data. Therefore, we are intimately involved in the planning and operational implementation of the Basic Attributes Test (BAT) which will be used as a selection and classification tool in the new Pilot Selection and Classification System (PSACs). Our role as the policy-maker will be to establish all testing requirements for the BAT under the Air Force test control system. We will provide guidance and regulate procedures for test administration, security of the test data, retest policy, and a variety of other issues.


The electroencephalogram (EEG) is used to screen pilot candidates for professional licensure irrespective of medical history in many European countries; applicants with paroxysmal abnormalities are excluded. The aim is to reduce the risk of later seizures in flight, which may cause accident or death, but there is no clear evidence that EEG screening can lead to any significant risk reduction. This is partly due to the low predictive value of the EEG, the low risk of seizure in healthy young adults, and the low risk of commercial aircraft accidents. Later-onset seizures, unrelated to the results of earlier screening, cannot be predicted or prevented by screening. Some benefit cannot be excluded, but may be difficult to demonstrate in prospective studies. Specialized screening of targeted populations, such as for photosensitivity in helicopter pilot candidates, may be justifiable; prospective studies with standardized recording and interpretation protocols are needed to assess this.