

Exploring the Equivalence of the Aviation Gender Attitude Questionnaire for South African and Australian Pilots: A Cross-Cultural Comparison

Leopold P. Vermeulen ^a; Pieter Schaap ^a; Jim I. Mitchell ^b and Alexandra Kristovics ^b

^a Department of Human Resources Management, University of Pretoria, Pretoria, South Africa

^b School of Management, University of Western Sydney, Sydney, Australia

Abstract

The aim of this study was to establish the differential item functioning and construct equivalence of the Aviation Gender Attitude Questionnaire (AGAQ) to determine the generalizability of the instrument across 2 national cultures (South African and Australian). Data were collected from 544 South African and 1,064 Australian aircraft pilots and analyzed by means of various analytical procedures. To establish the intercultural equivalence of the AGAQ, multivariate methods for item analysis and the comparison of factor structures were used. This included analysis of variance, exploratory factor analysis, Procrustean rotation, and the application of single- and multigroup confirmatory factor analysis. The item-level statistics revealed acceptable item discrimination values, satisfactory item and scale reliabilities, and the absence of item bias for the AGAQ across the 2 cultures. The scale-level statistics confirmed the equivalence of the AGAQ scales for both samples in the aviation contexts. It was unequivocally established that the AGAQ is a valid and culturally nonbiased measure that can be used to assess South African and Australian aviators' perceptions of gender-related pilot behavior.

Introduction

Aviation is the most global of industries. Professional aircraft pilots tend to follow a career path that takes them into employment opportunities that cross national and international boundaries. They take with them their training, qualifications, and experiences, and then build on these in diverse organizational and national cultural environments. They also carry with them their personal and professional attitudes, which then influence their behavior (Rollinson, Broadfield, & Edwards, 1998). Unfortunately, even today, many professional pilots still often display an attitude that is historically masculine (Davey & Davidson, 2000). This dominant attitude affects relationships on the flight deck, particularly when one of the pilots or a crew member is female. According to Sitler, Turney, and Wulle (1996) and Turney (1995), diverse crews must be supported and directed to adjust their attitudes, stereotypes, and behavior to meet the demands of a two-gender workplace.

Several researchers have documented various stereotypes, prejudices, and preconceptions regarding female aviators' abilities and the suitability of women for careers as professional pilots. See, for example, the research and publications of Bateman (1987), Davey and Davidson (2000), Mills (1998), Sitler (2004), Sitler et al. (1996), Turney (1995), Turney and Bishop (2004), and Wilson (2004). In a recent publication, Turney and Bishop (2004) stressed the importance of researching gender issues, barriers, and differences to promote effective crew teamwork and to improve the operational culture in aviation. However, in studies where cultural diversity is a factor, it is critical that the effect and influence of national, organizational, and professional cultures on the measurement of pilots' perceptions of crew behavior are considered (Helmreich & Merritt, 1998; Turney, 2004).

This article reports on some of the findings of a research project on gender issues in aviation coordinated by researchers at the University of Pretoria (South Africa) and the University of Western Sydney (Australia) in conjunction with researchers at the University of Tromsø (Norway). A questionnaire called the Aviation Gender Attitude Questionnaire (AGAQ) was developed as the basis for a cross-cultural study of perceptions of gender-related pilot behavior in general and, more specifically, of attitudes toward female pilots (Vermeulen & Wilson, 2002; Vermeulen, Wilson, & Mitchell, 2004). The study was done in several stages. In different stages, the questionnaire was distributed to private, commercial, and airline transport pilots in South Africa, Australia, the United States of America, and Norway. Results of the study have already been reported elsewhere (Kristovics, Mitchell, Vermeulen, Wilson, & Martinussen, 2006; Mitchell, Kristovics, Vermeulen, Wilson, & Martinussen, 2005; Vermeulen & Mitchell, 2007; Vermeulen, Vermeulen, & Le Kay, 2004; Vermeulen & Wilson, 2002; Vermeulen, Wilson, & Mitchell, 2004; Wilson, 2004).

Cross-cultural research by means of attitude surveys in the aviation context is not new. The Aerospace Crew Research Project at the University of Texas, Austin undertook an extensive study to establish any relationship(s) between a national culture and pilots' work values and attitudes (Helmreich, 1984, 1987; Helmreich & Merritt, 1998; Merritt, 1996, 2000; Merritt & Helmreich, 1996). During 1983 and 1984, Helmreich developed the Cockpit Management Attitude Questionnaire (CMAQ) and distributed it to flight crew members at several airlines and in military operations. The purpose of that survey was to gather data on attitudes concerning crew resource management (CRM) prior to and after completing formal CRM training. In 1993, the survey was updated and expanded to include issues involving multicultural dynamics.

The CMAQ was then refined and renamed the Flight Management Attitudes Questionnaire (FMAQ). Merritt (1993, 1996) used the FMAQ to explore cross-cultural similarities and differences with respect to pilots' attitudes toward flight management. Her findings indicated that there were strong and significant differences across cultural boundaries. Helmreich and Merritt (1998) published a substantive body of research results based on the data gathered by means of the FMAQ. They provided important evidence that culture (national and organizational culture, and professional attitudes and values) affects pilots' perceptions of flight management (Helmreich & Merritt, 1998).

In their publication "*Culture at Work in Aviation and Medicine*", Helmreich and Merritt (1998) reflected on the complexities of cross-cultural research methodology, including issues of culture-

sensitive data sampling, awareness of item- and culture-level response biases, decisions regarding the appropriate level of analysis, and the number of analyses necessary to fully understand the similarities and differences between and unique concepts among respondents of different cultures. The issues mentioned by Helmreich and Merritt refer to key concepts such as construct bias, method bias, and item bias and are related to scientific proof of the construct equivalence, measurement unit equivalence, and scalar equivalence of a measure used in cross-cultural research, as also discussed by Triandis and Berry (1980), Van de Vijver and Leung (1997b), and Van de Vijver and Rothmann (2004). According to Van de Vijver and Leung (1997a, 1997b), differential item functioning or item bias and construct equivalence should be established for measuring instruments in any setting where individuals from different cultural groups are compared in terms of a specific construct. If cultural influences are not accounted for, invalid conclusions could be reached regarding the constructs under review (Naudè & Rothmann, 2004).

Item bias in cross-cultural research refers to aspects of measurement artifacts at item level that might be the result of poor item translation, inadequate item formulation, or inappropriate items in a particular context (Van de Vijver & Leung, 1997b, p. 17). The assumption is that an item is unbiased if respondents from different culture groups with an equal standing on the theoretical construct underlying the instrument get the same expected score on the item. “Unbiased items define culture-common aspects of a construct and bias items denote cultural idiosyncrasies” (Van de Vijver & Leung, 1997b, p. 83).

Two types of bias can be distinguished, namely uniform and nonuniform bias (Van de Vijver & Leung, 1997b). Uniform bias refers to the effects of cultural differences on item scores that are more or less the same for all the score levels of the scale. More specifically, individuals from one cultural group might have consistently higher scores on an item than individuals from another cultural group that are on the same score level. Uniform bias can be explained in terms of an example of a metric scale used to measure the heights of people from different groups. If people's height is measured, but the measure consistently shows 5 cm more for one group than for the other, then the measure consistently “favors” one group over the other. The presence of a significant main effect of score level is usually taken as an indication of uniform bias.

Nonuniform bias refers to the interaction effects of a scale's score levels and culture on the variation of item scores of different groups across all score levels (Naudè & Rothmann, 2004; Van de Vijver, 2002; Van de Vijver & Leung, 1997b). In terms of the human height metric scale example used earlier, nonuniform bias is present when the measurement for one group is different from that for the other group at each scale level: 1 cm on the scale for one group is actually 1.1 cm for the other group, 2 cm for one group is actually 2.3 cm for the other group, 3 cm for one group is actually 3.6 cm for the other group, and so on. Cross-cultural comparisons involving human heights (taller or shorter) would therefore be distorted differently at each scale level. If item bias is not properly attended to, it might put the validity of cross-cultural comparisons at risk.

Construct equivalence is the most fundamental question when test responses are compared across different cultures, because it concerns the nature and essence of what is being measured (Owen, 1996). According to Reynolds (1982), constructs are equivalent when a test is shown to measure

the same hypothetical traits (psychological constructs) in one group as in another, or when it measures the same trait with similar degrees of accuracy. Bechger, Van den Wittenboer, Hox, and De Gloppe (1999) distinguished between two aspects of construct equivalence; first, the similarity of the theoretical mechanisms that underlie behavior in each group and, second, the similarity of the pattern of relations of a test to that of other measures in each group. Van de Vijver and Leung (1997a) defined construct equivalence—the focus of this article—as the similarity between the correlations (patterns of relations) of items of an instrument in respect of each group. If equivalence of the item discrimination values, test reliabilities, and factor structures of a measure can be demonstrated, the generalizability (Van de Vijver & Leung, 1997b) of the instrument can be claimed across the different cultures under review.

Proof of the generalizability of a measure across cultures is particularly crucial when a new instrument is developed and applied in a national cultural context that differs from the one in which the instrument was originally generated. Invariance of the psychometric properties of a new instrument cannot be assumed.

The aim of this study was therefore to examine the differential item functioning and construct equivalence of the AGAQ to establish the generalizability of this instrument for South African and Australian aircraft pilots.

METHOD

Research Design

According to Van de Vijver and Leung (1997b), “the interpretability of cross-culture research is optimized by a combination of substantive, methodological, and statistical considerations” (p. 145). In this study, a structure-oriented approach to cross-cultural comparison was followed and executed in terms of the generalizability study—the method, design, analyses, strengths, and weaknesses of a generalizability study are described by Van de Vijver and Leung (1997b, pp. 131-133). This approach required the application of analytical procedures that can detect item bias and confirm the construct equivalence of an instrument to be used on different culture groups. The data analysis of a generalizability study consists of two parts: The first part replicates the original analysis and the second part analyzes the agreement between the original and the new results (Van de Vijver & Leung, 1997b, p. 133).

However, ascertaining the generalizability of the factorial structure of an instrument across different national cultures is a complex process that requires multiple analytical strategies. In this regard, several multivariate methods of item analysis and the comparison of factor structures were used. This included analysis of variance, exploratory factor analysis (EFA), Procrustean rotation, and the application of single- and multigroup confirmatory factor analysis (CFA), as suggested by studies by Naudè and Rothmann (2004), and Meiring, Van de Vijver, Rothmann, and Barrick (2005).

Sample

To restrict method bias, the rationale used for the sampling in this study was to obtain a cross-section of pilots involved in flying fixed-wing aircraft for different operational purposes throughout South Africa and Australia and to ensure that the respondents in the different samples were all bona fide nationals. This decision was taken in view of Helmreich and Merritt's (1998, p. 234) contention that in cross-cultural comparisons, care must be taken to select culturally homogeneous samples.

Clear standardized instructions in English were provided to direct both the South African and the Australian participants in their completion of the questionnaire. A cover letter accompanied the questionnaire, stating the purpose of the research. It stressed voluntary participation and confidentiality. Anonymity was ensured by the fact that no names were asked for in the questionnaire.

The sampling frame for the South African survey was derived from a cross-section of a range of aviation organizations. A total of 1,800 questionnaires were distributed by hand to the pilots of 23 aviation organizations that were willing to participate in the research. This included the following:

- Seven flight training schools (150 surveys, 56 responses [37.3%]).
- Ten charter companies (350 surveys, 92 responses [26.3%]).
- Five airlines (1,100 surveys, 372 responses [33.8%]).
- The South African National Air Force (military; 200 surveys, 66 responses [33.0%]).

A total of 586 questionnaires (32.6%) were returned, of which 544 (30.2%) of the respondents were South African pilots by nationality. An earlier article by Vermeulen and Mitchell (2007) sets out more detailed information regarding the sampling procedures that were followed (that article discusses another aspect of the study reported here).

The sample for the Australian survey was selected on the basis of the membership of various pilot associations and interest groups in Australia. A total of 6,100 questionnaires were distributed to pilots via the newsletters of the following:

- The Aircraft Owners and Pilots Association (AOPA; 1,200 surveys, 168 responses [14.0%]).
- The Australian Federation of Air Pilots (AFAP; 2,300 surveys, 319 responses [13.9%]).
- The Australian and International Pilots Association (AIPA; 2,000 surveys, 544 responses [27.2%]).
- The Australian Women Pilots Association (AWPA; 600 surveys, 83 responses [13.8%]).

Access to military pilots was not granted. A total of 1,114 (18.3%) questionnaires were returned, of which 1,064 (17.4%) respondents were bona fide Australian pilots.

Participants

The research group represented a convenience sample of 544 South African and 1,064 Australian licensed aircraft pilots. The South African sample consisted of the following:

- 359 (66.0%) airline transport pilots (ATPs).
- 84 (15.4%) commercial pilots (CPLs).
- 64 (11.8%) flight instructors.
- 37 (6.8%) private pilots (PPLs).

Of this group, 513 (94.3%) were male and 31 (5.7%) were female. These numbers reflect the current low proportional representation of female pilots (5.9%) in commercial aviation licensed with the South African Civil Aviation Authority, SACAA for short (SACAA, 2007). The average age of pilots in the sample was 37.06 years ($SD = 11.056$), and their flying experience ranged between 1 and 43 years, with an average of 16.06 years ($SD = 10.874$). The mean number of flying hours of the sample was 6501.82 hr ($SD = 5626.125$). Only 112 (20.6%) of the South African respondents had any tertiary qualification(s), 74 (13.6%) a technical diploma, and 357 (65.7%) a high school certificate.

The Australian sample consisted of the following:

- 785 (73.8%) ATPs.
- 119 (11.2%) CPLs.
- 33 (3.1%) flight instructors.
- 126 (11.9%) PPLs.

There were 933 (87.7%) male and 131 (12.3%) female pilots. Their average age was 44.77 years ($SD = 11.987$). The flying experience of the pilots in the Australian sample ranged between 1 and 63 years, with an average of 22.075 years ($SD = 11.130$). The mean number of flying hours of the sample was 8681.946 hr ($SD = 6110.107$). Among the Australian respondents, 367 (35%) had tertiary qualification(s), 207 (19.8%) a technical diploma, and 473 (45.2%) a high school certificate.

Measuring Instrument

The AGAQ was used to survey the South African and Australian aviators' perceptions of pilot behavior related to gender. The AGAQ was originally developed by Vermeulen and Wilson (2002) and was refined by Vermeulen and Mitchell (2007) on a sample of 544 South African airline, commercial, military, and general aviation pilots. The original 72 items of the AGAQ were extensively analyzed by means of a series of EFAs. In these analyses, 34 of the items yielded a four-factor solution related to gender-related pilot behavior. CFA confirmed that the four-factor structure corresponded with the latent variables, and that the factors fit the data adequately (Vermeulen & Mitchell, 2007).

The four factors were labeled flying proficiency, safety orientation, flight confidence, and attrition of flight standards.

- Flying proficiency consisted of 16 items that refer to how proficient either gender is perceived to be at the task of piloting. The primary elements of this scale related to general piloting skills, accident proneness, learning ability, decision making in flying, and the pilot's perceived comfort level with regard to stick and rudder controls.
- Safety orientation incorporated seven items that related to perceptions about the level of safety awareness among pilots, pilots' attention to detail, risk-taking, and prudence.
- Flight confidence included seven items that deal with pilot confidence, assertiveness, taking charge, emotional stability, and coping with flight demands.
- Attrition of flight standards as a factor was defined by means of four items that referred to participant's perception that flight training standards and operational standards were being eroded by allowing female pilots some latitude when they are tested for their licenses and when they are given access to flying positions.

The AGAQ consists of two sections. Section A contains questions regarding anonymous biographical data, and Section B contains statements with regard to gender-related pilot behavior. The statements are rated on a 5-point Likert-type scale to assess the perceptions of participants at an approximate interval level. All the items that contained negative statements have been reverse coded to produce a measure where high scores indicate positive perceptions and low scores suggest negative perceptions toward female pilots.

Analytical Procedure

The equivalence of the AGAQ for the South African and Australian pilots was evaluated by means of various statistical analyses. This included the following:

- Examining the psychometric properties of the AGAQ at the item or instrument level for both samples (item analyses).
- Examining and comparing the internal consistency of the constructs of the AGAQ for both samples.
- Identifying the presence of anomalous items (item bias).
- Conducting structure-oriented techniques such as EFAs with target rotation.
- Conducting single-group and multiple-group CFA.

The Statistical Program for Social Sciences (SPSS) for Windows, Release 14 and the EQS structural equation program with maximum likelihood (ML) estimation (Bentler, 1995) were used to perform the required analyses.

Item analysis was performed to determine the characteristics of the items included in the four constructs of the AGAQ for each sample. The discrimination and reliability indexes of the individual items were calculated to establish the ability of each item to discriminate between high and low scores and to determine whether the items contributed to the underlying construct of the factors, as suggested by Gregory (2004).

To examine the internal consistency of the four factors of the AGAQ, Cronbach's alpha coefficients were calculated and compared for similarity across samples. Comparing the reliabilities of an instrument with respect to different cultural groups can be considered a

preliminary test for construct equivalence. The observation of reliability coefficients “can provide valuable clues about measurement accuracy and hence, the appropriateness of an instrument for cross-cultural comparison” (Van de Vijver & Leung, 1997a, pp. 271-272).

Item bias was identified using an extension of Cleary and Hilton's (1968) analysis of variance, as suggested by Van de Vijver and Leung (1997b). Two effects were tested by means of analysis of variance, namely the main effect of culture and the interaction of score level and culture, as suggested by Van de Vijver (2002). In cases where both the main effect of culture and the interaction of the score level and culture were significant, the item was regarded as biased, in terms of Van de Vijver and Leung's (1997b) criteria. With large samples, groups may be found to differ significantly with regard to a dependent variable; however, the effect of these differences may be small. Therefore partial eta square was used as a level of association for significant effects.

For the purposes of this study, both EFA and CFA methods were applied. This approach allowed us to take advantage of the relative strengths of each method in relation to those of other methods. Exploratory factor extraction procedures provide the most parsimonious explanation of the common variance underlying a correlation matrix. By contrast, confirmatory factor extraction procedures determine whether the data are consistent with a hypothesized factor structure. According to Van de Vijver and Poortinga (2002), confirmatory approaches might provide a statistically more adequate test of the structural equivalence of multiple-group data. Compared to EFA procedures, CFA is a more versatile tool to test for hierarchically linked hypotheses of cross-cultural invariance.

RESULTS

Before the main statistical analysis was undertaken, the presence of univariate and multivariate outliers was tested and the assumptions of normality were evaluated. Using z scores and Mahalanobi's distance, six South African and nine Australian cases were identified and deleted from the analysis due to extreme positive or negative values on most of the items or on three or more of the factors of the AGAQ. A small number of outliers (1 for every 100 respondents) are usually present in a normal distribution.

The responses of the remaining 538 South African and 1,055 Australian pilots on the 34 items of the AGAQ were examined to explore the quality of the cross-cultural data and to examine the psychometric properties of the AGAQ across groups.

Item Discrimination and Reliability Analyses

First, the item discrimination of the items was computed by comparing the item mean scores of the highest 25% and lowest 25% of responses on each scale, as suggested by Gregory (2004). To establish the differences between the item mean scores of the highest and lowest quartiles for each factor, the t test for independent groups was conducted. The item reliability index of each item was determined by computing the product of the item's standard deviation and the item's correlation with the total score (Gregory, 2004). The item analysis results for the AGAQ scales are set out in Table 1.

According to the item analysis (Table 1), all the items displayed acceptable item index values. The results indicated that all the items on each scale were able to discriminate significantly ($p < .001$) between high-scoring and low-scoring groups in both samples. The values of the item reliability index varied from .442 to .893 for the South African sample and from .474 to .856 for the Australian sample. The high item reliability values ($\geq .40$) served as evidence that the majority of the items exhibited acceptable internal consistency and produced a good distribution of responses across both samples. The results indicated that the item functioning was satisfactory for both groups and that the items contributed to the constructs that were measured.

The Cronbach's alpha coefficients for the four scales of the AGAQ were highly satisfactory, with reliability coefficients of .924, .813, .844, and .819 for the South African sample and .933, .824, .818, and .811 for the Australian samples, respectively. It should be noted that the alpha coefficients were all higher than the minimum level of $\alpha > .70$ recommended by Nunnally and Bernstein (1994). The statistical significance of differences between the Cronbach's alpha coefficients was determined by using Feldt's test statistic (Charter & Feldt, 1996). The significance value of Feldt's test statistic is sample-sensitive and increases with sample size. Although a small difference in the reliability coefficient for large independent samples might prove to be statistically significant, the difference could be so small that it has little practical significance. Feldt's statistic for testing the equality of two alpha coefficients indicated no difference ($p \geq .05$) between the reliability coefficients of the two groups. Each scale of the AGAQ was equally reliable and uniformly high for both the South African and Australian samples. This provided additional support for the psychometric adequacy of the AGAQ.

Item Bias Analysis

In the next step, the item bias for the individual items was computed. Bias for each of the 34 items was examined separately. Univariate analysis of variance was used to determine the main and interaction effect sizes of the culture and score levels on the scale items. The aim of the analysis was to test whether the average item scores were identical for respondents from the different culture groups with an equal score level, as proposed by Van de Vijver (2002). The item score was chosen as the dependent variable; nationality (two samples) and score levels were chosen as the independent variables. The total score on the different factors of the AGAQ was used to compose the score levels. The levels were obtained by using the equal grouping procedure of the SPSS. Seven to eight score levels with at least 50 respondents in each cell were created for each of the two national samples.

TABLE 1 Item Analysis: Differences Between the Mean Scores in the Lowest and Highest Quartiles and the Item Reliability Index for Each Item

<i>Scales/Items</i>	<i>Mean Score</i>		<i>Mean Score</i>		<i>Mean Score</i>		<i>Item</i>	
	<i>Lowest Quartile</i>		<i>Highest Quartile</i>		<i>Difference*</i>		<i>Reliability</i>	
	<i>SA</i>	<i>AUS</i>	<i>SA</i>	<i>AUS</i>	<i>SA</i>	<i>AUS</i>	<i>SA</i>	<i>AUS</i>

Note. SA = South African sample; AUS = Australian sample.

* $p < .001$.

Flight proficiency

Q6	2.993	3.264	4.472	4.368	-1.479	-1.104	.580	.474
Q13	2.408	2.678	4.221	4.192	-1.813	-1.514	.722	.644
Q17	2.908	3.112	4.362	4.427	-1.454	-1.315	.565	.539
Q19	2.743	2.830	4.229	4.372	-1.556	-1.542	.615	.661
Q22	2.415	2.685	4.063	4.220	-1.648	-1.535	.662	.683
Q26	2.618	2.794	4.260	4.277	-1.642	-1.484	.677	.635
Q30	2.862	3.130	4.386	4.524	-1.524	-1.394	.647	.583
Q34	2.197	2.681	3.913	4.217	-1.716	-1.535	.677	.655
Q37	2.671	2.830	4.039	4.296	-1.368	-1.466	.579	.622
Q38	2.915	2.841	4.315	4.323	-1.400	-1.482	.578	.643
Q41	2.382	2.757	3.827	4.073	-1.445	-1.316	.582	.576
Q42	2.770	3.036	4.260	4.405	-1.490	-1.369	.633	.585
Q46	2.572	2.772	4.173	4.348	-1.601	-1.576	.657	.660
Q55	2.559	2.790	4.016	4.055	-1.457	-1.265	.570	.544
Q57	3.217	3.217	4.378	4.348	-1.161	-1.131	.474	.482
Q63	2.553	2.953	4.370	4.463	-1.817	-1.510	.730	.641

Safety orientation

Q14	2.979	2.618	4.344	3.990	-1.365	-1.372	.542	.556
Q21	2.646	2.411	4.200	3.975	-1.554	-1.564	.627	.620
Q23	2.347	2.127	4.064	3.840	-1.717	-1.713	.692	.701
Q25	2.625	2.560	4.184	4.000	-1.559	-1.440	.629	.599
Q27	2.201	2.033	3.984	3.671	-1.783	-1.638	.684	.679
Q33	2.646	1.989	4.120	3.875	-1.474	-1.886	.578	.741
Q45	2.222	2.135	4.064	3.862	-1.842	-1.727	.711	.709

Flight confidence

Q15	1.890	2.022	3.539	3.747	-1.649	-1.725	.673	.703
Q24	1.890	2.405	3.898	4.136	-2.008	-1.730	.774	.695
Q39	1.929	2.015	3.797	3.837	-1.868	-1.822	.746	.745
Q43	1.864	2.011	3.875	3.891	-2.011	-1.880	.796	.790
Q49	2.058	2.494	3.828	3.861	-1.770	-1.367	.672	.589
Q59	1.935	2.279	3.898	4.021	-1.963	-1.742	.767	.707
Q61	1.994	2.030	3.055	3.124	-1.061	-1.094	.442	.493

Attrition of flight standards

Q40	1.954	2.491	4.146	4.577	-2.129	-2.086	.893	.828
Q52	2.035	2.572	4.364	4.555	-2.329	-1.983	.935	.827
Q56	1.843	1.898	4.046	4.201	-2.203	-2.303	.818	.856
Q69	2.355	2.871	4.473	4.757	-2.118	-1.886	.879	.782

The results of the item bias analysis and partial eta square values are reported in Table 2. According to Cohen's (1988) criteria the results show no practical significant partial eta square ($\eta^2 < .05$) values. Cohen regards a partial eta square as small when $\eta^2 < .06$.

This indicates that the mean scores of the two samples for the different score levels did not differ from zero in a systematic way. As such, the average responses of the two samples on the 34 items showed no uniform or nonuniform bias.

Exploratory Factor Analysis

Before the factor analysis was undertaken, it was determined whether or not the data of the two samples were suitable for factor analysis. To assess compliance with the distribution requirements, Bartlett's test of sphericity and the Kaiser-Meyer-Olkin measure of sampling adequacy (MSA) were used. The two diagnostic tests produced satisfactory results for both countries. The sample sizes for the South African and the Australian samples were both adequate. The MSA values were .948 and .961 for South Africa and Australia, respectively, and Bartlett's test confirmed ($p < .001$) that the data sets for both countries were highly suitable for factor analysis.

In this study, the principal axis factoring extraction method and promax rotation ($\kappa = 4$) were used to generate the factor solutions for the data. Based on Kaiser's (1961) criterion (eigenvalues larger than unity), four factors were extracted for both the South African and the Australian samples. The eigenvalues of the first four factors were remarkably similar for both countries. For the South African sample the values were 11.534, 3.812, 1.454, and 1.224; for the Australian sample they were 11.812, 3.667, 1.411, and 1.217.

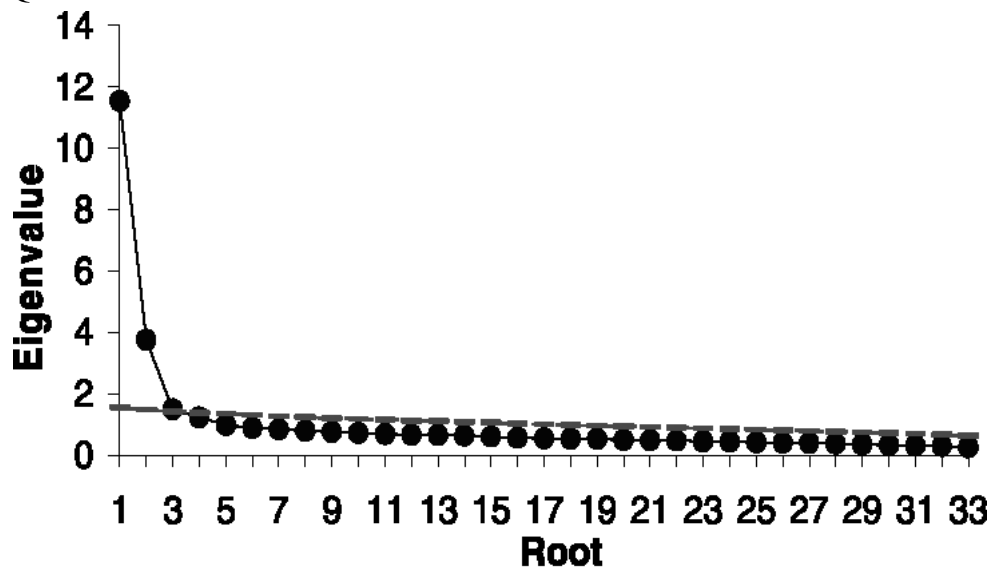
The parallel method of Horn (1965) and scree plots (Cattell 1966) were used to verify the number of significant factors with regard to each of the cultural groups. Horn's method entails contrasting the eigenvalues of a correlation matrix of random uncorrelated variables with those of the data set in question, based on the same sample size and the same number of variables. Factors of the matrix of interest that have eigenvalues greater than those of the random comparison matrix would be retained. According to Zwick and Velicer (1986), Horn's method provides the most accurate estimate of the number of true factors in a complex data set.

The scree plot and parallel analysis presented in Figures 1 and 2 confirmed that there were four significant constructs in the data set of each sample. Parallel analysis indicated a break in the scree plot between roots three and four. However, the curve of the eigenvalues of the random data set (the broken line) intersected the curve of the eigenvalues for the real data (the solid line) at root four, signifying four significant factors (Hayton, Allen, & Scarpello, 2004). To avoid underfactoring, it was decided to conclude with a four-factor solution. The four factors explained 53.011% of the variance in the factor space of the South African data and 53.255% of the variance in the factor space of the Australian data. The results of the principal factor extraction and the promax rotation on the AGAQ items for the South African and the Australian respondents are reported in Table 3.

TABLE 2 Item Bias Analysis of the Aviation Gender Attitude Questionnaire for the Two Cultures

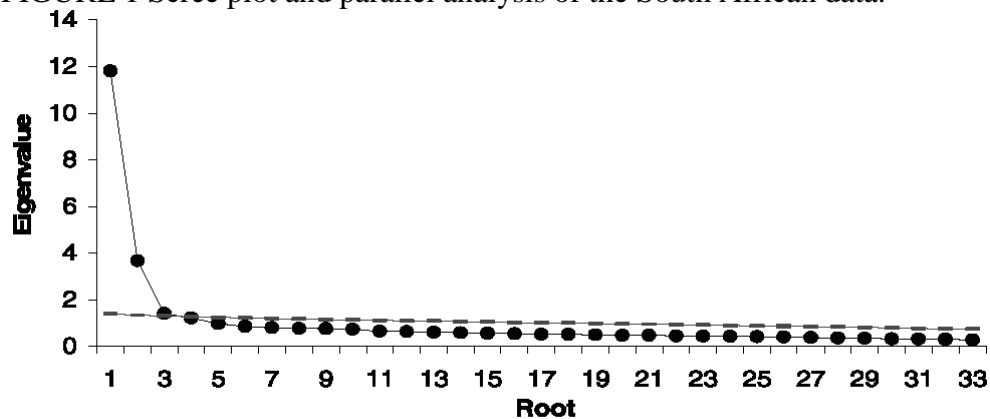
<i>Item</i>	<i>Tot_SS</i>	<i>Df_g</i>	<i>SS_g</i>	<i>F_g</i>	<i>Eta square_g</i>	<i>Df_i</i>	<i>SS_i</i>	<i>F_i</i>	<i>Eta square_i</i>
<i>Note.</i> g = culture; i = interaction; Tot_SS = correlated total sum of squares; Df_g = degrees of freedom for cultural groups; SS_g = summed square of the cultural groups; F_g = statistics for cultural groups; Eta square_g = partial eta square for cultural groups; Df_i = interaction (levels*culture); SS_i = sum of squares of the interaction (levels*culture); F_i = statistics for interaction (levels*culture); Eta square_i = partial eta square for interaction (levels*culture). $\eta^2 < .06$ (small effect).									
Flying proficiency									
Q6	1071.740	1	.312	.744	.000	7	5.121	1.745	.008
Q13	1433.787	1	.553	1.213	.001	7	4.009	1.257	.006
Q17	1152.468	1	.161	.371	.000	7	1.522	.502	.002
Q19	1411.062	1	.491	.998	.001	7	2.117	.615	.003
Q22	1272.983	1	3.494	9.282	.006	7	5.118	1.943	.009
Q26	1184.875	1	.264	.770	.000	7	.443	.184	.001
Q30	1012.212	1	.002	.009	.000	7	3.095	1.653	.007
Q34	1509.760	1	35.782	76.475	.046	7	2.873	.877	.004
Q37	1199.932	1	.316	.785	.000	7	9.158	3.254	.014
Q38	1247.135	1	13.572	31.874	.020	7	6.044	2.028	.009
Q41	1238.045	1	7.976	18.162	.011	7	2.858	.930	.004
Q42	1020.095	1	.990	3.444	.002	7	3.914	1.946	.009
Q46	1248.382	1	.190	.534	.000	7	2.292	.919	.004
Q55	1290.554	1	.151	.293	.000	7	3.356	.932	.004
Q57	910.465	1	1.551	4.353	.003	7	.779	.312	.001
Q63	1370.249	1	2.170	5.370	.003	7	.771	.273	.001
Safety orientation									
Q14	1327.873	1	6.571	12.886	.008	6	1.374	.449	.002
Q21	1268.280	1	.143	.340	.000	6	2.673	1.058	.004
Q23	1490.249	1	.060	.132	.000	6	3.108	1.143	.004
Q25	1373.876	1	2.579	4.915	.003	6	7.381	2.345	.009
Q27	1392.126	1	.664	1.563	.001	6	1.485	.582	.002
Q33	1669.930	1	7.721	14.042	.009	6	17.833	5.405	.020
Q45	1484.675	1	1.751	3.991	.003	6	1.643	.624	.002
Flight confidence									
Q15	1536.653	1	1.540	3.200	.002	6	3.938	1.364	.005
Q24	1666.254	1	13.688	28.586	.018	6	2.674	.931	.004
Q39	1706.107	1	1.647	3.166	.002	6	1.594	.511	.002

Q43	1684.005	1	1.611	3.653	.002	6	1.823	.689	.003
Q49	1370.983	1	.238	.502	.000	6	6.651	2.339	.009
Q59	1586.407	1	.171	.374	.000	6	3.562	1.302	.005
Q61	1279.842	1	2.543	4.299	.003	6	2.943	.829	.003
Attrition of flight standards									
Q40	2022.001	1	12.577	25.070	.016	6	3.553	1.180	.004
Q52	1783.783	1	.019	.052	.000	6	1.455	.661	.003
Q56	1914.254	1	10.235	21.735	.014	6	1.567	.554	.002
Q69	1548.321	1	.909	3.244	.002	6	1.243	.739	.003



Actual ◆-◆-◆ Random ----

FIGURE 1 Scree plot and parallel analysis of the South African data.



Actual ◆-◆-◆ Random ----

FIGURE 2 Scree plot and parallel analysis of the Australian data.

TABLE 3 Results of the Principal Factor Extraction and Promax Rotation on the Aviation Gender Attitude Questionnaire Items

<i>Item No.</i>	<i>Factor</i>	<i>Factor Loadings</i>	
		<i>South Africa</i>	<i>Australia</i>
Flying proficiency			
Q30	Male pilots make fewer mistakes when piloting than female pilots.	.845	.789
Q42	Female pilots are more likely to lose control following a stall than male pilots.	.808	.916
Q38	The most likely reason for accidents in which female pilots are involved is aircraft mishandling.	.724	.753
Q13	Male students learn piloting skills faster than female flight students.	.704	.682
Q6	Male pilots are less prone to incidents than female pilots.	.685	.564
Q19	The most likely reason for accidents involving women pilots is poor decision making.	.684	.564
Q37	Female students tend to experience difficulty in learning to use rudder control.	.682	.851
Q46	Male pilots are less likely to lose control when landing or taking off in a crosswind than female pilots.	.661	.794
Q26	Male pilots are less likely to make judgment errors in an emergency than female pilots.	.634	.711
Q57	Female flight students tend to experience more difficulty in learning radio communication procedures than male flight students.	.617	.695
Q17	Women often lack the endurance to complete flight school.	.545	.586
Q41	Male flight students tend to respond better to a “bounce” than female flight students.	.530	.692
Q63	Female pilots often lack the leadership ability required to pilot a multicrew flight.	.509	.440
Q22	Female pilots become fatigued sooner during stressful flights than male pilots.	.481	.635
Q34	Female pilots tend to worry too much about insignificant things when flying.	.450	.514
Q55	Female pilots' decision-making ability is as good in emergency situations as it is in routine flights.	.396	.484
Safety orientation			
Q23	Female pilots prefer to have information above the required minimum, more so than male pilots.	.729	.702
Q27	Female pilots prefer to have complete resolution to a problem before taking off, more so than male pilots.	.706	.736
Q45	When learning to fly, female pilots are more safety-oriented than male pilots.	.674	.686

Q21	Female flight students are more cautious than male flight students.	.621	.544
Q14	Female pilots tend to pay meticulous attention to detail.	.634	.637
Q21	Female flight students are more cautious than male flight students.	.621	.544
Q25	Male flight students take greater risks in flying than female flight students.	.532	.455
Q33	Women tend to learn to fly and preflight “by the book,” more so than men.	.494	.623
Flight confidence			
Q43	Male pilots tend to be more confident than female pilots.	.848	.766
Q39	Male pilots tend to be more assertive than female pilots.	.766	.696
Q15	Male pilots tend to “take charge” in flying situations more than female pilots do.	.671	.437
Q24	Male pilots are less nervous when piloting than female pilots are.	.557	.250
Q59	In a given situation, male pilots will often make decisions faster than female pilots do.	.544	.393
Q49	Male flight students tend to be less fearful of learning stall procedures than female students are.	.449	.265
Q61	Female flight students might feel intimidated when first learning to fly, more so than male flight students might.	.442	.518
Attrition of flight standards			
Q52	Flight program standards for the airlines and military have been relaxed to increase the number of female pilots.	.850	.821
Q69	Flight training standards have been relaxed so that it is easier for women to get their “wings.”	.780	.738
Q56	Supervisors of female pilots often let them get away with a little more because they are afraid of being branded sexist.	.620	.409
Q40	Professional female pilots are only in the positions they are in because airlines want to fill affirmative action quotas.	.538	.563

The results of the principal axis factor analysis performed on the AGAQ indicated that the factor structures for the South African and the Australian samples were very similar and that four factors had been properly determined for both groups.

A visual check of factor loadings indicated that Items 24 and 49 of the Australian sample seemed to cross-load on flying proficiency (F1) and flight confidence (F3). The factor loadings on the two factors were .347 (F1) and .250 (F3) for Item 24 and .503 (F1) and .265 (F3) for Item 49, respectively. These items are possible candidates for revision.

Van de Vijver and Leung (1997b) and McDonald (1985), however, suggested that Procrustean target rotation and Tucker's (1951) congruence coefficient should be used rather than visual assessment to determine the agreement between sets of factor loadings and to test for the stability of the factor structures. Accordingly, target rotation was used to determine the correspondence of

the factor structure of the AGAQ for the different culture groups. The factor loadings for the South African and the Australian samples were rotated to one target group. After the target rotation, factorial agreement was estimated using Tucker's coefficient of agreement. The Tucker's phi coefficients for the two culture groups are set out in Table 4.

Table 4 shows that the Tucker's phi coefficients for the South African and the Australian groups were all acceptable. Values of .90 and higher are seen as evidence for factorial similarity (Van de Vijver & Poortinga, 1994).

Confirmatory Factor Analysis

As a first step, a preliminary single-group CFA was conducted to test the extent to which the data fitted the measurement model with respect to each of the cultural groups. Byrne, Shavelson, and Muthén (1989) pointed out the necessity for determining the model fit for each of the groups separately before multiple-group comparisons are made. If a model fits very badly in a one-group analysis, it is likely that the model will not fit into a larger multisample analysis (Bentler, 1995).

EQS structural equation software was used to conduct CFA with a robust ML estimation procedure to test the extent to which the four-factor measurement model for the AGAQ fitted the data of the South African and the Australian samples. The Bentler-Bonnett Non-Normed Fit Index (NNFI), the Comparative Fit Index (CFI), the Bollen Incremental Fit Index (IFI), the root mean squared error of approximation (RMSEA), and the model chi-square were used as model fit indexes. The chance of obtaining a nonsignificant chi-square is extremely small with large sample sizes. The ratio of chi-square to degrees of freedom has therefore been proposed as a better measure, although it appears to suffer from arbitrary standards of interpretation (Kelloway, 1998). The NNFI, CFI, and IFI are considered relatively robust in terms of the effect of sample size (Bentler, 1990). Table 5 depicts the goodness-of-fit information for the robust ML model for the two samples.

TABLE 4 Construct Equivalence of the Aviation Gender Attitude Questionnaire for the Two Culture Groups

<i>Factor</i>	<i>Identity Coefficient</i>	<i>Proportionality Coefficient</i>
Flight proficiency	.97	.98
Safety orientation	.97	.97
Flight confidence	.91	.92
Attrition of flight standards	.90	.90

The Satorra-Bentler scaled chi-square statistic for the South African sample was highly significant, $\chi^2(521, N = 538) = 1048.43, p < .001$. This result indicated a poor model fit. However, the evaluation of fit on the basis of the χ^2/df ratio of 2:1 suggested that the model fits the data well—ratios between two and three can be interpreted as representing a good fit (Kelloway, 1998). The NNFI value (.917), the CFI value (.923), and the IFI value (.924) indicated that the model fits well—a value of = .90 is generally considered an indicator of a good model fit (Bentler, 1990; Bentler & Bonnett, 1980). The RMSEA value of .043 supported the

contention of a good model fit—according to Kelloway (1998), a value of less than or equal to .05 indicates an excellent fit. The 90% confidence interval of the RMSEA (.039-.047) confirmed the adequate fit of the structural model to the South African data.

With respect to the Australian group (Table 5), the Satorra-Bentler scaled chi-square, $\chi^2(521, N = 1,055) = 1728.35, p < .001$, was statistically significant; however, the χ^2/df ratio of 3.3:1 was within the acceptable range, indicating an adequate model fit. The NNFI value (.903), the CFI value (.909), and the IFI value (.910) exceeded the critical value of .90. The RMSEA value of .047 supported the notion of a good fit, and the 90% confidence interval of the RMSEA of .044 to .049 confirmed the adequate fit of the measurement model. The values of all the fit indexes indicated that the four-factor measurement model fit the data of the Australian sample well.

Overall, the single-group CFA produced very similar fit indexes for both samples, postulating models with the same number of factors for the South African and the Australian datasets. Multiple-group CFA was used to evaluate the equivalence of the factor model, the factor variances, factor covariances, the number of factors, and factor loadings for multiple groups. The equality of the covariance matrix is normally tested first, followed by subsequent analyses of nested models (Van de Vijver & Harsveld, 1994).

TABLE 5 Goodness-of-Fit Indexes for the Four-Factor Model of the Aviation Gender Attitude Questionnaire for South African and Australian Respondents

<i>Country</i>	χ^2	χ^2/df	<i>df</i>	<i>NNFI</i>	<i>CFI</i>	<i>IFI</i>	<i>RMSEA</i>
South Africa	1048.43	2.02	521	.917	.923	.924	.043
Australia	1728.35	3.31	521	.903	.909	.910	.047

Note. NNFI = Non-Normed Fit Index; CFI = Comparative Fix Index; IFI = Incremental Fit Index; RMSEA = root mean squared error of approximation.

TABLE 6 Overall Goodness-of-Fit Indexes for the Confirmatory Factor Analysis of the Multiple-Group Model

<i>Model</i>	χ^2	<i>df</i>	$\Delta\chi^2$	Δdf	<i>CFI</i>	<i>NNFI</i>	<i>IFI</i>	<i>RMSEA</i>
Null model	21599.481*	1,122						
Nested models								
Equal factor model	2800.034*	1,042	N/A	N/A	.914	.908	.914	.033
Equal factor variances	2807.156*	1,046	7.122	4	.914	.908	.914	.033
Equal factor covariances	2812.946*	1,052	5.790	6	.914	.908	.914	.032
Equal factor loadings	2867.642*	1,082	54.696*	30	.913	.910	.913	.032

In this study, a set of hierarchically nested models that successively increases the number of

equality constraints was used to test the equivalence of constructs with regard to the South African and Australian samples. The constraints were imposed in the sequence proposed by Vandenberg and Self (1993), starting with equality in the number of factors, followed by the equality of factor variances and factor covariances and ending with the equality of factor loadings. The incremental change in the chi-square with each constraint imposed provided an indication of the extent to which the constraints could be considered reasonable for the groups. The chi-square, CFI, NNFI, IFI, and RMSEA statistics were used as overall goodness-of-fit indexes for the nested models.

The results of the multiple-group analysis are presented in Table 6. The series of hierarchically nested models indicated statistically significant chi-square values for all the constraints. The chi-square is known to be sensitive to sample size and large samples tend to yield significant values with regard to small differences. The statistically nonsignificant change in the chi-square values of the test for equality of factor variance, $\Delta\chi^2(4) = 7.122, p > .05$, indicated that the factor variances could be considered equal for the South African and the Australian groups. The change in the chi-square values was statistically nonsignificant for equal factor covariances, $\Delta^2(6) = 5.790, p > .05$, but statistically significant for equal factor loadings, $\Delta^2(30) = 54.696, p \leq .05$, with respect to the South African and the Australian groups. Thus, factor covariances appear to be equal for the two groups, but differences do exist on the factor loadings. The CFI, NNFI, and IFI values in Table 6 for the equal factor model, the equal factor variance, equal factor covariance, and equal factor loading constraints were all above the critical value of .90, indicating a good model fit. The RMSEA values were all below .05 for all the constraints, indicating a good model fit and construct equivalence.

DISCUSSION

The principal aim of this study was to use multiple multivariate methods to test whether or not the construct of the AGAQ could be generalized across two national cultures.

The results of the preliminary analyses of the data indicated that the psychometric properties of the AGAQ can be considered acceptable for both the South African and the Australian samples. The discrimination and reliability indexes of the items on each of the four scales all meet the minimum requirements in terms of their distribution and their contribution to the construct that was measured. No statistical differences ($p > .05$) were found between the reliability coefficients of the scales for the two samples and no uniform or nonuniform bias was detected with regard to the responses on the items of the AGAQ. These results strongly support the presence of conceptual and measurement (metric) equivalence that are essential to perform cross-cultural comparisons, according to Berry and Lonner (1986) and Van de Vijver and Leung (1997b).

The EFAs of the data confirmed that there were four significant constructs in the data set of each sample. The results indicated a similar factor structure for both groups, in which the following factors were properly determined: flying proficiency, safety orientation, flight confidence, and attrition of flight standards. Although two items of the Australian sample cross-loaded on flying proficiency and flight confidence, Tucker's coefficient of factorial agreement presented acceptable evidence to prove factorial similarity, as described by Van de Vijver and Poortinga (1994). The agreement coefficients and proportionality indexes of .90 to .98 that were obtained

confirmed the correspondence of the four factors of the AGAQ for the South African and the Australian groups. This evidence of stability and invariance of the factor structure across different cultural contexts suggests that respondents from both countries hold similar beliefs regarding the construct that was measured.

The results of the CFA strongly supported the four-factor model, and provided conclusive evidence of the equality of factor variance and the equality of factor covariance of the AGAQ for the South African and the Australian samples. The chi-square statistic, which is sensitive to sample size, indicated a notable difference in the factor loadings of the items on the factors for the two samples. However, this difference did not seem to compromise the high level of construct equivalence between the South African and the Australian data. The CFI and IFI values (which are less sensitive to sample size) and the RMSEA values indicated a good model fit for all the constraints set. One can therefore infer that the four factors of the AGAQ were equivalent for the two samples and that the AGAQ is portable between South African and Australian culture. This might be the result of the fact that both samples (South African and Australian pilots) operate in Western aviation cultures where similar technical and human factor principles are applied. Both countries' pilots also communicate and are trained in English.

Overall, the results obtained in this study provide strong support for concluding that the English version of the AGAQ is a valid and culturally nonbiased measure to assess bona fide South African and Australian aviators' perceptions of gender-related pilot behavior and that the measure can be used to obtain relevant data to perform descriptive, comparative, and associational statistical analyses within and between the two countries. Further studies are, however, needed to investigate the cross-cultural generalizability of the AGAQ for aviators operating in non-English-speaking countries and even non-Western cultures.

Finally, this article is meant to encourage aviation psychologists to place more emphasis on applicable methods and data analysis to improve the effectiveness of their studies. Without a high degree of methodological sophistication, the application of appropriate research strategies, and the analyses of data by means of relevant statistical procedures, the results of cross-cultural research in aviation can be seriously flawed.

It is perhaps valuable to remind ourselves that in doing cross-cultural studies we need to be mindful of the complexity of our task and of our own expectations, which might cloud our objectivity:

The replicability of cross-cultural findings, often the Achilles heel of our empirical endeavors, will improve when we develop more sensitivity to our *partis pris* as psychologists and cross-cultural psychologists, when we put more emphasis on theory testing and development, and when our research is guided by appropriate methodological tools. (Van de Vijver & Leung, 2000, p. 48.)

ACKNOWLEDGMENTS

This article is part of a research project on gender issues in aviation, coordinated by Professor Dr. Leopold P. Vermeulen (University of Pretoria) and Dr. Jim Mitchell (University of Western

Sydney) in conjunction with Dr. Alexandra Kristovics (University of Western Sydney), Professor Dr. Monica Martinussen (University of Tromso), and Janice Wilson (United States).

REFERENCES

- Bateman, S. (Winter 1987) "The right stuff" has no gender. *Aerospace Power Journal* — Retrieved March 19, 2004, from <http://www.airpower.maxwell.af.mil/archronicals/apj/apj87/bateman.htm>
- Bechger, T. M. , Van den Wittenboer, G. , Hox, J. J. and De Gloppe, C. (1999) The validity of comparative educational studies. *Educational Measurement: Issues and Practice* **18** , pp. 18-26.
- Bentler, P. M. (1990) Comparative fit indexes in structural models.. *Psychological Bulletin* **107** , pp. 238-247.
- Bentler, P. M. (1995) *EQS structural equations program manual*. Multivariate Software , Encino, CA
- Bentler, P. M. and Bonnett, D. G. (1980) Significance tests and goodness of fit in the analysis of covariance structures.. *Psychological Bulletin* **88** , pp. 588-606.
- Berry, J. W. and Lonner, W. J. (eds) (1986) *Field methods in cross-cultural research*. Sage , Beverly Hills, CA
- Byrne, B. M. , Shavelson, R. J. and Muthén, B. (1989) Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance.. *Psychological Bulletin* **105** , pp. 456-466.
- Cattell, R. B. (1966) The scree test for the number of factors.. *Multivariate Behavior Research* **1** , pp. 245-276.
- Charter, R. A. and Feldt, L. S. (1996) Testing the equality of two alpha coefficients.. *Perceptual and Motor Skills* **82** , pp. 763-768.
- Cleary, T. A. and Hilton, T. L. (1968) An investigation of item bias.. *Educational and Psychological Measurement* **28** , pp. 61-75.
- Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*, 2nd ed., Lawrence Erlbaum Associates , Hillsdale, NJ
- Davey, C. L. and Davidson, M. J. (2000) The right of passage? The experiences of female pilots in commercial aviation.. *Feminism and Psychology* **10** , pp. 195-225. 13. Gregory, R. J. (2004) *Psychological testing: History, principles, and applications*, 4th ed., Pearson , Boston
- Hayton, J. C. , Allen, D. G. and Scarpello, V. (2004) Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis.. *Organizational Research Methods* **7** , pp. 191-

205.

Helmreich, R. L. (1984) Cockpit management attitudes.. *Human Factors* **26** , pp. 583-589.

Helmreich, R. L. (1987) Exploring flight crew behaviour.. *Social Behaviour* **2** , pp. 63-72.

Helmreich, R. L. and Merritt, A. C. (1998) *Culture at work in aviation and medicine: National, organizational and professional influences*. Ashgate, Aldershot, UK 18. Horn, J. L. (1965) A rationale and test for the number of factors in factor analysis.. *Psychometrika* **30** , pp. 179-185.
19. Kaiser, H. F. (1961) A note on Guttman's lower bound for the number of common factors.. *British Journal of Statistical Psychology* **14** — 1

Kelloway, E. K. (1998) *Using LISREL for structural equation modeling: A researcher's guide*. Sage , Thousand Oaks, CA

Kristovics, A. , Mitchell, J. I. , Vermeulen, L. P. , Wilson, J. and Martinussen, M. (2006) Gender issues on the flight-deck: An exploratory analysis.. *International Journal of Applied Aviation Studies* **6** , pp. 99-119.

McDonald, R. P. (1985) *Factor analysis and related methods*. Lawrence Erlbaum Associates , Hillsdale, NJ

Meiring, D. , Van de Vijver, A. J. R. , Rothmann, S. and Barrick, M. R. (2005) Construct, item and method bias of cognitive and personality tests in South Africa.. *South African Journal of Industrial Psychology* **31** , pp. 1-8.

Merritt, A. C. (October 1993) *The influence of national and organisational culture on human performance*. — Paper presented at the Australian Aviation Psychology Association Industry Human Factors Seminar, Sydney

Merritt, A. C. (1996) *National culture and work attitudes in commercial aviation: A cross-cultural investigation*. — Unpublished doctoral dissertation, The University of Texas at Austin, Austin, TX

Merritt, A. C. (2000) Culture in the cockpit: Do Hofstede's dimensions replicate?. *Journal of* **31** , pp. 283-301. — Cross-Cultural Psychology

Merritt, A. C. and Helmreich, R. L. (1996) Human factors on the flightdeck: The influences of national culture.. *Journal of Cross-Cultural Psychology* **27** , pp. 5-24.

Mills, A. (1998) Cockpit, hangars, boys and galleys: Corporate masculinity and the development of British Airways.. *Gender, Work & Organization* **5** , pp. 172-188.

Mitchell, J. I. , Kristovics, A. , Vermeulen, L. P. , Wilson, J. and Martinussen, M. (2005) How pink is the sky? A cross-national study of the gendered occupation of pilot. *Employment Relations Record* **5** , pp. 43-60.

Naudè, J. L. P. and Rothmann, S. (2004) The validation of the Maslach Burnout Inventory—Human services survey for emergency medical technicians in Gauteng. *South African Journal of Industrial Psychology* **30** , pp. 21-28.

Nunnally, J. C. and Bernstein, I. H. (1994) *Psychometric theory*, 3rd ed., McGraw-Hill , New York

Owen, K. Owen, K. and Taljaard, J. J. (eds) (1996) Test bias and test fairness.. *Handbook for the use of psychological and scholastic tests of the HSRC* pp. 77-95. Human Sciences Research Council , Pretoria, South Africa

Reynolds, C. R. Reynolds, C. R. and Gutkin, T. B. (eds) (1982) The problem of bias in psychological assessment.. *The handbook of school psychology* pp. 178-208. Wiley , New York

Rollinson, D. , Broadfield, D. and Edwards, A. (1998) *Organisational behaviour and analysis*. Addison-Wesley , Singapore

SACAA (July 2007) *South African licensed pilots' information*. [Statistics per license by gender, XLS]. Licensing office of the South African Civil Aviation Authority

Sitler, R. and Turney, M. A. (ed) (2004) Gender differences in learning to fly.. *Tapping diverse talents in aviation: Culture, gender, and diversity* pp. 77-88. Ashgate, Aldershot, UK

Sitler, R. , Turney, M. A. and Wulle, W. Hayward, B. J. and Lowe, A. R. (eds) (1996) Attitudes reflective of gender-based issues in the aviation and transportation workplace.. *Applied aviation psychology: Achievement, change and challenge*. pp. 332-342. Ashgate, Aldershot, UK — Proceedings of the Third Australian Aviation Psychology Symposium)

Triandis, H. C. and Berry, J. W. (eds) (1980) *Handbook of cross-cultural psychology: Methodology (Vol. 2)*. Allyn & Bacon , Boston

Tucker, L. R. (1951) *A method for synthesis of factor analysis studies (Personnel Research Section Rep. No. 984)* U.S. Department of the Army , Washington, DC

Turney, M. A. McDonald, N. , Johnston, N. and Fuller, R. (eds) (1995) Women's learning and leadership styles: Impact on CRM.. *Applications of psychology to the aviation system: 1* , pp. 262-268. Ashgate, Aldershot, UK — Proceedings of the 21st Conference of the European Association for Aviation Psychology

Turney, M. A. (2004) *Tapping diverse talents in aviation: Culture, gender, and diversity*. Ashgate, Aldershot, UK

Turney, M. A. and Bishop, J. C. Turney, M. A. (ed) (2004) Women's learning and leadership styles: Implications for air crews.. *Tapping diverse talents in aviation: Culture, gender, and diversity* pp. 61-75. Ashgate, Aldershot, UK

Van de Vijver, F. J. R. (2002) *Analysis of variance and item bias [Unpublished notes]*

Van de Vijver, F. J. R. and Harsveld, M. (1994) The incomplete equivalence of the paper-and-pencil and computerized versions of the General Aptitude Test Battery.. *Journal of Applied Psychology* **79** , pp. 852-859.

Van de Vijver, F. J. R. and Leung, K. Berry, J. W. , Poortinga, Y. H. and Pandey, J. (eds) (1997a) Methods and data analysis of comparative research.. *Handbook of cross-cultural psychology*, **1** , pp. 257-300. 2nd ed., Allyn & Bacon , Boston

Van de Vijver, F. J. R. and Leung, K. (1997b) *Methods and data analysis for cross-cultural research*. Sage , Thousand Oaks, CA

Van de Vijver, F. J. R. and Leung, K. (2000) Methodological issues in psychological research on culture.. *Journal of Cross-Cultural Research* **31** , pp. 33-51

Van de Vijver, F. J. R. and Poortinga, Y. H. Perris, C. , Arrindell, W. A. and Eisemann, M. (eds) (1994) Methodological issues in cross-cultural studies on parental rearing behavior and psychopathology.. *Parental rearing and psychopathology* pp. 173-197. Wiley , Chichester, UK

Van de Vijver, F. J. R. and Poortinga, Y. H. (2002) Structural equivalence in multilevel research.. *Journal of Cross-Cultural Research* **33** , pp. 141-156.

Van de Vijver, F. J. R. and Rothmann, S. (2004) Assessment in multicultural groups: The South African case.. *SA Journal of Industrial Psychology* **30** , pp. 1-7.

Vandenberg, R. J. and Self, R. M. (1993) Assessing newcomers' changing commitments to the organization during the first 6 months of work.. *Journal of Applied Psychology* **78** , pp. 557-568.

Vermeulen, L. P. and Mitchell, J. I. (2007) Development and validation of a measure to assess perceptions regarding gender-related behavior.. *International Journal of Aviation Psychology* **17** , pp. 75-96.

Vermeulen, L. P. , Vermeulen, P. and Le Kay, U. (September 2004) *The perceptions of flight instructors with regard to gender-related pilot behaviour*. — Paper presented at the 52nd International Congress on Aviation and Space Medicine, Sun City, South Africa

Vermeulen, L. P. and Wilson, J. (2002) *The measurement of perceptions with regard to gender-based issues in aviation*. — Unpublished research report, University of Pretoria, Pretoria, South Africa

Vermeulen, L. P. , Wilson, J. and Mitchell, J. I. D'Oliveira, T. (ed) (2004) The measurement of perceptions regarding gender-related pilot behaviour: An exploratory study.. *Aviation Psychology-Costs and Benefits. Proceedings of the 26th Conference of the European Association for Aviation Psychology (CD)* pp. 205-210. ISPA , Lisbon

Wilson, J. (2004) *Gender-based issues in aviation, perceptions towards female pilots: A cross-cultural analysis*. — Unpublished doctoral dissertation, University of Pretoria, Pretoria, South Africa

Zwick, W. R. and Velicer, W. F. (1986) A comparison of five rules for determining the number of components in complex datasets. *Psychological Bulletin* **99**, pp. 432-442.