

# Development and Validation of a Measure to Assess Perceptions Regarding Gender-Related Pilot Behavior

Leopold P. Vermeulen

James I. Mitchell

There are several stereotypes, prejudices and preconceptions regarding female aviators' abilities and the suitability of women for a career as professional aircraft pilots. This study reports on the development and validation of the Aviation Gender Attitude Questionnaire (AGAQ), which was designed to assess aviators' perceptions about gender-related pilot behavior. A total of 544 South African aircraft pilots participated in the validation of the questionnaire. A four-factor measurement model was established by using exploratory factor analysis and structural equation modeling. Confirmatory Factor Analysis confirmed a good fit between the multifactor solution for the AGAQ and the data. The four factors associated with perceptions about gender-related pilot behavior were Flying Proficiency, Safety Orientation, Flight Confidence and Flight Standards.

Although women have been flying aircraft for almost as long as men have, aviation is still largely a male-dominated sphere. Stereotypically masculine traits are more often perceived to be desirable in pilots than stereotypically female characteristics are, and a career in aviation is not deemed as feasible for women as for men (Helmreich & Merrit, 1998; Sitler, Turney & Wulle, 1996). Although women pilots have proven their worth and ability, prejudices such as "women should not fly" and "a woman in the cockpit is one less in the kitchen" still survive. As soon as issues are raised such as whether women should take up flying in a professional capacity or whether they should take up combat aviation roles, many people,

particularly men, continue to display negative attitudes (Barker, 1999a, 1999b). Wilkening (1973), as cited in Hunter (2005, p. 24) defined attitude as "a learned and relatively enduring perception, expressed or unexpressed, influencing a person to think or behave in a fairly predictable manner toward objects, persons, or situations".

Traditionally, "masculine" beliefs, values and attitudes (having the "right stuff") appear to dominate the aviation industry and the profession of pilot (Nagel & Wiener, 1988). The notion of "the right stuff, which originated in the United States military air force, is epitomized by aviators and astronauts who "possess extreme levels of confidence, assertiveness and competitiveness" and who are driven to achieve personal and organizational goals (King, Retzlaff & McGlohn, 1997, p. 695). This attitude and behavior was popularized in movies such as *The Right Stuff* and *Top Gun*. Although the nature of the "right stuff is being challenged (Bate-man, 1987), a cultural milieu permeates by this notion favors men in the recruitment of aviators, their induction and promotion within the aviation industry. Despite a history of legendary female aviators (Moolman, 1981; Cadogan, 1992; Yount, 1995), a career as a pilot still does not seem to attract as many women as men.

In South Africa, labor policies are encouraging more women to become professional pilots. Since the general democratic election in South Africa on 27 April 1994, attempts to make South Africa a more just society have increased. The government has realized that legislation is necessary to guide organizations in promoting justice and equity in the workplace. Hence, three Acts related to affirmative action have been promulgated, namely the Employment Equity Act (1998), the Promotion of Equality and the Prevention of Unfair Discrimination Act (2000) and the Preferential Procurement Policy Framework Act (2000). These Acts are designed to govern the promotion of social justice and to eradicate inequalities in the workplace, including the aviation industry (Bendix, 2001; Department of Labour, 2001). However, despite the fact that the aviation community in South Africa and abroad has instituted policies and procedures that meet equal opportunity and anti-discrimination legislation, the masculine culture of the industry still results in a workplace where female pilots experience sexism, discrimination, prejudice and hostility, and are exposed to inappropriate responses (Davey & Davidson, 2000).

Female pilots have to function in a profession steeped in patriarchal systems of dominance in terms of organizational power and the existing labor processes. In this domain, a certain construction of gender prevails, embedded in discourses that convey various contradictions, inconsistencies and ambivalences (Hearn, 1993). Arising from this discourse are gender-related perceptions that influence airline managers' and pilots' behavior, and that underpin the professional culture of pilots and the aviation industry. As concepts of masculinity dominate perceptions within the industry, female pilots are confronted with perceptions that are often not articulated, but that are acted out through associated behaviors. In particular, "sex-role stereotyping" (Nicholson, 1996) implies that often women are identified first in terms of their gender and only then in terms of their professional role (Sitler et al., 1996). From a broader perspective, these negative perceptions can lead to incidents or episodes of

gender bias, prejudice, and discrimination, in spite of legislators' best intentions and organizational initiatives designed to overcome such manifestations of unacceptable behavior (Davey & Davidson, 2000; Nicholson, 1996; Sitler et al., 1996).

One female captain with a large European airline expressed concern that male captains expected less of female co-pilots, or were reluctant to give the necessary corrective feedback to them. She believes that the flying standards of female pilots are being eroded (Johnson, 1995). Davey and Davidson (2000) have shown that the first women to join airlines in Europe experienced high visibility on the one hand, but were exposed to sexism, harassment, and isolation on the other. According to Davey and Davidson's (2000) study, many female pilots working for European commercial aviation companies reported that gender-related obstacles caused them to experience difficulties and distress. They encountered strong prejudices and preconceptions concerning their abilities and suitability for a career as professional pilot. Although the extent of these difficulties has declined (Davey & Davidson 2000), female aviators still come across sexism and prejudice. Such experiences have important implications for flight crew effectiveness and aviation safety.

Thus, because perceptions based on gender differences (real or alleged) have a pervasive and powerful influence on behavior, it is important to manage gender diversity properly, particularly in the aviation field. If stereotypes, negative attitudes and prejudices are not addressed in a formal setting such as Crew Resource Management (CRM) training, they can lead to problems in a multi-crew environment. According to Sitler et al. (1996) and Turney (1995), every member of the crew must be supported and directed to adjust his/her stereotypes, attitudes and behavior to meet the demands of a two-gender workplace.

In a recent publication, Turney and Bishop (2004) stress the importance of researching gender issues to promote effective crew teamwork and improve the operational culture in aviation. These authors recommended the development of a "survey to ascertain perceptions regarding gender-based differences and the awareness of barriers" that may cause gender-related bias in an aviation environment (Turney & Bishop, 2004, p. 71).

Very few scientific surveys about gender issues in aviation have been published so far. The ones that have appeared include those by Baker, Lamb, Grabowski, Rebok & Li (2001), Barker (1999a; 1999b), Bartsch and Hawkins (1996), Davey and Davidson (2000), Johnson (1995), King et al. (1997), McCarthy and McGinn (1995), Sitler et al. (1996), Sitler (2004), Tumey (1995; 2004), and Tumey and Bishop (2004). All these researchers have provided evidence to support their findings. However, not one of these publications refers to the existence of any standardized instrument that can be used to assess pilots' gender-related attitudes.

This study arose from this gap in the literature and research. The objective of this study was therefore to develop a valid and reliable instrument to assess female and male aviators' perceptions regarding gender-related pilot behavior.

## METHOD

### Measuring Instrument

Because there was no specific conceptual theoretical framework or model with regard to gender-related pilot behavior, surveys by various researchers (Baker, et al., 2001; Barker, 1999a; 1999b; Bartsch & Hawkins, 1996; Davey & Davidson, 2000; Johnson, 1995; McCarthy & McGinn, 1995; Sitler et al., 1996; Turney, 1995) were used as points of departure for developing the Aviation Gender Attitude Questionnaire (AGAQ).

The original questionnaire consisted of 72 statements (items). These statements refer to behavioral domains such as performance during flight training, flying skills, safety orientation, hazardous behavior, decision-making, leadership, professionalism, confidence, and flight standards (Vermeulen & Wilson, 2002; Vermeulen, Wilson & Mitchell, 2004). All the statements (except the demographic variables) were rated on a five-point Likert-type scale to measure the perceptions of participants at an approximate interval level. All the items that consisted of negative statements were reverse coded to produce a measure where high scores indicate positive perceptions and where low scores suggest negative perceptions towards female pilots.

### Procedure

A list of all the registered aviation organizations that employ aircraft pilots in South Africa was obtained from the South African Civil Aviation Authority (SACAA). Thirty organizations whose contact details were available were approached and asked to participate in the research on a voluntary basis. The South African Airline Pilot Association was asked to distribute the AGAQ questionnaire to their members, but the Association refused, as its executive did not want to get involved in research on gender issues in aviation.

The first author of this study distributed a total of 1800 questionnaires by hand to a cross-section of 23 aviation organizations that were prepared to participate in the research. All the organizations were involved in flying fixed wing aircraft for different operational purposes. They included seven flight training schools, ten charter companies, five airlines, and the South African Air Force (military). In order to maximize the sample size, the heads of Flight Safety and the CRM training sections of the organizations concerned were requested to forward the questionnaires, each accompanied by a prepaid return envelope, to their pilots. A covering letter accompanied the questionnaire, stating the purpose of the research and stressing voluntary participation and confidentiality. Anonymity was ensured by the fact that no names were asked for in the questionnaire. The completed questionnaires were collected both manually and via mail. A total of 586 (32.6 percent) questionnaires were returned, of which 544 (30.2 percent) were usable. In Table 1, details are given of the number of questionnaires distributed to and the response rate for each of the

different areas of flight operations.

### Participants

The research group represented a non-probability convenience sample of 544 South African licensed aircraft pilots. The sample consisted of 359 airline transport pilots (ATPs), 82 commercial pilots (CPLs) and 37 private pilots (PPLs), as well as 66 military pilots. Of this group, 513 (94.3 percent) were male and 31 (5.7 percent) were female. The small percentage of females was due to the fact that women have only recently started choosing professional flying as a career in South Africa. These numbers also reflect the low proportional representation of female pilots in commercial aviation in South Africa. The average age of the respondents was 37.06 years ( $SD = 11.056$ ). The respondents' number of years of flying experience ranged between one year 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 hours ( $SD = 5626.125$ ). Only 20.6 percent (112) of the respondents had any tertiary qualification(s).

TABLE 1  
Number of Questionnaires Distributed and Response Rate  
by Flight Operations

<i>Flight Operations</i>	<i>Number of Questionnaires</i>	<i>Response</i>	
		<i>Frequency</i>	<i>Percentage</i>
Flight training schools	150	56	37.3
Charter companies	350	92	26.3
Airlines	1,100	372	33.8
South African Air Force	200	66	33.0

## Analytical Procedure

A statistical analysis was performed using the Statistical Program for Social Sciences (SPSS) for Windows, Release 11 and Release 12. For the purposes of this study, both exploratory and confirmatory factor analytical methods were applied to "demonstrate that a measurement model derived from the internal structure of the construct and the operational design intentions of the instrument fits the data obtained on the instrument" (Theron 2006, p. 2).

Exploratory factor analysis (EFA) was used to explore the internal structure and validity of the AGAQ. First-order EFA was carried out by means of principal axis factoring and rotated using the promax procedure with Kaiser's Normalization to an oblique solution to generate a factor solution for the AGAQ. To assess compliance with the distribution requirements, Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy were used. In order to estimate the number of significant item factors, Kaiser's criterion (1961), Cattell's (1966) scree-plot and Horn's (1965) parallel analysis were used.

Horn's (1965) method of parallel analysis entails contrasting the eigenvalues of a correlation matrix of random uncorrelated items with the eigenvalues of the matrix of the actual data, based on equal sample size and an equal number of variables. Factors in or dimensions of the matrix are retained if "the *i*th eigenvalue from the actual data is greater than the *i*th eigenvalue from the random data" (O'Conner, 2000, p. 397). According to Hayton, Allen and Scarpello (2004), parallel analysis provides the most accurate estimate of the number of true factors in a complex dataset.

The Confirmatory factor analysis (CFA) procedure in Structural Equation Modeling (SEM) was used to validate the constructs of the measurement model (Garson, 2002; Gorsuch, 1997). CFA was performed to confirm that the observed variables (indicators) sort themselves into factors corresponding to the latent variables (Garson, 2002); and to test the overall quality of the solution and the specific parameters (factor loadings) that constitute the measurement model (Kelloway, 1998). CFA was done, using the EQS structural equation program with maximum likelihood (ML) estimation, described by Bentler (1995).

The EQS provided 11 Goodness of Fit Indices (GFIs). In this study, the following six GFIs were used to analyze the degree of likeness between the covariance matrices of the latent variables and the observed variables: the Model chi-square, the Root Mean Squared Error of Approximation (RMSEA), the 90 percent confidence interval of the RMSEA, the Bentler-Bonnett Non-normed Fit Index (NNFI), the Comparative Fit Index (CFI), and the Bollen Incremental Fit Index (IFI). The last three indices are "based on a ratio of the sum of the squared discrepancies to the observed variance" (Kelloway, 1998, p. 27) and they are normed to yield numbers that range from zero to one. Values above .9 indicate a good fit with the data. The RMSEA is based on an analysis of residuals and it estimates the lack of fit in a

model compared to a perfect model. Hair, Anderson, Tatham and Black (1998) regard RMSEA values between .05 and .08 as indicative of an acceptable fit.

The Model chi-square is the most common fit index test. Garson (2002) has referred to the chi-square as a "badness of fit" measure, in that a finding of significance ( $p < .05$ ) means that the given model's covariance structure is significantly different from the observed covariance matrix. One of the limitations of the Model chi-square fit index is that it is unlikely that an insignificant chi-square will be obtained in a large sample, even when the model fits the empirical data well (Krafft, Engelbrecht & Theron, 2004). To address this problem, the relative chi-square, also called the practical chi-square fit index, was devised (Kelloway, 1998). This fit index is directly related to the chi-square value and it is based on the ratio of the chi-square to the degree of freedom ( $\chi^2/df$ ) (Tabachnick & Fidell, 2001).

The item-discrimination index and the item-reliability index of the individual items were calculated to ascertain the ability of each item to discriminate between high and low scores and whether the items contributed to the underlying construct of the factors (Gregory, 2004). Cronbach alpha coefficients, as well as the average correlations between the items of each scale, were calculated to examine the internal consistency and unidimensionality of the retained factors of the AGAQ (Cortina, 1993; Clark & Watson, 1995).

## RESULTS

### Exploratory Factor Analysis

Before the main statistical analysis was undertaken, the sample size and the relationship between the responses to the items were examined to determine whether the data were suitable for factor analysis. The number of subjects (544) was larger than seven times the number of variables (72). The ratio of 7.5:1 of the sample size to the number of items met the guidelines established by Bryant and Yarnold (1995), and the terms of Lawley and Maxwell's (1971) "significance rule," which requires 51 more cases than the number of variables to support chi-square testing. The intercorrelation matrix revealed that underlying structures do exist. Both Bartlett's test of sphericity (19429.296;  $p < .001$ ) and the KMO measure of sample adequacy (.952) confirmed that the properties of the correlation matrix of the item scores were suitable for factor analysis, in terms of the guidelines recommended by Hair et al. (1998).

In the first round of Exploratory Factor Analysis, the responses on the 72 items of the AGAQ were inter-correlated and rotated by means of the promax procedure ( $\kappa = 4$ ) to an oblique solution. (Because of the limitations on the length of the article, the inter-correlation matrices are not reproduced here.) Based on Kaiser's (1961) criterion (eigenvalues larger than unity), 14 factors were extracted. The 14

factors explained 60 percent of the variance in the factor space of the data. The factor analyses yielded more factors in the real test space than was expected. This is probably due to the presence of differentially skewed items (Schepers, 1992). However, the scree-plot and parallel analysis presented in Figure 1 suggest that there are actually four significant constructs. Parallel analysis indicated a break in the scree-plot between roots three and four. The curve of the eigenvalues of the random data set (the broken line) intersects the curve of the eigenvalues for the real data (the solid line) at root four, signifying three significant factors (Hayton et al., 2004). Compared to Cattell's scree-test and parallel analysis, Kaiser's criterion overestimated the number of true factors for the dataset. To avoid underfactoring, it was sufficient to conclude with a four-factor solution. The four retained factors explained 42.57 percent of the total variance of the data.

Next, the items included in the four factors were scrutinized. All the items which had factor loadings less than .40, or which seemed to be exceedingly similar in content, were omitted. Only 42 items were retained and they were subjected to a second round of Exploratory Factor Analysis. The KMO value of .954 and the significant sphericity (10881.912,  $p < .001$ ) of the data were sufficient for factor analysis. Six factors with eigenvalues greater than one were extracted. These factors explained 49.74 percent of the total variance in the data. However, three of the items did not load on the four factors and five items loaded significantly across two or more factors. These eight items were disregarded and a third round of Exploratory Factor Analysis with promax rotation was performed. Bartlett's test of sphericity [ $\chi^2(561) = 8718.960, p < .001$ ] and the KMO measure of sample adequacy (.952) once more indicated that the properties of the correlation matrices of the 34 item scores were likely to factor well. (The inter-correlation matrices of the 34 items were also considered too large to reproduce here.) Inspection of the eigenvalues, the scree-plot and parallel analysis confirmed that four factors had been properly determined.

The results of the principal axis factor analysis for the 34 items of the AGAQ are summarized in Table 2. The factor loadings, percentage variance, squared multiple correlations and factor correlations are depicted.

Table 2 shows that four factors were extracted that explain 53.09 percent of the total variance in the data. Factors 1, 3 and 4 correlated highly with one another ( $r = .475$  to  $.694$ ), which might be an indication of an overlap in the factors, or of a lack of clarity in these factors. However, the Squared Multiple Correlations (SMCs) of .974 to .994 between the item scores and the factor scores indicated that all the factors were internally consistent and were well defined by the relevant items.



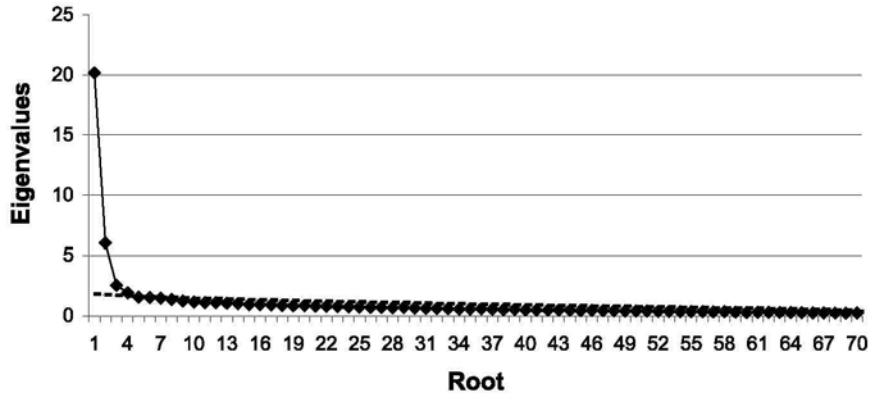


FIGURE 1 Scree Plot of the Actual and the Random Data (Actual  $\blacklozenge$ - $\blacklozenge$ - $\blacklozenge$  Random -----)

Scales were created for each factor and these were labeled according to the general content of their significant related items. The four scales or dimensions of gender-related pilot behavior were labeled Flying Proficiency, Safety Orientation, Flight Confidence and Flight Standards respectively.

### Confirmatory Factor Analysis (CFA)

CFA was used to test the extent to which the four-factor measurement model for the AGAQ fitted the data. First, the presence of univariate and multivariate outliers was tested and the assumptions of normality were evaluated. Using z-scores and Mahalanobi's distance, one case was identified and deleted from the analysis due to extreme values on two of the factors. In addition, Mardia's coefficient and the normalized estimate of the coefficient suggested that not all the measured variables distributed normally. Therefore, the robust maximum likelihood (ML) estimation, which adjusts the standard errors and offers the Satorra-Bentler scaled chi-square statistics, was employed, as suggested by Tabachnick and Fidell (2001). The Satorra-Bentler scaled chi-square is an adjusted chi-square statistic that attempts to correct for the bias introduced when data are markedly non-normal in distribution (Garson, 2002). Table 3 depicts the goodness-of-fit information for the robust ML model.

The results of the Satorra-Bentler scaled chi-square statistic [ $\chi^2(521, N = 543) = 1050.71, p < .001$ ] revealed a poor overall fit of the original measured four-factor

AGAQ model. Given the sample size and chi-square/df ratio, it would be hasty to accept a poor model fit based on the significance of the chi-square alone (Kelloway, 1998). The chi-square/df ratio was 2:1. Ratios between two and three can be interpreted as representing a good fit, and a value less than two as representing an over-fit (Kelloway, 1998). In terms of these guidelines, the measurement model seems to fit the data well.

All the other comparative fit indices indicated suitable values that were within the required limits for an acceptable model fit. The values for CFI, NNFI, and IFI were .923, .917 and .924 respectively. These values are all greater than .90, which is indicative of a model that fits well (Bentler, 1990; Bentler & Bonnett, 1980; Garson, 2002). The CFI, NNFI and the IFI are considered to be relatively robust to the effect of sample size (Bentler, 1990). The RMSEA value .043 supported the contention of a good model fit. A value of less than or equal to .05 indicated an excellent fit. The 90 percent confidence interval of the RMSEA (.039 to .047) confirmed the adequate fit of the measurement model to the data.

TABLE 2  
The Results of the Principal Factor Extraction and Promax Rotation  
on the AGAQ items

Item No.	Description	Factor Loading			
		1	2	3	4
30	Male pilots make fewer mistakes when piloting than female pilots.	.876			
42	Female pilots are more likely to lose control following a stall than male pilots.	.796			
38	The most likely reason for accidents in which female pilots are involved is aircraft mishandling.	.748			
19	The most likely reason for accidents involving women pilots is poor decision-making.	.730			
13	Male students learn piloting skills faster than female flight students.	.729			
6	Male pilots are less prone to incidents than female pilots.	.722			
26	Male pilots are less likely to make judgment errors in an emergency than female pilots.	.684			
37	Female students tend to experience difficulty in learning to use rudder control.	.674			
46	Male pilots are less likely to lose control when landing or taking off in a crosswind than female pilots.	.651			
57	Female flight students tend to experience more difficulty in learning radio communication procedures than male flight students.	.591			
17	Women often lack the endurance to complete flight school.	.580			
63	Female pilots often lack the leadership ability required to pilot a multicrew flight.	.531			
22	Female pilots become fatigued sooner during stressful flights than male pilots.	.519			
41	Male flight students tend to respond better to a "bounce" than female flight students.	.495			
34	Female pilots tend to worry too much about insignificant things when flying.	.482			
55	Female pilots' decision-making ability is as good in emergency situations as it is in routine flights.	.459			
45	When learning to fly, female pilots are more safety-oriented than male pilots.	.693			
23	Female pilots prefer to have information above the required minimum, more so than male pilots.	.689			
27	Female pilots prefer to have complete resolution to a problem before taking off, more so than male pilots.	.669			
21	Female flight students are more cautious than male flight students.	.650			

14	Female pilots tend to pay meticulous attention to detail.	.645	
25	Male flight students take greater risks in flying than female flight students.	.539	
33	Women tend to learn to fly and preflight "by the book," more so than men.	.493	
43	Male pilots tend to be more confident than female pilots.	.829	
39	Male pilots tend to be more assertive than female pilots.	.762	
15	Male pilots tend to "take charge" in flying situations more than female pilots do.	.664	
24	Male pilots are less nervous when piloting than female pilots are.	.558	
59	In a given situation, male pilots will often make decisions faster than female pilots do.	.537	
61	Female flight students may feel intimidated when first learning to fly, more so than male flight students might.	.482	
49	Male flight students tend to be less fearful of learning stall procedures than female students are.	.428	.851
52	Flight program standards for the airlines/military have been relaxed in order to increase the number of female pilots.		.761
69	Flight training standards have been relaxed so that it is easier for women to get their "wings."		.629
56	Supervisors of female pilots often let them get away with a little more because they are afraid of being branded sexist.		.551
40	Professional female pilots are only in the positions they are in because airlines want to fulfill affirmative action quotas.		
	Eigenvalues	11.527	3.820
	Percentage variance	33.904	11.234
	Squared multiple correlation	.994	.974
	Factor correlations		
	Flying proficiency	1.000	
	Safety orientation	-.159	1.000
	Flight confidence	.694	-.397
	Flight standard	.627	.017
			1.000

TABLE 3  
Goodness-of-Fit Indexes for the Four-Factor Model of the AGAQ

$\chi^2$	$\chi^2/df$	<i>df</i>	<i>NNFI</i>	<i>CFI</i>	<i>IFI</i>	<i>RMSEA</i>
1050.71	2.02	521	.917	.923	.924	.43

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

### Item Analysis of the Four Factors of the AGAQ

Next, the items were subjected to an item analysis to provide evidence that the items produced a statistically acceptable dispersion of scores and that the items contributed to the overall measurement goal of the AGAQ. For this purpose, the item-discrimination index and item-reliability index were computed for each item included in the four scales.

According to Gregory (2004, p. 131), "an effective test item is one that discriminates between high scores and low scores on the entire test" or scale. As suggested by Gregory (2004), the item-discrimination index was computed by comparing the item mean scores of the highest 25 percent and lowest 25 percent responses for each scale. The t test for independent groups was conducted to establish differences between the item mean scores of the highest and lowest quartiles for each factor. The item-reliability index of each item was determined by computing the product of the item's standard deviation and item's correlation with the total score (Gregory, 2004). Table 4 shows the results of the item-discrimination analysis and the item-reliability index for every item.

According to the item analysis (Table 4), almost all the items displayed acceptable item index values. The results indicated that every one of the items on each scale was able to discriminate significantly ( $p < .001$ ) between high- and low-scoring groups in the present sample. The values of the item-reliability index varied from .338 to .779 and they served as evidence that the majority of the items exhibit acceptable internal consistency and produce a good distribution of responses. Based on the results mentioned above, all the items of the four scales were retained to be included in the AGAQ to measure and assess pilots' perceptions of pilot behavior related to gender.

In addition, the mean, standard deviation, skewness, kurtosis, mean inter-item correlation and coefficient alpha were calculated for the sample scores on the four scales of the AGAQ. The results are set out in Table 5.

From Table 5 it is evident that the summated scores of the sample on the four scales are approximately normally distributed, with a tendency towards negative skewness and platykurtic distributions. The assumption of normality requires that the key statistics, skewness and kurtosis be less than 2.5 times the standard error (Morgan & Griego, 1998).

TABLE 4  
Item Analysis: Differences in the Mean Scores between Lowest  
and Highest Quartiles and the Item-reliability Index for Each Item

<i>Items</i>	<i>Lowest Quartile</i>	<i>Highest Quartile</i>	<i>Mean Difference*</i>	<i>Item Reliability Index</i>
Flight proficiency				
6	2.968	4.485	-1.517	.558
13	2.396	4.239	-1.842	.681
17	2.883	4.377	-1.494	.542
19	2.727	4.315	-1.588	.584
22	2.597	4.277	-1.689	.643
26	2.838	4.400	-1.680	.651
30	2.188	3.915	-1.562	.637
34	2.649	4.062	-1.727	.620
37	2.890	4.331	-1.412	.554
38	2.377	3.831	-1.441	.562
41	2.747	4.277	-1.454	.534
42	2.558	4.192	-1.530	.652
46	2.539	4.008	-1.634	.629
55	3.208	4.392	-1.469	.524
57	2.539	4.385	-1.185	.446
63	2.968	4.485	-1.846	.689
Safety orientation				
14	2.973	4.344	-1.371	.419
21	2.642	4.200	-1.558	.527
23	2.351	4.064	-1.713	.579
25	2.601	4.184	-1.583	.511
27	2.182	3.984	-1.802	.579
33	2.615	4.120	-1.505	.443
45	2.189	4.064	-1.875	.597
Flight confidence				
15	1.878	3.573	-1.694	.594
24	1.878	3.924	-2.045	.687
39	1.917	3.817	-1.900	.643
43	1.853	3.901	-2.048	.726
49	2.058	3.855	-1.797	.578
59	1.923	3.924	-2.001	.680
61	1.994	3.076	-1.083	.338
Flight standards				
40	1.948	4.152	-2.204	.676
52	2.023	4.375	-2.352	.779
56	1.833	4.027	-2.193	.641
69	2.339	4.482	-2.143	.748

\* $p < .001$ .

The reliability of the four scales was highly satisfactory, with Cronbach alpha coefficients varying between .82 and .93. The alpha coefficients surpassed the minimum level of .70 recommended by Morgan and Griego (1998). The mean inter-item correlations of three scales (Flying Proficiency, Safety Orientation and Flight Confidence) were considered acceptable. All the mean inter-item correlations were within the desirable range of .15 to .50 suggested by Clark and Watson (1995). The high mean inter-item correlation (.54) obtained for Factor 4 (Flight Standards) may be the result of the specificity of the target construct. According to Clark and Watson (1995), a much higher mean inter-item correlation (.40 to .50) can be expected when one is measuring a narrow construct. The scores on the four scales of the AGAQ seem to satisfy the requirements of homogeneity and unidimensionality. Together, these results demonstrated that the items of each scale assess the underlying factors or constructs associated with the measurement (Cortina, 1993; Clark & Watson, 1995).

## DISCUSSION

A review of the literature available revealed that, so far, no substantive research had been done on the development and validation of a questionnaire to allow the aviation industry to determine pilots' attitudes towards female aviators.

TABLE 5  
Descriptive Statistics and Reliability of the Scales of the AGAQ

<i>Descriptive Statistics</i>	<i>Flying Proficiency</i>	<i>Safety Orientation</i>	<i>Flight Confidence</i>	<i>Flight Standards</i>
<i>M</i>	55.525	23.525	18.877	12.322
<i>SD</i>	10.235	4.541	5.009	3.555
Skewness	-.248	-.431	.529	-.112
Sk error	.105	.105	.105	.105
Kurtosis	.031	-.343	-.278	-.650
Ku error	.209	.209	.209	.209
<i>r(M)</i>	.45	.33	.45	.54
Alpha	.93	.82	.85	.82

The primary objective of this study was therefore to develop a valid and reliable instrument to assess female and male aviators' perceptions regarding gender-related pilot behavior.

A questionnaire named the Aviation Gender Attitude Questionnaire or AGAQ was devised to survey aviators' perceptions regarding gender-based pilot behavior. A total of 72 items were initially included in the AGAQ. After three applications of exploratory factor analysis, 34 of the original items yielded a four-factor solution. CFA confirmed that the observed factors corresponded with the latent variables, and that a good fit had been achieved between the multifactor solution for the AGAQ and the data.

The four factors or dimensions associated with gender-related pilot behavior were labeled Flying Proficiency, Safety Orientation, Flight Confidence and 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, the second dimension, was made up of seven items that related to perceptions about the level of safety awareness among pilots, their attention to detail, risk-taking and prudence. The third dimension was labeled Flight Confidence and included seven items that deal with pilot confidence, assertiveness, taking charge, emotional stability and coping with flight demands. The fourth dimension looked at perceptions that standards had been compromised to accommodate female pilots and it was labeled Flight Standards. The content of the four items that defined this scale referred to beliefs that flight training standards and operational standards were being eroded by allowing female pilots latitude when they are tested for their licenses and when they are given access to flying positions.

Several researchers have directly or indirectly confirmed the four dimensions of the AGAQ. Literature cited by Morgan (2004) in the field of women in the professions indicated that although women are sometimes slower than men in completing their work, they work more safely and more accurately, and deliver a superior product. Sitler (2004, pp. 82-85), an academic and a flight Instructor, reported that she had observed and studied women pilots over a period of thirty years and had found that female flight students

- "May tend to be slower to gain confidence;
- May tend to be more apprehensive about first solo;
- May be quicker to grasp instrument flight;
- Have far fewer fatal aircraft accidents;
- May tend to learn a procedure correctly, and be more consistent, rarely varying from the procedure;
- May be more fearful of stalls, spins and other unusual attitudes;



- Tend to operate the controls of the airplane with more smoothness and skill;
- May have less technical background from past experience;
- Rarely fly into dangerous weather;
- Rarely show-off for spectators;
- Accidents usually involve touchdown or taxiing".

Baker, et al. (2001) at the Bloomberg School of Public Health have found that male pilots who fly general aviation (private) aircraft in the United States of America are more likely to crash due to inattention or flawed decision-making, while female pilots are more likely to crash because they mishandle the aircraft. The researchers have found that loss of control on landing or takeoff was the most common source of crashes for both sexes, leading to 59 percent of crashes by female pilots and 36 percent of crashes by male pilots.

On the basis of a qualitative and quantitative analysis of data collected from 33 male and female airline, corporate and military pilots, Turney (1995) and Turney and Bishop (2004) identified a number of gender-related differences and issues. The pilots they interviewed generally expressed a sense of differences between men and women pilots' learning and leadership styles. Male pilots "were reported to be more task-oriented and to exhibit more confidence as crew members. Female pilots were reported to be better communicators, to exhibit more sensitivity to people and to be good negotiators. Women were also reported to work harder at learning technical information to compensate for any possible lack of background in mechanical training" (Turney, 1995, pp. 264). Barriers to women's integration into the flight crews and areas of conflict included a lack of understanding of gender differences, a "macho" pilot image, quota hiring, pregnancy, leadership roles, attitudes that only males are professionals and the view that mistakes made by women are gender weaknesses rather than personal mistakes (Turney, 1995, pp. 264-265).

Taken together, the research reported by Baker et al. (2001), Morgan (2004), Sitler (2004), Turney (1995) and Turney and Bishop (2004) have all provided evidence of observed gender differences that parallel several of the items that define the scales of the AGAQ.

### Implications and Practical Application

The statistical analysis of the AGAQ has indicated that the questionnaire has acceptable factorial validity, internal consistency and unidimensionality. Hence aviation human factor specialists and aviation psychologists can use the instrument with confidence to gather valid and reliable data about gender-related attitudes held by aircraft pilots in South Africa.

The four scales of the AGAQ also demonstrate adequate power to discriminate between the higher and lower scoring groups. During the process of item analysis and the examination of the distribution of the responses, it was noticed that a mean-

ingful number of the present sample expressed negative judgments about female pilots. Based on Gregory's (2004) description of the percentage that constitutes a lower-scoring group, the number of pilots who scored at the 33 percentile and below was calculated. The number of respondents in the low-scoring groups were 194 for Flying Proficiency, 206 for Safety Orientation, 215 for Flight Confidence and 217 for Flight Standards. A relatively high number of South African pilots thus seem to hold attitudes that can potentially be detrimental to crew teamwork and the operational culture in aviation.

Contemporary aviation organisations have become increasingly reliant on knowledge and on the management of information to ensure optimal performance and aviation safety. To be able to objectively measure and assess gender related attitudes, stereotypes and prejudices, an instrument like the AGAQ may be able to help researchers, human factor specialists, aviation psychologists, instructors and employers in understanding and managing key concepts and issues associated with gender-related diversity in the aviation industry.

First, the AGAQ can provide the data needed to convince management to address any unjust attitudes, stereotypes and prejudices in the aviation industry. For example, male pilots may label an increase in the number of female professional pilots in South Africa an affirmative action effort aimed at filling a quota. Such a perception tends to breed resentment and an unfounded belief that the female pilots who are hired are incompetent and that standards are compromised. Because the effective assimilation of females as pilots into the industry is important for crew teamwork and flight safety (Turney & Bishop, 2004), it is important to address such negative perceptions.

Second, data collected by means of the AGAQ can be disseminated for discussion, to create an awareness of diversity in operational contexts. Turney and Bishop (2004) suggested that, to improve the operational culture in aviation, air crews, check pilots and instructors should receive regular and current information about research regarding gender differences. Knowing and understanding people's attitudes and perceptions toward pilot behavior related to gender can do much to advance positive interactions and communication among diverse flight crews.

Third, knowledge and greater understanding of the attitudes and beliefs relevant to gender will also benefit aviation-training organizations to develop better-targeted training material. This applies to Human Factor (HF) training and Crew Resource Management (CRM) training, where topics such as gender differences and diversity management must be addressed to improve CRM training and promote a gender-sensitive multi-crew environment (Turney, 1995). The AGAQ may also provide valuable information that can be used as input in the Diverse Workforce Model. This model was developed by Karp and Turney (2004) to support aviation education and training institutions in their education efforts and to enhance successful learning within a diverse workforce.

To summarize, the AGAQ can be a valuable tool to capture data to

- Confirm or reject the existence of gender-related biased perceptions among aviation personnel;
- Make pilots aware of their perceptions of gender-based differences and the way such perceptions may promote compatibility or discord;
- Improve and promote better understanding and communication between female and male pilots;
- Improve and advance gender sensitivity and diversity training in aviation education and crew resource management (CRM) programs; and
- Develop action plans to address gender bias, prejudice and discrimination in aviation (Vermeulen et al., 2004).

### Limitations and Future Research

A shortcoming of this study was the small number of female participants in the South African survey. Institutions in Norway, Australia and the United States have since participated in a broader survey, and this will result in the inclusion of more female pilots in future reported studies.

Cross-cultural comparisons of perceived gender-related pilot behavior will be possible when the research data for the participants in other countries has been processed. According to Van de Vijver and Leung (1997), structural equivalence and item bias should be computed 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 regarding the constructs under study could be made (Naude & Rothmann, 2004). Additional research to determine the factorial validity, structural equivalence, item bias and internal consistency of the AGAQ in measuring pilots' attitudes across cultures is currently being done on the basis of the results coming in from the Norwegian, Australian and United States data.

Future research also needs to explore the relationship and main effects of pilot-related factors such as gender, age, flying time, type(s) of license, area(s) of operation, level(s) of command, experience of flying with female pilots, and so on, on the perceptions of pilots regarding gender-related pilot behavior. This could add value to our understanding of the dynamics of pilots' attitudes towards female pilots. Furthermore, a qualitative dimension to the research might be valuable: an open-ended opportunity was provided to the pilots to include any comment(s) with regard to stereotypes and/or prejudices (either positive or negative) about female pilots, and a qualitative analysis of this data can enhance knowledge on the key issues that may influence crew relationships and performance in a diverse multi-crew environment.

Finally, the process that was described in validating the measuring instrument represents the initial rather than the final steps in the development of the AGAQ. There is strong evidence that the items of the AGAQ are related to two broad factors. An examination of the initial unrotated factors and the inter-correlations of a significant number of the rotated factors suggest that a two-factor measurement model exists. To delineate the construct boundaries accurately and to ensure optimal scale development, second order factor analysis should be considered (Schepers, 1992).

It is perhaps good to conclude with the following caveat: "In spite of what has been learned, cultural, social and organizational factors affecting women's employment in the aviation industry remain largely undefined. If effective air crew diversity is the goal, then it becomes necessary to continue to study differences and integrate diversity in the work place" (Turney & Bishop, 2004, p. 72).

## ACKNOWLEDGEMENT

This article is part of a research project on gender issues in aviation co-ordinated by Professor Dr. Leopold P. Vermeulen (University of Pretoria) and Dr. James I. Mitchell (University of Western Sydney) in conjunction with Dr. Alexandra Kristovics (University of Western Sydney), Professor Dr. Monica Martinussen (University of Tromsø) and Dr. Janice Wilson (Independence Air, USA).

## REFERENCES

- Baker, S. P., Lamb, M. W., Grabowski, J. G., Rebok, G., & Li, G. (2001). Characteristics of general aviation crashes involving mature male and female pilots. *Aviation, Space, and Environmental Medicine*, 72(5), 447–452.
- Barker, D. (1999a). Women pilots in operational combat: an emotive issue. *Salut*, August, 30-34.
- Barker, D. (1999b). Women pilots in operational combat: an emotive issue. *Salut*, September, 30-34.
- Bartsch, J., & Hawkins, R. M. F. (1996). The influence of sex, attractiveness and qualifications on hiring practices in aviation. In B. J. Hayward, & A. R. Lowe (Eds.), *Applied aviation psychology: Achievement, change and challenge. Proceedings of the Third Australian Aviation Psychology Symposium* (pp. 343–350). Aldershot, England: Ashgate.
- Bateman, S. (1987). "The right stuff" has no gender. *Aerospace Power Journal*, Winter, 1987-88. Retrieved March 19, 2004, from the World Wide Web: <http://www.airpower.maxwell.af.mil/airchronicals/apj/apj87/bateman.htm>.
- Bendix, S. (2001). *Industrial relations in South Africa* (4th ed.). Lansdowne, South Africa: Juta.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–247.
- Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.
- Bentler, P. M., & Bonnett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588–606.

- Bryant, F. B., & Yarnold, P. R. (1995). Principal component analysis and exploratory and confirmatory factor analysis. In L. G. Grimm, & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 99–136). Washington, DC: American Psychological Association.
- Cadogan, M. (1992). *Women with wings*. Chicago: Academy Chicago Publishers.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behaviour Research*, *1*, 245–276.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104.
- Davey, C. L., & Davidson, M. J. (2000). The right of passage? The experiences of female pilots in commercial aviation. *Feminism and Psychology*, *10*(2), 195–225.
- Department of Labour. (2001). Employment Equity Report. Pretoria, South Africa.
- Employment Equity Act, 55 (1998). The South Africa Department of Labour. Retrieved August 29, 2004, from the World Wide Web: [http://www.labour.gov.za/legislation/original\\_act.jsp?legislationDetail\\_id=5954](http://www.labour.gov.za/legislation/original_act.jsp?legislationDetail_id=5954).
- Garson, G. D. (2002). An online textbook. PA765. Retrieved May 26, 2002, from the World Wide Web: <http://www2.chass.ncsu.edu/garson/p9765/statnote.htm>.
- Gorsuch, R. L. (1997). Exploratory factor analysis: its role in item analysis. *Journal of Personality Assessment*, *68*(3), 532–560.
- Gregory, R. J. (2004). *Psychological testing: history, principles, and applications*, (4th ed.). Boston: Pearson Education Group.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). New York: Prentice-Hall.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191–205.
- Hearn, J. (1993). Emotive subjects: Organizational men, organizational masculinities and the (de)construction of “emotions”. In S. Fineman (Ed.), *Emotion in organizations* (pp. 142–166). London: Sage.
- Helmreich, R. L., & Merrit, A. C. (1998). *Culture at work in aviation and medicine: National, organizational and professional influences*. Aldershot, England: Ashgate.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185.
- Hunter, D. R. (2005). Measurement of hazardous attitudes among pilots. *The International Journal of Aviation Psychology*, *15*(1), 23–43.
- Johnson, D. (1995). Cultural differences: Flight deck reality and problems. In B. J. Hayward, & A. R. Lowe (Eds.), *Applied Aviation psychology: Achievement, change and challenge. Proceedings of the Third Australian Aviation Psychology Symposium* (pp. 191–194). Aldershot, England: Ashgate.
- Karp, M. R., & Turney, M. A. (2004). Making everyone part of the team: A model. In M. A. Turney (Ed.), *Tapping diverse talents in aviation: Culture, gender, and diversity* (pp. 221–229). Aldershot, England: Ashgate.
- Kaiser, H. F. (1961). A note on Guttman’s lower bound for the number of common factors. *British Journal of Statistical Psychology*, *14*(1), 1.
- Kelloway, E. K. (1998). *Using LISREL for structural equation modeling: a researcher’s guide*. Thousand Oaks, CA: Sage.
- King, R. E., Retzlaff, P. D., & McGlohn, S. E. (1997). Female United States Air Force pilot personality: The new right stuff. *Military Medicine*, *162*(10), 695–697.

- Krafft, P., Engelbrecht, A. S., & Theron, C. C. (2004). The influence of transformational and transactional leadership on dyadic trust relationships through perceptions of fairness. *South African Journal of Industrial Psychology*, 30(1), 10–18.
- Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworth.
- McCarthy, E., & McGinn A. (1995). Organisational culture, job satisfaction and work stress: the case of Irish women pilots. In N. McDonald, N. Johnston, & R. Fuller, (Eds.), *Applications of psychology to the aviation system. Proceedings of the 21<sup>st</sup> Conference of the European Association for Aviation Psychology (EAAP)*, 1, 81–91. Aldershot, England: Ashgate.
- Morgan, A. G., & Griego, O. V. (1998). *Easy use and interpretation of SPSS for Windows: Answering research questions with statistics*. London: Lawrence Erlbaum.
- Morgan, M. J. (2004). Women in a man's world: Gender differences in leadership at the Military Academy. *Journal of Applied Psychology*, 34(12), 2482–2502.
- Moolman, V. (1981). *Women aloft*. Alexandria, Va: Time-Life Books.
- Nagel, D. C., & Weiner, E. L. (1988) *Human Factors in Aviation*. San Diego: Academic Press Inc.
- Naudé, J. L. P., & 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(3), 21–28.
- Nicholson, P. (1996). *Gender, power and organisation*. London: Routledge.
- O'Conner, B. P. (2000). SPSS and SAS programs for determining the number of components: using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments & Computers*, 32, 396–402.
- Preferential Procurement Policy Framework Act, 5. (2000). *Republic of South Africa Government Gazette*. (Vol. 416, No. 20854).
- Promotion of Equality and the Prevention of Unfair Discrimination Act, 4. (2000). *Republic of South Africa Government Gazette*. (Vol. 416, No. 20876).
- Schepers, J. M. (1992). *Toetskonstruksie: Teorie en praktyk. (Test construction: Theory and practice)*. Johannesburg, South Africa: Rand Afrikaans University.
- Sitler, R. (2004). Gender differences in learning to fly. In M. A. Turney (Ed.), *Tapping diverse talents in aviation: Culture, gender, and diversity* (pp. 77–88). Aldershot, England: Ashgate.
- Sitler, R., Turney, M. A., & Wulle, W. (1996). Attitudes reflective of gender-based issues in the aviation and transportation workplace. In B. J. Hayward, & A. R Lowe (Eds.), *Applied aviation psychology: Achievement, change and challenge. Proceedings of the Third Australian Aviation Psychology Symposium* (pp. 332–342). Aldershot, England: Ashgate.
- Tabachnick, B. G., & Fidell, C. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn and Bacon.
- Theron, C. C. (2006). *Referee report of manuscript 06017: The construct validity of the revised job diagnostic survey*. (Available from the Scientific Editor of the South African Journal of Business Management, University of Stellenbosch, P.O. Box 610, Bellville 7535, South Africa).
- Turney, M. A. (1995). Women's learning and leadership styles: Impact on CRM. In N. McDonald, N. Johnston, & R. Fuller, (Eds.), *Applications of psychology to the aviation system. Proceedings of the 21<sup>st</sup> Conference of the European Association for Aviation Psychology (EAAP)*, 1, 262–268. Aldershot, England: Ashgate.
- Turney, M. A. (2004). *Tapping diverse talents in aviation: Culture, gender, and diversity*. Aldershot, England: Ashgate.
- Turney, M. A., & Bishop, J. C. (2004). Women's learning and leadership styles: Implications for air crews. In M. A. Turney (Ed.), *Tapping diverse talents in aviation: Culture, gender, and diversity* (pp. 61–75). Aldershot, England: Ashgate.
- Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.

- Vermeulen, L. P., & Wilson, J. (2002). *The measurement of perceptions with regard to gender-based issues in aviation. Unpublished research report.* University of Pretoria, Pretoria. 1–52.
- Vermeulen, L. P., Wilson, J., & Mitchell, J. I. (2004). *The measurement of perceptions regarding gender-related pilot behaviour: An exploratory study. Proceedings of the 26<sup>th</sup> Conference of the European Association for Aviation Psychology.* 205–210.
- Wilkening, H. E. (1973). *The psychology almanac.* Monterey, CA: Brooks/Cole.
- Yount, L. (1995). *Women aviators.* New York: Facts on File.