CHAPTER FIVE:

RESULTS

5.1 INTRODUCTION

The primary purpose of the study was to develop a valid and reliable instrument to measure airline pilots' perceptions of the training climate associated with advanced automated aircraft. The fundamental goal was to develop a questionnaire by operationalizing a hitherto unobserved hypothetical construct (perceptions of the advanced automated aircraft training climate) based on the three hypothesised levels of analysis (the person, the group and the organisation) that conceptualised the construct. In addition, the purpose of the research was to explore the relationships between the data and the characteristics of the data further in terms of the latent constructs that emerged.

The earlier part of the discussion provided a theoretical overview of the background in which the hypothetical measurement construct was devised and developed. Chapters 5 and 6 report on the results, interpret and discuss them.

Factor analysis was the statistical method used to provide scale descriptors, and it played a pivotal role in the development of an appropriate psychological measurement tool. Furthermore, the results of an analysis of the items forming the scale and the reliability of the constructs are reported on and discussed. To enhance the strategic statistical exploration of the phenomena present in the data set, initially, basic non-parametric comparative and associational analyses were completed. Subsequently, more in-depth and complex analyses, such as non-parametric MANOVAs and a stepwise logistic regression, were conducted to round off a thorough examination of the data set.

5.2 FACTOR ANALYSIS

Developing a scale for any psychological measurement instrument generally entails reducing the attribute space of a large number of variables into correlated factors or dimensions (Clason, & Dormody, 2001). Thus, it was assumed that the explanation for any correlations between the many variables in the dataset could be obtained from a small number of sub-constructs, as further substantiated by Corston and Colman (2003). In order to determine the number and nature of the latent factors responsible for most of the covariance in the current data, the process followed in this study consisted of the following more complex steps which is formulated within the algorithm for data reduction in SPSS (version 17). According to Brown *et al.* (2012):

- Step 1 is to calculate the covariance matrix from the raw data;
- Step 2 then computes a correlation matrix by transforming the aforementioned covariance matrix (in this case, referred to as matrix, A);
- Step 3 is then to extract eigenvectors (in this case, x_i) from the aforementioned correlation matrix by the method of successive squaring;
- Step 4 obtains eigenvalues (λ_i) from each eigenvector (x_i) computed, based on the characteristic equation, $A.x_i = x_i.\ \lambda_i$ (the equation shows that there exists a unique eigenvalue in each linear transformation);
- Step 5 finally generates the factor loadings (f_i), by a normalisation of the eigenvectors (x_i) to λ_i, that is, $f_i = x_i\ (\sqrt{\ \lambda_i})$;
- Step 6 finalises the results of the above computations, and are displayed in terms of the factor product matrices and factor residuals.
- Step 7 is required to examine the diagonal of the final matrix of residuals to determine the amount of variance that was not accounted for in each variable, and also to examine the off-diagonal elements so as to determine how well the level of correlations amongst each variable pair was subsequently reproduced by the factor solution; and
- Step 8 creates the final matrix of factor loadings (F) from the factors (f_i) extracted. The factor loadings were used as the basis for imputing labels to the explanatory dimensions of the construct under investigation.

The technique therefore uses linear combinations of the empirically obtained variables to explain the sets of observations in the dataset.

5.2.1 Sample size rationale used for the factor analysis

The initial challenge in planning a factor analysis is to determine the size of the sample frame that is likely to result in the most stable solution. Unfortunately, the literature contains a multitude of divergent opinions on the question of the ideal sample size required for factor analysis, which made the selection of an appropriate technique difficult (Gorsuch, 1983; Hayton *et al*., 2004; MacCallum *et al*., 1999). Nonetheless, the main problem to be addressed in this study was in fact, understanding the impact that sampling error, rather than the sample size *per se*, may have had on the final factor analytic solution. Although sampling error can be directly related to sample size, sampling error can be mitigated even when the sample is relatively small, provided that the sample's elements are of a high quality (Haworth, 1996). A high quality final factor solution was deemed critical for this study, because the accuracy and validity of subsequent analyses required a strong foundation. Thus it was important to understand the impact that the final sample frame would have on the factor structure.

Typically, sampling error influences the estimation of factor loadings because the sample covariance of the measured variables is expressed as a function of the population's common and unique loadings (MacCallum *et al*., 1999). This then implies that prescriptive sample sizes, as a ratio of the population size, or in terms of the number of items, might be misleading. The methodology chapter discussed the various rules of thumb proposed by some authors to facilitate the determination of the ideal sample size (see Section 4.10.1). For instance, according to Cooper and Schindler (2003), a sample five times the number of variables is sufficient to produce relatively stable factor solutions, whereas Comrey and Lee (1992:67) suggest that between 200 and 300 participants in a survey using factor analysis should be regarded as only a "mediocre" sample size. Because of these divergent expert opinions, the various rules of thumb have all had their fair share of critics (MacCallum *et al*., 1999). On the basis of the suggested sample sizes discussed above, the final cohort for this study consisted of 229 participants, which is more than the 210 participants that would be five times the number of item variables – a rule of thumb in many similar studies, as recommended originally by Gorsuch (1983) and later by Cooper and Schindler (2003).

MacCallum *et al*. (1999) argue that if the communalities between the variables are relatively high, a small sample size will produce high quality sample factor solutions (in other words, effective replicability and recovery of population factors). Communalities refer to the proportion of each variable's variance that can be explained by the factors (Ledyard, 1966); hence, high communalities in the current study indicated that the final sample, in terms of the validity of the questionnaire items, was highly stable. Using a similar method, Corston and Colman (2003) found that even analysing only a small portion of Thurstone's (1947) data set did not significantly affect their own results.

The average initial communalities of the 42 items which operationalised the main research construct were computed at 0.630. MacCallum *et al*. (1999) consider communality strengths of above 0.60 to be very good. In summary, it appears from various findings reported in the literature that when communalities are relatively good, and the content of the hypothetical framework has been validated (as is the case in the present study), sampling error is reduced (Corston & Colman, 2003). Therefore, using between 200 and 300 returns can produce high quality sample factor analytic solutions. It is also reasonable to assume that very small differences would have occurred if the study had used a sample of 500 versus a sample of 200. The first step in the research process was to validate the content of the construct by using a panel of experts. This choice may be the reason for strengthened communalities amongst the variables. The item communalities (h^2) are reported in Tables 32 and 33.

5.2.2 Factor analytic computation

Using SPSS Version 17.0, in the various rounds of exploratory factor analyses, the 42 items in respect of the content of the hypothetical construct were inter-correlated and analytically rotated to an oblique simple structure by means of the promax method (factors were permitted to optimally correlate with one another). The rotations were raised to a Kappa power of 4, which Hendrickson and White (1964) consider appropriate for most analyses in the social sciences. Kappa is a statistical parameter, which was used to control the calculation of the promax rotation (Morgan, *et al*., 2007).

Gorsuch (1983) found that when values of Kappa are high, correlations among factors tend to be high, leading to a simpler structure of the loadings. It was concluded that raising the loadings to a power of 4 produced an optimised solution for the dataset, which resulted in a simpler structure with the lowest correlation amongst the factors. An inter-correlation matrix consisting of 1 764 variables was subsequently produced. However, due to its size, this matrix is not reported here.

Prior to the aforementioned factor analytic choice, two diagnostic tests confirmed that the data passed the minimum criteria required to conduct an initial exploratory factor analysis. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (which assesses whether the partial correlations among variables are small) and Bartlett's test of sphericity (which evaluates the null hypothesis that the variables in the population correlation matrix are uncorrelated) both produced satisfactory results for the current data set.

The calculated KMO value of 0.927 was greater than 0.70, which is the normal cut-off recommended for a factor analysis to proceed (Gorsuch, 1983). Similarly, Bartlett's test of sphericity [χ^2 (861) = 6896.895, $p < 0.001$] was significant and therefore confirmed that the properties of the inter-correlation matrix of the 42 item scores were suitable for factor analysis.

Subsequently, using Kaiser's (1961) criterion, principal axis factoring (PAF) postulated seven factors, where the initial eigenvalues (greater than unity) accounted for 66.248% of the variance in the factor space (see Table 31). The extraction was conducted on a correlation matrix, after conversion from a covariance matrix of the raw data. Therefore, the variables were standardised and the total variance (100%) would be found from the remaining items. The loadings and cross-loadings of the items associated with the seven eigenvalues greater than one were then examined. All items that had loadings of less than 0.40 or which were found to have a high cross loading on more than one factor were deleted. A cut-off point of 0.40 is the criterion generally used in the social sciences (Welman & Kruger, 1999). The loadings used in this round of analysis are not reported, due to space constraints. Items whose properties appeared extremely similar were also discarded, in line with the recommendations of Tabachnick and Fidell (2007). In total, seven items were deleted (see Table 32).

Finally, 35 items in a clean matrix were subjected to a second round of exploratory factor analysis with promax rotation and Kaiser Normalization (which is generally the default during rotation). This method decreased the standard errors of the loadings for variables with small communalities.

Table 31: Variance explained by eigenvalues greater than one (42 items)

Initial eigenvalues				Extraction sums of squared loadings	
n	Total	% of variance	Cumulative %	Total	% of variance
1	16.431	39.122	39.122	16.431	39.122
2	3.872	9.219	48.341	3.872	9.219
3	2.174	5.177	53.518	2.174	5.177
4	1.888	4.495	58.013	1.888	4.495
5	1.226	2.919	60.932	1.226	2.919
6	1.174	2.795	63.726	1.174	2.795
7	1.059	2.522	66.248	1.059	2.522

In general, most statistical software programs, and more specifically, the SPSS factor analysis program, defaults to Kaiser's (1961) criterion (eigenvalues greater than unity) for factor extraction. This method and other factor retention methods were examined further. Substantive evidence was found in the literature that demonstrated that surprisingly, Kaiser's criterion was correct only 22% of the time (Hayton *et al.*, 2004), which makes it problematic as a factor retention method for scale development at this level.

Hayton *et al.*'s (2004) data also show that Kaiser's criterion will only become accurate when the sample size approaches infinity, and therefore it was expected that any sampling error in the current study would generally produce more factors in the sample space. Deleting some items and subjecting the remaining variables to further rounds of factor analysis assisted in reducing the problems associated with factor over-estimation.

Table 32: Statements deleted in the first round of exploratory factor analysis

Item	Statement
Q35	The standard operating procedures (SOPs) for learning to fly this aircraft are adequate.
Q37	The simulators my company trains its pilots in are in good condition.
Q43	I learn better when I work as a member of the crew.
Q46	I tend to communicate well with my simulator partner.
Q47	The instructor is committed.
Q48	Instructors are very similar in how they teach pilots to fly this aircraft.
Q59	I reflect on my learning experience after a simulator session.

5.2.3 Results of the factor retention method

Cattell's scree plot and Kaiser's (1961) criterion were used to determine tentatively the number of factors to retain, as these appeared to be the most common methods reported in the literature. In Figure 24, the graphs plot the eigenvalues against each component number. From the third or fifth factor onwards, it appears that the plot has flattened, which suggests that each successive factor accounts for less and less of the total variance explained. According to Cattell (1966), factors before the one starting the bend or elbow where the plot levels off should be retained. Inspection of the curve indicating the number of factors to retain showed that the graph was very difficult to decipher, due to the ambiguity in the shape of the curve. Unfortunately, the scree test has been shown to suffer from a degree of subjectivity. For instance, Hayton *et al.* (2004) point out that when there are no clear breaks, or, as in this case, when there are two or more apparent breaks, confusion can arise.

In the current study, this confusion made it difficult to determine unambiguously the optimum number of factors to keep which can accurately describe the latent structure of the measured construct. A further problem was that Kaiser's method appeared to overestimate the number of factors to retain, because six eigenvalues exceeded unity for 35 items (see Table 33).

Figure 24: Thirty-five item scree plot

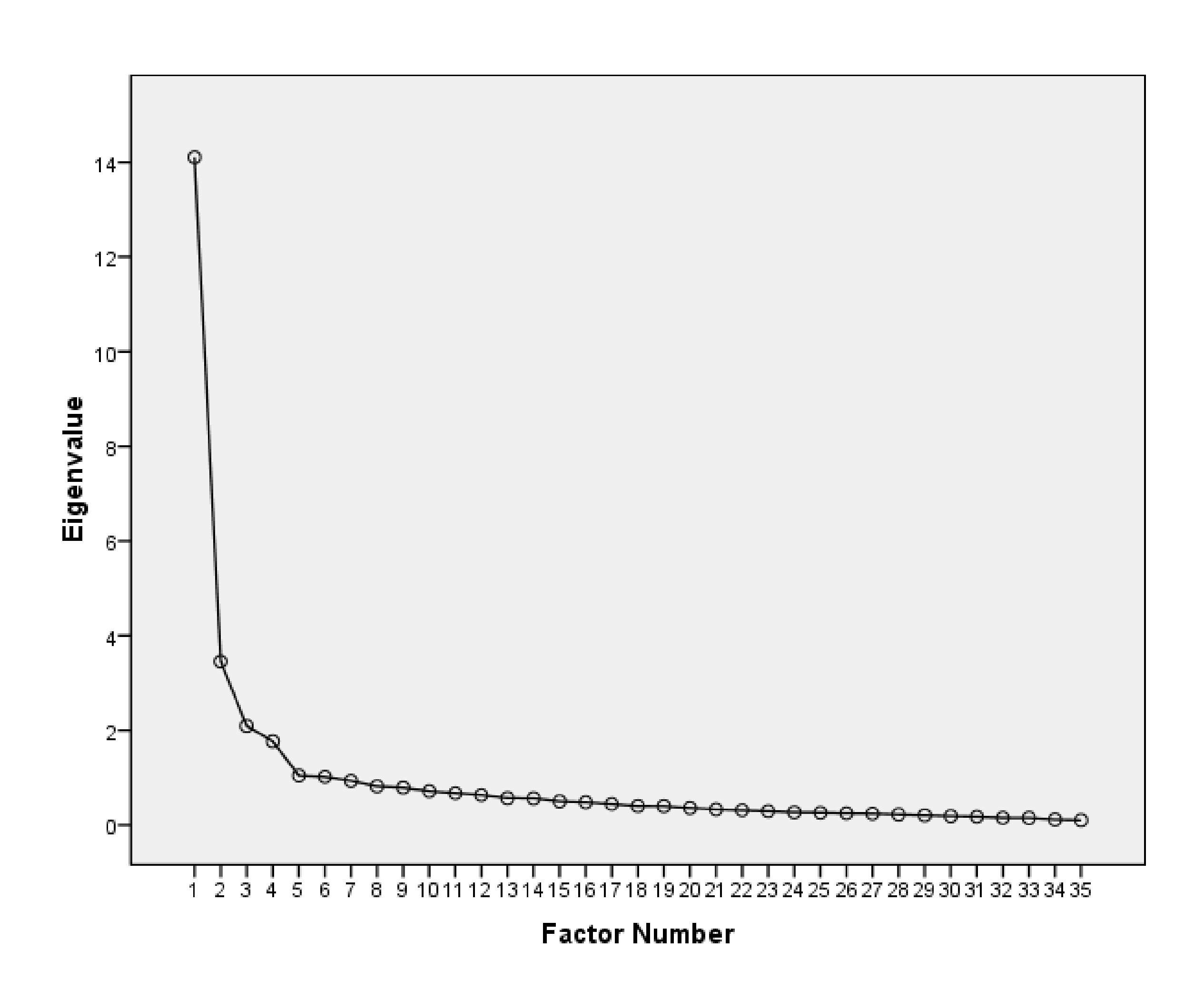


Table 33: Variance explained by eigenvalues greater than one (35 items)

Initial eigenvalues				Extraction sums of squared loadings		
VAR	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	14.105	40.301	40.301	13.713	39.180	39.180
2	3.454	9.869	50.170	3.148	8.994	48.175
3	2.092	5.978	56.148	1.727	4.934	53.109
4	1.767	5.048	61.196	1.454	4.156	57.264
5	1.044	2.984	64.180	0.639	1.825	59.089
6	1.018	2.909	67.089	0.611	1.745	60.834

It was observed that the first attempt at exploratory factor analysis overestimated the factors in the real test space for this data set using traditional methods. One source of this overestimation may be possible sampling error or differentially skew items (Hendrickson & White, 1964). According to Hayton *et al*. (2004), some factors may have eigenvalues greater than one only because they originate from a finite sample, as opposed to an infinite population. In an infinite population, initial eigenvalues tend to be greater than one, whilst later eigenvalues tend to be smaller than one. Possibly, the most important decision a researcher developing a scale and using factor analysis has to make, is deciding on the number of factors to retain (Glorfeld, 1995; Watkins, 2006).

Accurate factor retention was considered to be a core requirement for the overall success of this study. Both the scree plot and Kaiser's method were inconclusive for the current dataset. To determine the maximum number of significant factors, which would reasonably explain the variability of the main research construct, a modified version of Horn's (1965) parallel analysis (PA) based on a Monte Carlo simulation, was conducted, as recommended by Velicer (1976).

Hayton *et al*. (2004:194) explain that the "rationale underlying PA is that nontrivial components from real data with a valid underlying factor structure should have larger eigenvalues than parallel components derived from random data having the same sample size and number of variables". A comparative method for exploration in factor analysis is considered one of the most accurate factor retention methods available (Rencher, 2002). Unfortunately, most statistical software packages do not offer Horn's (1965) procedure to determine the maximum number of factors to retain (Watkins, 2006). For the current study, the computational algorithm created by Brian O'Connor (2000) was used; and by manipulating this complex syntax code within the SPSS editor the required analysis was achieved.

Figure 25: O'Connor plot of the actual, mean and permutated eigenvalues

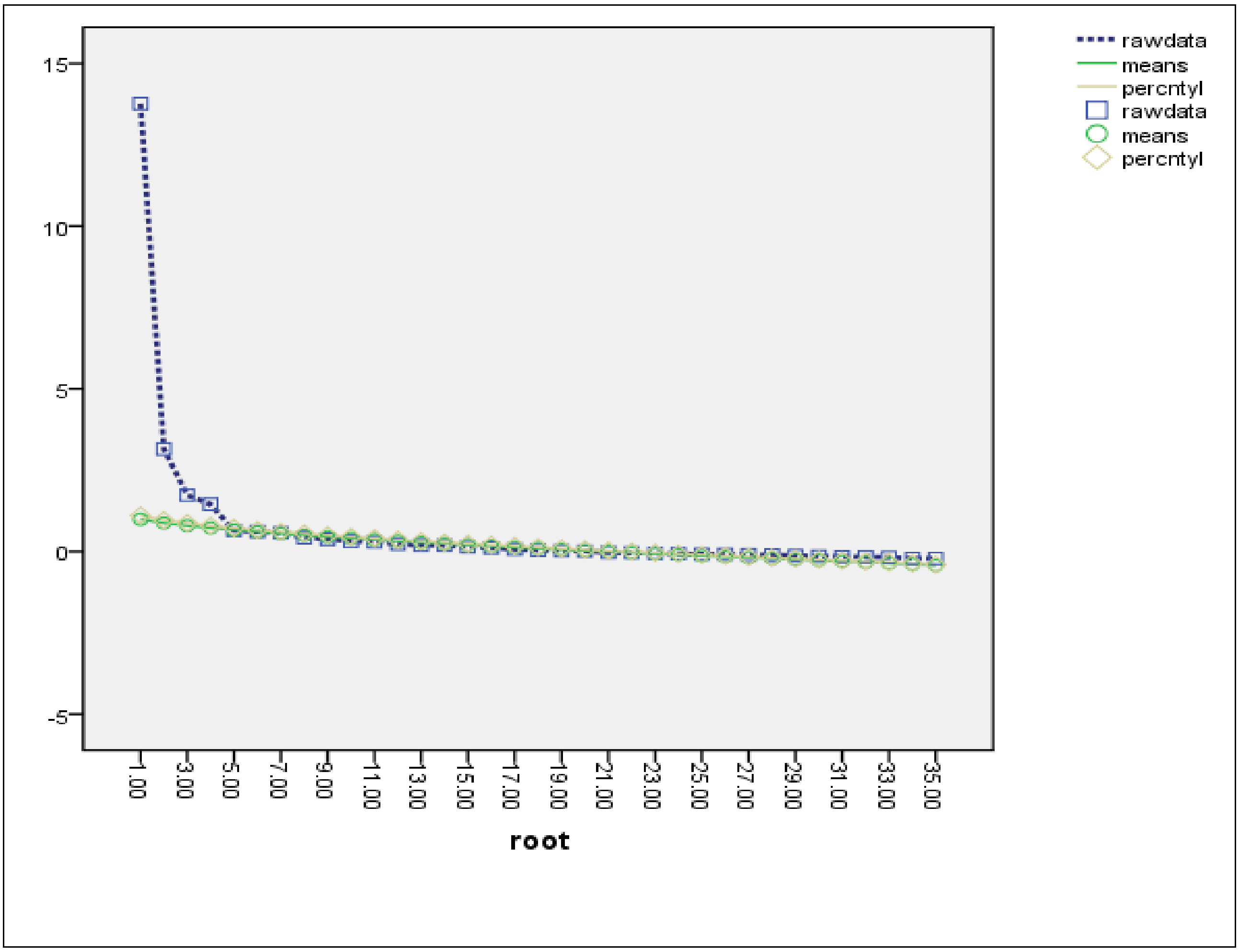


Table 34 and Figure 25 show the actual eigenvalues drawn from the real data space of a 35-item AATC-Q for 229 participants, together with the mean and 95th percentile eigenvalues extracted from 1 000 permutations of the original data set generated in a Monte Carlo simulation. In this simulation, in cases where the raw data are not normally distributed or where they do not meet the assumption of multivariate normality, permutations are deemed more accurate and relevant than randomised data (O'Connor, 2000)

Table 34: Actual and permutated eigenvalues on 35 items based on O'Connor's (2000) algorithm

Root	Actual eigenvalue	Mean eigenvalue	95th percentile eigenvalue
1.	13.763320	0.991263	1.110948
2.	3.152418	0.877747	0.964691
3.	1.734102	0.794188	0.876276
4.	1.451794	0.723979	0.792523
5.	0.649697	0.660416	0.722079
6.	0.612930	0.601232	0.659002
7.	0.574928	0.548531	0.604958
8.	0.422722	0.498339	0.553514
9.	0.381465	0.450765	0.499703
10.	0.320640	0.404774	0.451020
11.	0.297724	0.360918	0.404150
12.	0.233578	0.318256	0.361014
13.	0.199624	0.278398	0.317899
14	0.196496	0.239460	0.277779
15.	0.140422	0.202216	0.237528
16.	0.116488	0.166291	0.204303
17.	0.074143	0.129561	0.166742
18.	0.042527	0.094801	0.129188
19.	0.024271	0.060795	0.095014
20.	0.014329	0.027945	0.060390
21.	-0.023788	-0.004203	0.027235
22.	-0.035305	-0.035648	-0.003258
23.	-0.053125	-0.067442	-0.036161
24.	-0.065674	-0.097213	-0.068280
25.	-0.078143	-0.127373	-0.099662
26.	-0.084013	-0.157061	-0.130434
27.	-0.096836	-0.185802	-0.160826
28.	-0.105393	-0.213957	-0.188567
29.	-0.124817	-0.242558	-0.218763
30.	-0.144008	-0.270705	-0.247826
31.	-0.159285	-0.299199	-0.276946
32.	-0.164820	-0.328068	-0.304783
33.	-0.175846	-0.357945	-0.333908
34.	-0.209793	-0.390079	-0.363294
35.	-0.214610	-0.428936	-0.399151

The permutation algorithm based on Castellan's (1992) recommendation was used to compute the data set. The algorithm ensures that for *N* elements there are *N!* permutations, and each one is equally likely (Table 34). In accordance with the previous extraction methods, a principal axis or common factor analysis was requested for the parallel analysis. A difference can be clearly observed between the eigenvalues produced by the SPSS programme and those produced using the O'Connor (2000) algorithm. The SPSS statistical software program computes a correlation matrix, which differs slightly from that being used in the syntax of the parallel analysis. For instance, Kendall correlation coefficients are generally used for ordinal variables when assumptions of normality are violated, instead of Spearman correlation coefficients, as in the case of normalised data.

Furthermore, when the matrix is not Gramian, in that squared multiple correlations are inserted into the diagonal of the matrix as communalities; a principal factor extraction tends to produce a number of negative eigenvalues (Comrey & Lee, 1992). According to Gentle (2007:288), a Gramian matrix is "a (real) matrix A such that for some (real) matrix B, $A = B^TB$". Because B^TB is symmetric, any non-negative definite matrix is Gramian. In addition, "each element of a Gramian matrix is the dot products of the columns of the constituent matrix" (Gentle, 2007:228). It was determined that in the present SPSS algorithm, the issue is automatically resolved by adjusting the off-diagonal elements of the correlation matrix before factorisation, which is not the method adopted in O'Connor's (2009) program. However, making an adjustment in the off-diagonal elements will, in turn, result in an underestimation of the values of communalities. The parallel analysis syntax uses a more accurate shuffling transformation technique, as recommended by Castellan (1992), for the generation of comparable permutated matrices. According to Castellan (1992:72), the algorithm, "initially generates a random integer between 1 and N, swaps the first element with the generated element, then generates a second random integer between 2 and N, swaps the second element with the second generated element, and so forth". Coding in this manner in order to generate permutated data sets is regarded as efficient, in that it requires only N-1 random integers to permutate the array of N elements; furthermore, the elements are sampled without replacement (Castellan, 1992). In the current study, eigenvalues after the 21st variable in the O'Connor output were negative (dropped below zero). This result is due to the difference between the computation formulae adopted by the SPSS programmers for factor analysis and that used in the O'Connor syntax. Both formulae are acceptable for the purposes of exploratory factor analysis (Rencher, 2002). In addition, an iterative principal axis factoring technique (as in a Monte Carlo type simulation) is more likely to produce negative eigenvalues than a standard principal components method (Watkins, 2006). Also, anomalies can occur when eigenvalues and eigenvectors are computed after first estimating the initial values of the covariance matrix, as in the case of a principal axis factoring method (Rencher, 2002). The results set out in Figure 25 and Table 34 suggest that there are, at most, four eigenvalues greater than those generated by the parallel analysis for both the mean and the 95th percentile criterion. The raw data eigenvalues greater than chance ranged from 13.7 to 1.4. The 95th percentile eigenvalues for this range fell between 1.1 and 0.79. Therefore, four factors were deemed statistically significant ($p < 0.05$), suggesting that a maximum of four components is possible in the factor space of the current data set (Velicer, 1976), and not seven, as was originally postulated using the Kaiser (1961) criterion. O'Connor (2000) suggests that the eigenvalues from a parallel analysis should be used to determine the maximum real data eigenvalues that can occur beyond chance, and additional procedures should then be used to trim trivial factors. In other words, the current parallel analysis indicated the maximum possible number of factors, but it did not impose the final number of factors to retain, as Hayton *et al.* (2004) also found.

5.2.4 Finalised factor analytic solution

The aim in the final factor solution was to obtain a simple structure, after an exhaustive retention process (see also Appendix C, for the final 3-scale solution). Only by determining the probable number of real factors that could exist in the variable space from a Monte Carlo simulation could a maximum four-factor solution be requested. Unfortunately, a four-factor structure resulted in generalised Heywood cases, which, according to Harman and Yoichiro (1966:563), are mathematical anomalies that occur "when a correlation matrix yields a suitable factor solution with several common factors for which one of the communalities exceeds unity". Such cases indicate that the solution may be unstable, possibly because are too many or too few factors. Therefore, this result suggested that the solution was far from simple at this point. With this in mind, a three-factor solution was then requested, because "if the researcher is interested in using only demonstrably reliable factors, the fewest possible factors are retained" (Tabachnick & Fidell, 2007:646).

Table 35: The factor loadings and communalities (h^2) for the principal factors extraction and promax rotation for the final 33-item cohort

Item	Statement	Factor			
		1	2	3	h^2
Q29	Training on this aircraft is well organised.	0.809	-0.041	-0.021	0.708
Q27	Training on this aircraft is professional.	0.808	0.082	0.015	0.767
Q23	My company's training produces world class pilots.	0.785	0.121	-0.078	0.770
Q24	Training at my airline is in line with company goals.	0.785	0.115	-0.057	0.725
Q38	The airline is very supportive of its pilots' learning requirements for this aircraft.	0.763	-0.059	0.108	0.703
Q34	There is sufficient training guidance from the company.	0.762	-0.143	0.109	0.667
Q28	Management follows the rules and regulations appropriately.	0.755	-0.053	-0.102	0.646
Q39	My company's culture supports training for new technology aircraft.	0.731	-0.008	-0.045	0.594
Q30	I understand what the company expects of me when training.	0.700	0.198	0.000	0.706
Q26	My company has talented people in training.	0.695	0.012	0.060	0.633
Q33	If I had to experience a problem in training, it's easy for me to appeal.	0.686	-0.250	0.291	0.681
Q25	I know what my company's training goals are.	0.678	-0.036	0.146	0.697
Q31	Training at my airline produces safe pilots.	0.644	0.217	-0.043	0.670
Q40	There is sufficient feedback about my training on this aircraft.	0.606	0.029	0.090	0.546
Q42	My company uses only current training material.	0.589	-0.106	0.035	0.428
Q41	Training is in line with civil aviation regulations.	0.586	0.281	-0.078	0.595
Q32	The airline gives its pilots an appropriate amount of preparation work for training.	0.531	-0.028	0.189	0.525
Q45	My instructor is willing to listen.	0.528	-0.028	0.277	0.639
Q50	Pilots are in direct control of the training outcome.	0.522	-0.111	0.397	0.696
Q36	I'm given sufficient time to prepare for training on this aircraft.	0.499	0.046	0.192	0.558
Q61	It's a good idea to know more than what is required.	-0.341	0.759	0.340	0.689
Q52	I try never to be late for a training session.	0.208	0.746	-0.257	0.828
Q53	I co-operate when training in a simulator.	0.209	0.745	-0.238	0.810
Q62	I aim to gain a deeper understanding of this aircraft.	-0.312	0.734	0.328	0.773
Q51	Preparation improves performance.	0.254	0.655	-0.065	0.733
Q60	I read to understand so as to gain a deeper understanding of this aircraft's systems.	-0.218	0.626	0.258	0.672
Q55	I have a positive relationship with my colleagues.	0.224	0.586	-0.011	0.619
Q44	I operate well as a crew member in the simulator.	0.182	0.521	0.148	0.595
Q58	I enjoy studying the technical aspects of the aircraft.	-0.089	0.493	0.316	0.594
Q63	I'm comfortable undergoing training for this aircraft.	0.182	0.130	0.594	0.601
Q57	I'm in control of the outcome of a training session.	0.326	-0.017	0.529	0.656
Q64	I can control my anxiety so as to perform well in training.	-0.073	0.103	0.523	0.497
Q49	The instructors on this aircraft don't overload us with information.	0.398	-0.076	0.461	0.591
Eigenvalues = 14.105; 3.454; 2.092; % variance = 40.301; 4.869; 5.978; Cumulative % = 56.148					

A quantitatively superior solution was found by retaining only three factors (with no further Heywood anomalies). However, in the final 35-item cohort, Item 54 ("After training I feel a sense of mastery") and Item 56 ("Pilots who come prepared have no problems training for this aircraft") had weak loadings (< 0.40) in the final rotated matrix. They were therefore also discarded from the final factor solution.

The factor loadings and the communalities (h^2) of the 33 items, as well as the total variance explained by the different factors, are depicted in Table 35. The factor loadings are set out in descending order for easier interpretation. Inspection of these results suggests that the three factors were well determined and that the three factors explained approximately 56.148% of the total variance in the data.

According to the results, the factor scores of the factor solution ranged from excellent to fair, with factor scores varying from 0.809 to 0.461:

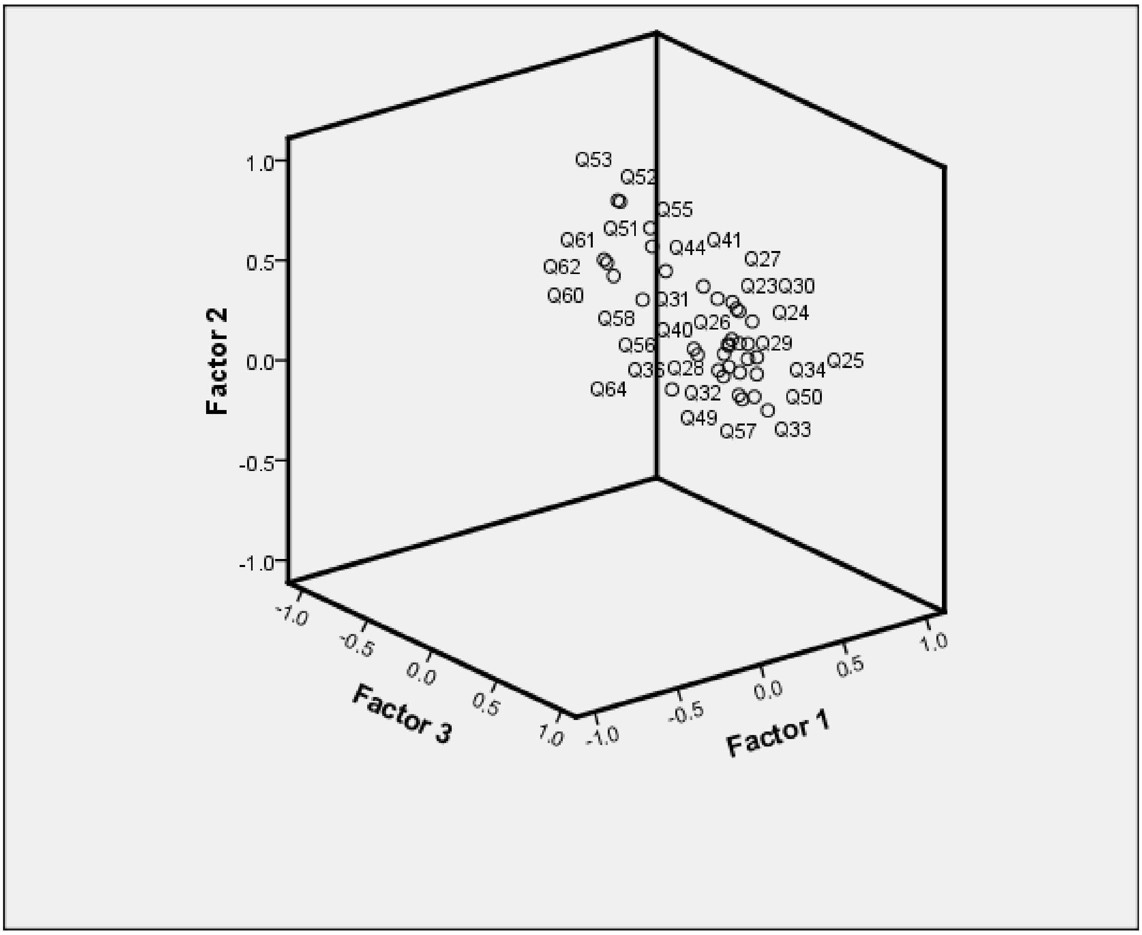
- 0.809 to 0.499 for Factor 1;
- 0.759 to 0.493 for Factor 2; and
- 0.594 to 0.461 for Factor 3.

Comrey and Lee (1992:203) suggest that loadings in excess of 0.70 be considered "excellent", 0.63 or more "very good", 0.55 or more "good", 0.45 or more "fair" and 0.32 or more "poor". Loadings lower than 0.32 should be disregarded. High factor loadings are a clear indication that a variable is a "pure" measure of the factor (Tabachnick & Fidell, 2007:649).

In the current study, the average communality loadings (h^2) of 0.701 to 0.586 indicated that the items define the factors relatively well, representing high levels of consistency (Leech *et al.*, 2005). In this kind of context, Tabachnick and Fidell (2007:660) point out that "[c]ommunalities indicate the percentage of variance in a variable that overlaps variance in the factors". Comrey and Lee (1992:12) define the communality for a variable in factor analysis as "the sum of squares of the factor loadings over all the factors". The final factor loadings and communalities that emerged from the analysis were dependent on the original estimated communality values used in the early correlation matrix (Comrey & Lee, 1992). Therefore, high communalities (see Table 35) indicate homogeneity and the factorial purity of the final scale (Appendix C).

The grouping of variables in the rotated space also indicates that the factors were sufficiently described by their items, because a “clustering of variable points reveals how clearly defined a factor is” (Tabachnick & Fidell, 2007:647). The linear combinations of the variables are presented in the rotated factor pattern matrix (see Figure 26). However, only one clear factor cluster (Factor 1) can be seen within the rotated space. The remaining two factors may not be as clear, because values of the loadings are based on the common variance, which is generally lower than the total variance.

Figure 26: Factor plot in the rotated space



In addition, the internal consistency of the factor solution was verified by calculating the Squared Multiple Correlations (SMCs). This index indicates that the variable data points were fixed in the factorial space to a high degree of certainty (Bentler & Raykov, 2000), as set out in Table 36.

The importance of each factor was assessed by the percentage of variance that it represented. The Sum of Squared Loadings (SSLs) from the rotated factors is the redistributed variance during rotation. Each of the factors accounted for between 6% and 40% of the covariance. As expected, the first factor (F1) accounted for the bulk of the covariance.

The SMCs values and the intercorrelation between the three factors are depicted in Table 36. The results suggest that the three factors intercorrelated significantly with one another (r = 0.242 to 0.446). The strength of the correlations indicates that the three factors are closely related in measuring the constructs associated with the advanced automated aircraft training climate. Although the relatively high intercorrelations may also suggest overlapping variability, the SMCs nonetheless indicate that all the factors were sufficiently defined by the relevant items.

Table 36: Item regression model and factor correlations

SMC Model (Factor)	R	R Square	Adjusted R Square	Standard error of the estimate
1	0.995	0.990	0.989	0.10083248
2	0.991	0.983	0.982	0.12711818
3	0.900	0.809	0.806	0.40278467
	Factor correlations			
Factor	F1		F2	F3
1	1.000		0.446	0.431
2	0.446		1.000	0.242
3	0.431		0.242	1.000

The SMCs or R^2 values in Table 36 were calculated by means of the regression method described by Tabachnick and Fidell (2007). Items were taken as predictors of the factor constructs to produce the SMC models shown. The high R^2 values, ranging from 0.809 to 0.990, between the item scores and the factor scores testify to a good fit between the item scores and the latent factors. The R^2 values also represent the proportion of shared variance between each of the items and its related factor. The results indicate that the items could account for between 80% and 98% of the variation in the three factors. The model provides the necessary evidence to demonstrate that the items are significant explanatory variables of the latent constructs of pilots' perceptions of the advanced aircraft training climate.

The results of the analysis discussed above suggest that the items of the questionnaire exceeded the required adequacy in measuring the factors that they were related to. The item clusters were used as a guide for the factor labelling process.

5.3 SCALE LABELLING AND FACTOR DESCRIPTION

In order to communicate the nature of the underlying constructs, factors must be appropriately labelled (Ledyard, 1966). From an inspection of the content of the sub-scales, it is possible to gain a deeper understanding of the nature of the underlying factor constructs (Comrey & Lee, 1992). In addition, factors are represented by different sets of variables and therefore each factor should be well differentiated from the others.

"The variables that have high loadings on the rotated factor are studied carefully to determine what they share in common" (Comrey & Lee, 1992:11). The description and naming of the latent factors that accounted for most of the variability in the main research construct was based on the three to five statements with the highest explanatory connotations within each grouping such that:

- *Factor 1*

  This factor essentially relates to the organisational aspects of the training climate. Items from both the macro domain (the airline) and the intermediate (instructor-trainee) domain loaded substantively onto Factor 1. Essentially, the factor expresses a component of the theoretical construct in terms of the efficiency, effectiveness and professionalism of both the company and its flight instructors. The elements dominating this factor relate to organisational co-ordination, trainee support, rules, regulations, sufficient learner feedback and guidance. This factor is referred to as *Organisational Professionalism*.

- *Factor 2*

  This factor contains elements representing the micro level of analysis (the person). The factor predominantly reflects the individual trainee's ability and eagerness to learn and understand complex concepts relating to advanced aircraft. Learning the aspects of a complex technology is regarded as a structured and iterative quantitative increase in knowledge. The fundamental aspects of this factor relate to an individual's learning approach, preparedness, and willingness to participate and co-operate in training to gain a knowledgeable and workable understanding of the advanced aircraft. This factor has been labelled *Intrinsic Motivation*.

- *Factor 3*

  The third factor represents an individual trainee's own perceived level of control in terms of stress levels and decision making, regarding the training required to operate an advanced automated aircraft. Four items found in the micro domain (the person) loaded meaningfully onto this factor. The principal elements of Factor 3 relate to the levels of perceived comfort experienced by trainees during training, their belief in their ability to control the outcome of a training session, their capacity to control their levels of stress (eustress or anxiety) in order to perform well, and ultimately their grasp of the amount of information required to cope with their training (intelligent decision making). This factor is referred to as *Individual Control of Training Outcomes.*

Comrey and Lee (1992:11) suggest that a researcher go beyond simply doing a factor analysis and labelling the factors, because a factor analysis should be considered "a way of generating hypotheses about nature". The derived factor constructs and rotated factor matrix can therefore be considered as only one interpretation of the latent behaviour of the present population (Appendix E).

5.4 RELIABILITY ANALYSIS

"The reliability of a measuring instrument is defined as its ability to consistently measure the phenomenon it is designed to measure" (Ho, 2006:239). However, it is expected that data variables will not be perfectly reliable because they are not perfectly correlated (Comrey & Lee, 1992). In order to determine the level of internal consistency (the interrelatedness of a set of items and the extent to which the items in a scale measure the same construct), Cronbach's coefficient alpha was computed for each of the sub-scales. In addition, the mean inter-item correlations were computed to assess the internal homogeneity and unidimensionality of the item clusters in each factor scale, as recommended by Pett *et al*. (2003). Furthermore, the mean, standard deviation, skewness, and kurtosis, for each of the items were calculated to assess the distribution of the reposes of the present sample on the 33 item AATC-Q

As described in Tables 37 to 39, the Cronbach's coefficient alphas (α) of all three factors were relatively high and exceeded the recommended threshold of 0.70, which is the accepted standard requirement for levels of internal consistency in social sciences studies of this nature and for the number of items in a scale (Comrey & Lee, 1992; Cortina, 1993; Cronbach, 1951; Morgan & Griego, 1998; Nunnally & Bernstein, 1994). Furthermore, the findings indicate that each item contributed significantly to the high reliability coefficient within each factor. It was also apparent that none of the items adversely reduced the value of alpha if the item was removed from the cluster.

The high Cronbach coefficient alphas were also interpreted as an indication that there is very little variance specific to individual items. In other words, the sets of items were found to conform to Cronbach's (1951) original definition of *equivalence*, where the value of Cronbach's alpha takes into account all the information contained in the items, namely the number of items, their variance and covariance.

Table 37: Reliability and item statistics for Factor 1: *Organisational Professionalism* (n =229)

Item statistics	Sk	Ku	Mean	Standard deviation	Corrected item total correlation	Cronbach's alpha if the item is deleted
Q23	-1.42	2.17	5.97	1.162	0.776	0.948
Q24	-1.22	2.26	6.04	1.023	0.781	0.949
Q25	-1.18	1.23	5.67	1.355	0.722	0.949
Q26	-1.60	3.65	6.03	1.151	0.700	0.949
Q27	-1.81	4.19	6.15	1.066	0.828	0.948
Q28	-1.29	1.23	5.46	1.497	0.668	0.950
Q29	-1.42	1.99	5.65	1.370	0.767	0.948
Q30	-1.43	2.33	6.08	1.079	0.759	0.949
Q31	-1.41	2.17	6.00	1.116	0.693	0.949
Q32	-1.21	1.14	5.56	1.449	0.600	0.951
Q33	-1.09	0.69	5.39	1.609	0.701	0.950
Q34	-1.36	1.78	5.65	1.367	0.736	0.949
Q36	-1.61	2.55	5.91	1.383	0.595	0.951
Q38	-1.33	1.53	5.66	1.417	0.783	0.948
Q39	-1.38	1.48	5.75	1.385	0.689	0.949
Q40	-1.25	1.31	5.65	1.288	0.659	0.950
Q41	-2.24	6.07	6.38	1.018	0.649	0.950
Q42	-1.51	2.71	5.90	1.201	0.553	0.951
Q45	-1.53	2.78	5.77	1.275	0.623	0.950
Q50	-0.94	0.57	5.21	1.401	0.637	0.950
Reliability statistics				N of items 20	Mean inter-item correlation 0.663	Cronbach's alpha 0.952
Scale statistics				Mean 115.86	Variance 348.249	Standard deviation 18.661

Table 38: Reliability and item statistics for Factor 2: *Intrinsic Motivation* (n = 229)

Item statistics	Sk	Ku	Mean	Standard deviation	Corrected item total correlation	Cronbach's alpha if the item is deleted
Q44	-1.01	0.92	6.16	0.844	0.599	0.865
Q51	-1.89	3.27	6.52	0.820	0.645	0.861
Q52	-3.29	11.20	6.76	0.694	0.624	0.865
Q53	-2.69	6.98	6.74	0.643	0.635	0.865
Q55	-1.17	1.13	6.41	0.724	0.603	0.866
Q58	-1.28	2.39	5.72	1.131	0.560	0.873
Q60	-1.25	2.09	5.92	1.099	0.632	0.864
Q61	-1.73	3.83	6.31	0.944	0.678	0.858
Q62	-1.27	2.15	6.17	0.926	0.711	0.855
Reliability statistics				N of items 9	Mean inter-item correlation 0.632	Cronbach's alpha 0.877
Scale statistics				Mean 56.71	Variance 31.910	Standard deviation 5.649

Table 39: Reliability and item statistics for Factor 3: *Individual Control of Training Outcomes* (n =229)

Item statistics	Sk	Ku	Mean	Standard deviation	Corrected item total correlation	Cronbach's alpha if the item is deleted
Q49	-1.22	1.38	5.36	1.387	0.544	0.701
Q57	-1.15	1.56	5.52	1.255	0.546	0.692
Q63	-2.23	7.48	6.26	0.968	0.691	0.634
Q64	-1.54	3.08	5.83	1.139	0.452	0.741
Reliability statistics				N of items 4	Mean inter-item correlation 0.558	Cronbach's alpha 0.750
Scale statistics				Mean 22.97	Variance 13.109	Standard deviation 3.621

An examination of the reliability analyses in Tables 37 to 39 showed that the corrected item correlations of all three factors indicate that the specific items would make good components of a summated rating scale. The mean inter-item correlations scores on the three factors also satisfy the requirements of homogeneity and unidimensionality suggested by Clark and Watson (1995). An examination of the scores of the mean inter-item correlations in Tables 37 to 39 indicates that the items measured a narrow or well-defined construct (Zeller & Carmines, 1980). The average inter-item correlations for the three factors all yielded exceptionally high values (Factor 1 = 0.663; Factor 2 = 0.632 and Factor 3 = 0.558).

Clark and Watson (1995) suggest that the specificity of the target construct is prominent when the average inter-item correlation exceeds 0.50. Using a panel of experts to initially scrutinise construct items (from an application of the Lawshe method) considerably improved the quality of the items and, more importantly, trimmed invalid items, thereby retaining superior final variables for scale development.

Based on the results reported above, all the items of the three factors in each table were retained as separate indicators to measure airline pilots' perceptions of the training climate associated with advanced automated aircraft. An inspection of the items in Factor 1 (Table 37), Factor 2 (Table 38) and Factor 3 (Table 39) reveals that all the item means are between 5 and 7, with an approximate standard deviation of 0.9 to 1.1, and that all the skewness coefficients are negative, ranging from -0.940 to -3.29. A normal distribution has a skewness index of zero. Thus, the distribution of each factor would have a long "tail" to the left and can be said to deviate somewhat from normality (Morgan & Griego, 1998:49).

Kurtosis determines whether the peak of the distribution is higher or lower than the ideal normal curve. In a normal distribution, kurtosis is equal to zero. An examination of the results revealed high kurtosis values, ranging from 0.57 to 11.20. Here, the positive kurtosis index implies very peaked curves. These deviations from the normal distribution were expected due to the homogeneous nature of the sample. The normality of the data and the implications of non-normal distributions for the selection of statistical procedures are explored in Section 5.6.

5.5 ITEM DISCRIMINATION ANALYSIS

The goal of the study from here on is to understand the nature and finer characteristics of the scaled construct. According to Brown (1983), researchers should ensure that their measurements are as valid and as reliable as possible when phenomena are measured on the basis of psychological scales. An item discrimination analysis was deemed necessary in order to fulfil Brown's (1983) requirement for increased proof of validity and reliability.

A scale is considered more effective when its items differentiate successfully between the proportions of high and low scorers (Cortina, 1993; Leech *et al*., 2005). A well-formulated item should have the predictive power to place top scoring participants into an upper group and lower scoring participants into a lower group.

To determine the level of item-discrimination in the scale of the current study, a discriminant function analysis was used to classify a dichotomous dependent variable based on information from continuous or ordinal independent variables (in this case, the factor items). The latent construct factors for this process were considered independent for the purposes of the analysis (in other words, reversal). Cases were allocated a place in a dichotomous group labelled *Discriminant Category*. The formulated *Discriminant Category* group became the dependent variable for the purposes of this computation. The group contained two dummy variables, 0 and 1. Each placement was based on whether the mean inter-item score was in the upper bound (>5.0, Dummy Variable 1) or the lower bound (<4.99, Dummy Variable 0). Prior to conducting the analysis, however, a matrix scatter plot (see Figure 27) was used to analyse the basic assumptions required for proper model specification.

Figure 27 depicts a comparison of the scatter plots for the upper bound (Dummy Variable 1) and the lower bound (Dummy Variable 0) groups. It appears that the scatter plots for the same variables are similar with regard to their variability for the two groups. Therefore, the assumption of the homogeneity of variance-covariance matrices was met and additional quantitative analyses based on the discriminant function could proceed. Table 40 sets out the results of the discriminant significance tests based on Wilks's lamda for each item. Wilks's lamda is the statistic of choice to separate variable classes (Mardia, Kent, & Bibby, 1979).

Figure 27: Matrix scatterplot for the discrimination of classes

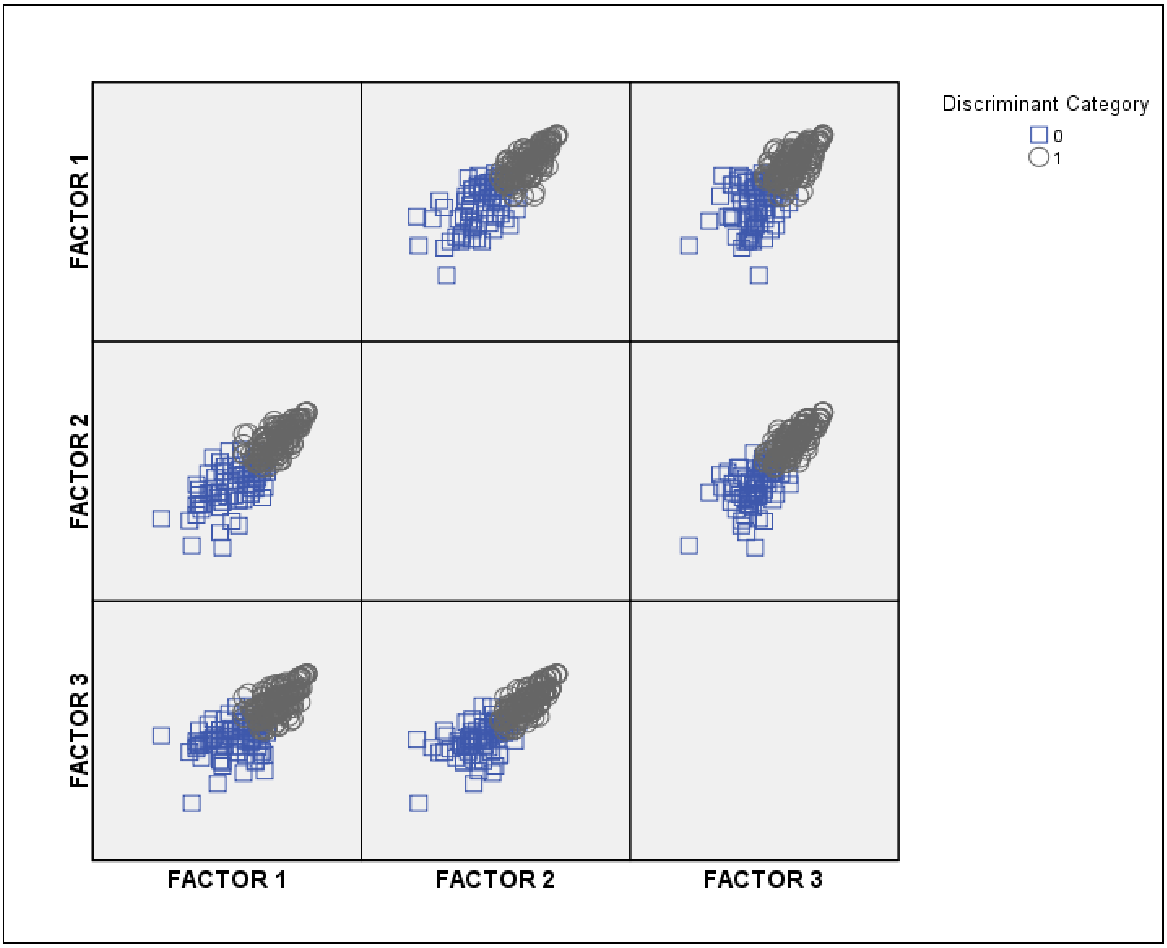


Table 40: Tests of equality of the discriminant group means

Item		Wilks's lambda	F	df1	df2	Sig.
Q23	My company's training produces world-class pilots.	0.820	49.742	1	227	0.000
Q24	Training at my airline is in line with company goals.	0.838	43.956	1	227	0.000
Q25	I know what my company's training goals are.	0.846	41.381	1	227	0.000
Q26	My company has talented people in training.	0.867	34.969	1	227	0.000
Q27	Training on this aircraft is professional.	0.808	54.019	1	227	0.000
Q28	Management follows the rules and regulations appropriately.	0.865	35.433	1	227	0.000
Q29	Training on this aircraft is well organised.	0.841	42.902	1	227	0.000
Q30	I understand what the company expects of me when training.	0.801	56.445	1	227	0.000
Q31	Training at my airline produces safe pilots.	0.827	47.510	1	227	0.000
Q32	The airline gives its pilots an appropriate amount of preparation work for training.	0.911	22.099	1	227	0.000
Q33	If I had to experience a problem in training, it's easy for me to appeal.	0.880	30.986	1	227	0.000
Q34	There is sufficient training guidance from the company.	0.866	35.184	1	227	0.000
Q36	I'm given sufficient time to prepare for training on this aircraft.	0.850	39.908	1	227	0.000
Q38	The airline is very supportive of its pilots' learning requirements for this aircraft.	0.839	43.424	1	227	0.000
Q39	My company's culture supports training for new technology aircraft.	0.879	31.332	1	227	0.000
Q40	There is sufficient feedback about my training on this aircraft.	0.860	37.087	1	227	0.000
Q41	Training is in line with civil aviation regulations.	0.860	37.088	1	227	0.000
Q42	My company uses only current training material.	0.942	13.885	1	227	0.000
Q44	I operate well as a crewmember in the simulator.	0.731	83.422	1	227	0.000
Q45	My instructor is willing to listen.	0.793	59.407	1	227	0.000
Q49	The instructors on this aircraft don't overload us with information.	0.814	51.897	1	227	0.000
Q50	Pilots are in direct control of the training outcome.	0.743	78.615	1	227	0.000
Q51	Preparation improves performance.	0.801	56.292	1	227	0.000
Q52	I try never to be late for a training session.	0.874	32.751	1	227	0.000
Q53	I co-operate when training in a simulator.	0.862	36.387	1	227	0.000

Item		Wilks's lambda	F	df1	df2	Sig.
Q55	I have a positive relationship with my colleagues.	0.779	64.554	1	227	0.000
Q57	I'm in control of the outcome of a training session.	0.719	88.890	1	227	0.000
Q58	I enjoy studying the technical aspects of the aircraft.	0.741	79.141	1	227	0.000
Q60	I read to understand so as to gain a deeper understanding of this aircraft's systems.	0.718	89.171	1	227	0.000
Q61	It's a good idea to know more than what is required.	0.770	67.908	1	227	0.000
Q62	I aim to gain a deeper understanding of this aircraft.	0.700	97.463	1	227	0.000
Q63	I'm comfortable undergoing training for this aircraft.	0.678	107.95	1	227	0.000
Q64	I can control my anxiety so as to perform well in training.	0.711	92.195	1	227	0.000

According to the values set out in Table 40, Wilks's lambda varied from 0.678 to 0.942, indicating that the variables differentiated moderately between the upper and lower groups (Mardia, *et al.*, 1979). The F-test of Wilks's lambda shows the variables' contributions that are significant. The F-test significance levels for all the items were good ($p < 0.001$), which in turn indicated that each of the items in the scale was a significant discriminant group predictor by itself.

In addition, the computation of the chi-square test statistic for the data also confirmed the inequality of the location of the mean scores of the upper bound and the lower bound groups (χ^2 [1,127] = 278.347; $p < 0.001$), and attested that the 33 items in the combined scale (predictors) are able to separate the upper and lower discriminant groups effectively and significantly, suggesting that the scale has an appropriate discriminant ability.

5.6 DATA EXPLORATION: ANALYSIS OF DISTRIBUTION

Normality of data is an underlying assumption of parametric statistical testing; therefore testing that the scores are normally distributed and variances of groups are equal is a prerequisite for studies of this nature (Field, 2005). In order to decide on the most appropriate family of statistics to analyse the summated scale scores further, variable score distributions were examined. Table 41 depicts the descriptive statistics of the instrument and some fundamental demographics.

Variables were considered continuous scales when measuring their means, the standard deviations, kurtosis and skewness of scores (because each item was based on a Likert design). The rejection of the assumption of normality can be determined from key statistics. In cases where the skewness and kurtosis are more than 2.5 times the standard error, the assumption of normality of data is violated (Morgan & Griego, 1998:49).

Next, the Kolmogorov-Smirnov and Shapiro-Wilks tests were used to test goodness-of-fit to obtain a statistic (a z-value) and a p-value. Apart from the three scales, the test was also applied to important demographic variables that described the sample. These were further collapsed into categories of the airline company size (small, medium, large) in Table 42 for further scrutiny.

It was important to determine the extent of non-normality amongst the airline carriers in terms of their sizes, because the literature demonstrates that the number of aircraft operated by an airline can directly influence organisational phenomena such as training. For instance, there is a direct correlation between the number of aircraft in a fleet and the preponderance of schedule frequencies, the number of pilots and the resources available (Wilson & Weston, 1989). These results are reported in Table 41.

The results show that the null hypothesis that either the demographic or the factor space samples come from a one-dimensional normal probability distribution should be rejected ($p < 0.05$) for all three factors and for the majority of the demographic variables. The two components of normality (Sk and Ku) were assessed (see Table 41). The results show that the mean was not in the centre of the distribution for either the continuous independent variables or the latent factors. In fact, the negative skewness computed for the three scales attested to the clustering of cases to the right of the distributions. The opposite was true for the distributions of the independent variables. Similarly, an investigation of kurtosis reveals that the distributions of the three scales were strongly peaked, with short, thick tails. Kurtosis values below zero indicated that the data associated with this index had a flat distribution, with many of the cases in the tails.

Table 41: Descriptive and distribution statistics of the three scales and continuous independent variables (n=229)

Measurement scale	Mean score	Standard deviation	Skewness		Kurtosis	
			Sk	Std. error	Ku	Std. error
Organisational professionalism	115.86	18.661	-1.241	0.161	2.316	0.320
Intrinsic Motivation	56.71	5.649	-1.392	0.161	1.506	0.320
Individual Control of Training Outcomes	22.97	3.621	-1.328	0.161	2.613	0.320
Age	41.28	11.36	0.375	0.161	-0.948	0.320
Experience (years)	20.75	11.66	0.500	0.161	-0.877	0.320
Flying time (hours)	9753.29	6116.72	0.731	0.161	-0.215	0.320
Digital time (hours)	4176.23	3216.05	0.875	0.161	0.307	0.320

Basic sample descriptive statistics, such as age, experience and flying time, are conceptually related to one another and exhibit similar distributive characteristics. Non-normality in the independent variables was expected, because, given the nature of the industry, the sample was extremely homogeneous, and all the participants had high levels of flight experience. Because the skewness and kurtosis of data do not indicate how close to normality a distribution actually is, and these indicators tend to deal with only one aspect of non-normality each (Field, 2005, 2009), to be truly useful, demographics relating to the size of the organisation had to be examined using the Kolmogorov-Smirnov and Shapiro-Wilks goodness-of-fit tests, bearing in mind the guideline that the “[s]creening of continuous variables for normality is an important early step in almost every multivariate analysis” (Tabachnick & Fidell, 2007:79).

The results of the objective statistical tests are set out in Table 42.

Table 42: Statistical tests for normality

Variables (omnibus)	Kolmogorov-Smirnov			Shapiro-Wilks			
	Statistic	df	Sig.	Statistic	df		Sig.
Age	0.098	229	0.000	0.953	229		0.000
Experience	0.117	229	0.000	0.938	229		0.000
Flying time	0.093	229	0.000	0.934	229		0.000
Digital time	0.141	229	0.000	0.918	229		0.000
Organisational Professionalism	0.096	229	0.000	0.950	229		0.000
Intrinsic Motivation	0.087	229	0.000	0.955	229		0.000
Individual Control of Training Outcomes	0.063	229	0.030	0.976	229		0.001

Variables (categorised)	Size of carrier	Mean	Kolmogorov-Smirnov			Shapiro-Wilks		
			Statistic	df	Sig.	Statistic	df	Sig.
Age (years)	Large	44	0.092	135	0.008	0.958	134	0.000
	Medium	33	0.164	48	0.002	0.858	49	0.000
	Small	41	0.112	46	0.189*	0.932	46	0.010
Experience (years)	Large	24	0.094	135	0.005	0.957	134	0.000
	Medium	13	0.176	48	0.001	0.836	49	0.000
	Small	20	0.193	46	0.000	0.870	46	0.000
Flying time (hours)	Large	11993	0.076	135	0.057	0.969	134	0.004
	Medium	5708	0.176	48	0.001	0.851	49	0.000
	Small	7537	0.188	46	0.000	0.801	46	0.000
Digital time (hours)	Large	5378	0.071	135	0.095*	0.972	134	0.008
	Medium	2663	0.225	48	0.000	0.829	49	0.000
	Small	2287	0.259	46	0.000	0.690	46	0.000
Organisational Professionalism	Large	5.86	0.096	135	0.004	0.897	134	0.000
	Medium	5.67	0.080	48	0.200*	0.972	49	0.284*
	Small	5.47	0.090	46	0.200*	0.923	46	0.005
Intrinsic Motivation	Large	5.35	0.158	135	0.000	0.886	134	0.000
	Medium	5.39	0.187	48	0.000	0.863	49	0.000
	Small	5.19	0.204	46	0.000	0.873	46	0.000
Individual Control of Training Outcomes	Large	4.33	0.118	135	0.000	0.914	134	0.000
	Medium	4.32	0.193	48	0.000	0.846	49	0.000
	Small	4.38	0.167	46	0.003	0.916	46	0.003

Significant when P < 0.05

* Not significant

The two-tailed test for significance in Table 42 shows that for the vast majority of the continuous independent variables and latent dependent factors, the distribution of the scores was statistically non-normal ($p < 0.01$).

The Shapiro-Wilks test provides a far superior computation of the test statistic (Field, 2009). Therefore, an appropriate family of non-parametric methods was used to explore phenomena in the data further, and determine the statistical significance of the relationship between the various characteristics of South African airline pilots and their perceptions of the advanced automated aircraft training climate.

5.7 RESULTS OF THE NON-PARAMETRIC COMPARATIVE STATISTICS USED TO EXPLORE PHENOMENA

The previous section has dealt with an in-depth exploratory analysis of the underlying characteristic phenomena based on the observed evidence, which manifested from within the data. However, to gain a deeper understanding of what other phenomena might be present, it was decided to conduct a series of non-parametric comparative tests. This helped the researcher, as a first step, to enhance the subsequent more complex statistical analysis in the current study. The nature of the study is primarily exploratory. In this section, non-parametric comparisons were conducted on broad categories of the independent variable. Any significance was then followed up with appropriate *post hoc* univariate examinations. Thereafter, a detailed exploration of the phenomena was conducted using a non-parametric MANOVA technique, as reported in Section 5.9 and its sub-sections.

The Kruskal-Wallis test was conducted to compare multivariate data. When data are deemed ordinal and the assumptions of the equality of group variances are violated, appropriate non-parametric statistics such as these are recommended (Field, 2005, 2009). The mean rank scores of various independent categorical groups were compared and examined to guide the more complex analyses in the study.

Where significant differences were found, a *post hoc* non-parametric test, the Mann-Whitney test, was used to determine the statistical significance of the actual difference between the highest and lowest ranking categories. When the means of the various demographic categories are tested, the most appropriate statistic used in parametric statistics is Student's t-test. However, in a case such as this, either the Mann-Whitney U or the Wilcoxon W would more be useful or appropriate as a non-parametric alternative (Green & Salkind, 2008). The Mann-Whitney test was therefore applied to the data as the equivalent to the independent sample's t-test to evaluate the differences between the medians. This test is useful in determining the significance of differences between the mean ranks for dyadic categorical independent variables and continuous dependent variables (Field, 2005; Leech *et al.*, 2005). In addition, sample sizes in excess of 40 are deemed more appropriate and large enough for this type of non-parametric comparison to yield accurate p-values (Green & Salkind, 2008). Reporting the effect of loss in statistical power may then not be necessary for Mann-Whitney comparative tests when sample sizes are sufficiently large (Field, 2009; Green & Salkind, 2008).

Although Field (2005:550) recommends applying the "Bonferroni correction" as a preventative measure against inflated Type 1 errors when conducting numerous *post hoc* tests, it was not a requirement for the present data set being examined. Therefore, combining independent variable classes to obtain sample categories that were sufficiently large and relevant for comparison mitigated the possibility of inflating Type 1 errors when conducting *post hoc* tests. Alternatively, the formula ($r = z/\sqrt{N}$) provided by Field (2005) was used as an indication of the effect size associated with any significant differences between the mean ranks of samples. It was assumed that $r \leq 0.10$ suggests no practical significance, $r > 0.10$ suggests a small effect, $r \geq 0.30$ suggests a medium effect, and $r \geq 0.50$ suggests a large effect.

The effect sizes in terms of small, medium or large effect sizes, as a method of power and practical significance analysis should, however, be read in context, because one should beware of an overly broad categorisation, based on what Ellis (2010:230) calls "T-shirt sizes". The computation of effect sizes and practical significance has been criticised by some authors as becoming increasingly generalised, which may then mask important alternative explanations or diminish useful considerations (Ellis, 2010). Thus, even where the effect size in the data seemed to diminish the impact of tests yielding significant p-values, it was considered in context for the purpose of the discussion of the nature of the phenomena observed. The concept of effect size, as originally proposed by Cohen (1988), has been adopted in the interpretation of the current research results in an effort to illuminate and complement the statistically significant findings.

The information provided in Tables 43 to 52 describes the strength in the differences or similarities between the various groups with regard to the latent behavioural factors observed in the construct. First, broad assessments on the behavioural scales in respect of the demographic groupings were made using a multivariate non-parametric comparative procedure (the Kruskal-Wallis or K-W procedure). Thereafter, a drill-down was done, using a two-sample non-parametric statistical analysis to examine differences in the mean ranks (M-W).

To examine the presence of any potentially ordered pattern in the medians of each group, the Jonckheere-Terpstra (J-T) test was used to assess the possibility of such data trends (Field, 2005, 2009). This test is useful when a researcher suspects that the order of the independent groups may be meaningful. Significant Jonckheere-Terpstra tests provide a valuable and useful basis for a more detailed understanding of the phenomena present within data. These analyses provided initial observable contact with the data set and also provided an indication of the correct path to follow for conducting further exploration. It was also then necessary to repeat the *post hoc* Mann-Whitney tests for some aspects of the data after discovering potentially important phenomena, such as Jonckheere-Terpstra trends.

Tables 43 to 52 set out the values for the chi-square and z-scores for the relevant significance tests, together with the two-tailed asymptotic levels. Some authors have argued that the one-tailed significance test cannot play an effective role in an exploratory examination of phenomena in nature (Clark & Watson, 1995). When sample sizes are relatively large, statistical significance is more likely, therefore good practice warrants an assessment of the effect associated with such significant findings (Cohen & Lea, 2004). The effect sizes were therefore generated for all *post hoc* analyses to determine the level of practicality whenever differences were detected in the data.

Table 43: Kruskal-Wallis test for the grouping variables flight deck position and size of carrier

Behavioural scale	Flight deck position	N	Mean rank	Chi-square	df	Asym. Sig. (2-tailed)
Organisational Professionalism	Co-Pilot Long-Range	39	121.46			
	Co-Pilot Short-Range	74	102.73			
	Captain Long-Range	31	121.65			
	Captain Short-Range	80	113.63			
	Total (5 missing)	224				
				3.071	3	0.381
Intrinsic Motivation	Co-Pilot Long-Range	39	104.49			
	Co-Pilot Short-Range	74	106.68			
	Captain Long-Range	31	119.32			
	Captain Short-Range	80	119.14			
	Total (5 missing)	224				
				2.393	3	0.495
Individual Control of Training Outcomes	Co-Pilot Long-Range	39	107.62			
	Co-Pilot Short-Range	74	113.97			
	Captain Long-Range	31	78.39			
	Captain Short-Range	80	126.74			
	Total (5 missing)	224				
				12.847	3	0.005*
Organisational Professionalism	Size of carrier	N	Mean rank			
	Large	135	126.37			
	Medium	48	104.33			
	Small	46	92.76			
	Total	229				
				10.416	2	0.005*
Intrinsic Motivation	Large	135	117.72			
	Medium	48	119.17			
	Small	46	102.67			
	Total	229				
				2.010	2	0.366
Individual Control of Training Outcomes	Large	135	113.48			
	Medium	48	114.44			
	Small	46	120.04			
	Total	229				
				0.341	2	0.843

*P < 0.05

Table 44: Mann-Whitney *post hoc* significance tests for the grouping variables flight deck position and size of carrier

Behavioural scale	Flight deck position	N	Mean Rank	M-W U	Z	2-tail Asymp. Sig. (effect size)
Individual Control of Training Outcomes	Captain Long-Range	31	39.98			
	Captain Short-Range	80	62.21			
	Total	111				
				743.5	-3.281	0.001* (0.31)
	All Short-Range Pilots	154	120.60			
	All Long-Range Pilots	70	94.67			
	Total (5 missing)	224				
				4 142.0	-2.791	0.005* (0.19)
	Co-Pilot Long-Range	39	40.06			
	Captain Long-Range	31	29.76			
	Total	70				
				426.5	-2.118	0.034* (0.26)
	All Co-Pilots	109	115.32			
	All Captains	120	114.71			
	Total	229				
				6 505.0	-0.070	0.944
Behavioural Scale	**Size of carrier**	**N**	**Mean rank**	**M-W U**	**Z**	**2-tail Asymp. Sig. (effect size)**
Organisational Professionalism	Larger operators	135	138.60			
	Smaller operators	94	81.71			
	Total	229				
				3202.0	-6.405	0.000* (0.42)

*P < 0.05

A non-parametric comparison of the subgroups for pilots' flight deck positions showed that the behavioural scale *Individual Control of Training Outcomes* was statistically affected by one of four flight deck positions that a pilot may occupy [$H(3) = 12.847$, $p < 0.05$]. The Mann-Whitney *post hoc* test was used to follow up on specific differences within the subgroups. The results show that short-range captains' scores were statistically higher than those of their long-range counterparts ($U = 743.50$, $p = 0.001$). It appears that the captains on short-range aircraft in the present sample felt that they had more control than the group of long-range captains over their levels of comfort in training and the related learning outcomes for the advanced aircraft they operate. The effect of this significance is regarded as medium. One possible reason for this difference may be the fact that short-range pilots have opportunities to fly more sectors than long-range pilots (therefore, they fly with a higher frequency, resulting in more exposure to their aircraft). Familiarity with an aircraft is directly related to the number of take-offs and landings performed (experience), and can therefore improve pilots' levels of confidence and comfort in working with the technology.

Similarly, there was a statistically significant difference between the mean ranked scores of *all* the short-range pilots and *all* the long-range pilots, but the effect is regarded as small ($U = 4142$, $p < 0.05$). Further examination of the long-range pilot group reveals that the co-pilots' scores were significantly higher than those of the captains ($U = 465.50$, $p < 0.05$, small effect). In general terms, captains have more overall flight experience. However, it appears that because a captain is normally a senior employee at an airline, much of a captain's flight experience is accounted for on analogue type aircraft (older generation aircraft) as opposed to digital or glass flight deck aircraft (which are more recent acquisitions at airline organisations). This difference may have affected the final behavioural scale scoring, and reduced their levels of perceived personal confidence in the advanced aircraft training they receive or have received.

Overall, pilots' perceptions on only the organisational professionalism scale were significantly affected by the size of the carrier where they were employed ($H[2] = 10.416$, $p < 0.01$). An analysis of a further *post hoc* two grouping of the large, medium and small categories (where South African Airways combined with BA Comair were considered larger operators, whilst the others were considered smaller operators, as defined in section 4.11), showed that perceptions of organisational professionalism were significantly affected (U = 3202.0, $p < 0.001$, medium effect). In addition, the order of the groupings (large, medium, small) was statistically significant (see Table 43). The Jonckheere-Terpstra test showed that this pattern was meaningful in that, as the size of the carrier increased, scores on the latent behavioural scale focused on perceptions of *Organisational Professionalism* increased statistically (J-T[3] = 4403, Std. J-T[3] = -6.036, $p < 0.001$). In other words, the larger the company in which the participant was employed, the more they would perceive their training as professional.

Table 45: Kruskal-Wallis test for the grouping variable interaction effect between experience in advanced aircraft and computer literacy

Behavioural scale	Interaction effect	N	Mean rank	Chi-square	df	2-tail Asymp.Sig.
Organisational Professionalism	Low Experience*Low Computer Literacy	19	76.00			
	Low Experience*High Computer Literacy	66	109.71			
	High Experience*Low Computer Literacy	73	130.42			
	High Experience*High Computer Literacy	71	114.49			
	Total	229				
				10.972	3	0.012*
Intrinsic Motivation	Low Experience*Low Computer Literacy	19	82.26			
	Low Experience*High Computer Literacy	66	125.74			
	High Experience*Low Computer Literacy	73	105.15			
	High Experience*High Computer Literacy	71	123.90			
	Total	229				
				2.393	3	0.025*
Individual Control of Training Outcomes	Low Experience*Low Computer Literacy	19	79.61			
	Low Experience*High Computer Literacy	66	138.09			
	High Experience*Low Computer Literacy	73	98.60			
	High Experience*High Computer Literacy	71	119.87			
	Total	229				
				18.485	3	0.000*

*P < 0.05

Table 46: Mann-Whitney *post hoc* significance tests for the grouping variable interaction effect between experience in advanced aircraft and computer literacy

Behavioural scale	Interaction effect	N	Mean rank	M-W U	Z	2-tail Asymp. Sig. (effect size)
Organisational Professionalism	Low Experience*High Computer Literacy	66	63.39			
	High Experience*Low Computer Literacy	73	75.97			
	Total	139				
				1973.0	-1.840	0.066
Intrinsic Motivation	Low Experience*High Computer Literacy	66	76.86			
	High Experience*Low Computer Literacy	73	63.80			
	Total	139				
				1956.5	-1.916	0.055
Individual Control of Training Outcomes	Low Experience*High Computer Literacy	66	82.48			
	High Experience*Low Computer Literacy	73	58.72			
	Total	139				
				1585.5	-3.492	0.000* (0.30)

*P < 0.01

Table 46 shows that overall, the effect of the interaction between a pilot's level of experience in advanced aircraft and her or his perceived computer literacy, significantly affects the pilot's perceptions of the advanced aircraft training climate ($p < 0.05$). Additionally, the mean rank scores on the Individual Control of Training Outcomes (Factor 3 score) behavioural scale is significantly different between pilots who reported low experience in advanced aircraft, combined with a high level of computer literacy (Mean Rank = 82.48) and the scores of pilots who reported high experience in advanced aircraft, combined with a low level of computer literacy (Mean Rank = 58.72).

The effect of the difference between pilots who reported low experience in advanced aircraft, combined with a high level of computer literacy, was regarded as medium ($U=1585.5$, $p < 0.001$; $r = 0.30$), and may have important consequences for airline training organisations. It appears that many hours of flight experience alone is not necessarily an important attribute in learning to operate the most advanced commercial aircraft. Organisations should consider that technologically averse individuals might have difficulty training for advanced automated aircraft.

It appears that the trainees' perceptions of their own computer literacy have a much larger influence in terms of interacting with pilots' flight experience levels at whether they feel that they may have more control of their training outcomes. More in-depth *post hoc* analysis on this interaction effect is reported in an analysis of the general linear model in Section 5.9.

Further examination of the rating on the computer literacy that the participants allocated to themselves was necessary. The results set out in Table 47 indicate that pilots' levels of computer literacy (poor, average, above average, excellent) significantly affects their *Intrinsic Motivation* and *Individual Control of Training Outcomes* regarding training for flying advanced technology aircraft ($H[3] = 13.291, 19.450$, $p < 0.05$). A significant difference was found between the mean ranked scores of pilots who perceived their computer literacy as low and the scores of those who perceived their computer literacy as higher ($U=4961.0, 4432.5$, $p < 0.01$, small to medium effect). Once again, these results show the importance of basic computer skills and competence in technology for trainees' perceptions of their training for flying advanced aircraft (see Table 47).

It was also necessary to explore whether a pattern existed in participants' perceived levels of computer literacy in terms of their learning for new technology aircraft. The Jonckheere-Terpstra test suggests that the order of computer literacy ratings (poor, average, above average, excellent) is statistically significant on the two behavioural scales at the individual level of analysis labelled *Intrinsic Motivation* and *Individual Control of Training Outcomes* ($J\text{-}T[4] = 10178.50, 10888.0$; Std. $J\text{-}T[4] = 2.948, 4.272$; $p < 0.01$).

These results attest that as pilots' perceived levels of computer literacy increase or improve, so too will their motivation to learn about new technology aircraft, and this affects their personal feelings about their ability to control the outcomes related to their training for flying advanced aircraft.

Table 47: Kruskal-Wallis test for the grouping variable computer literacy

Behavioural scale	Computer literacy	N	Mean rank	Chi-square	df	2-tail Asymp. Sig.
Organisational Professionalism	Poor	5	99.10			
	Average	87	120.34			
	Above average	92	110.92			
	Excellent	45	114.78			
	Total	229				
				1.202	3	0.752
Intrinsic Motivation	Poor	5	32.90			
	Average	87	104.30			
	Above average	92	122.76			
	Excellent	45	128.94			
	Total	229				
				13.291	3	0.004*
Individual Control of Training Outcomes	Poor	5	45.40			
	Average	87	97.51			
	Above average	92	123.25			
	Excellent	45	139.68			
	Total	229				
				19.450	3	0.000*

*P < 0.01

Table 48: Mann-Whitney *post hoc* significance tests for the grouping variable computer literacy

Behavioural scale	Computer literacy	N	Mean rank	M-W U	Z	2-tail Asymp. Sig. (effect size)
Intrinsic Motivation	Low Competence	92	100.42			
	High Competence	137	124.79			
	Total	229				
				4961.0	-2.738	0.006* (0.18)
Individual Control of Training Outcomes	Low Competence	92	94.68			
	High Competence	137	128.65			
	Total	229				
				4432.5	-3.823	0.000* (0.25)

*P < 0.01

Table 49 presents the aircraft that pilots operated according to manufacturer name. It shows that, overall, the type of aircraft operated by the pilot statistically affected their behaviour on the *Organisational Professionalism* scale and the *Individual Control of Training Outcomes* scale ($p < 0.05$). Furthermore, Table 49 clearly shows that when one compares the two largest subgroups that operate one of the main manufacturers' advanced aircraft (Boeing or Airbus), there is a significant difference in the pilots' scores on their *Individual Control of Training Outcomes* with regard to the climate for training on these aircraft ($U=2181.5$, $p < 0.05$, small effect). It appears that the Boeing pilots in the sample felt that they were more in control of their training outcomes than the Airbus pilots for this sample did. However, the effect of this difference should be regarded as small and needs to be interpreted with caution. It is recommended that this phenomenon be explored using a larger sample in future in order to obtain a more accurate effect size level.

Table 49: Kruskal-Wallis test for the grouping variable manufacturer

Behavioural scale	Manufacturer	N	Mean rank	Chi-square	df	2-tail Asymp. Sig.
Organisational Professionalism	Boeing	57	131.85			
	Airbus	95	129.53			
	Embraer	11	33.23			
	Canadair	9	87.44			
	De Havilland	7	103.93			
	Other	50	92.69			
	Total	229				
				32.454	5	0.000*
Intrinsic Motivation	Boeing	57	108.18			
	Airbus	95	116.41			
	Embraer	11	84.09			
	Canadair	9	103.50			
	De Havilland	7	119.86			
	Other	50	128.30			
	Total	229				
				5.403	5	0.369
Individual Control of Training Outcomes	Boeing	57	125.21			
	Airbus	95	104.15			
	Embraer	11	86.00			
	Canadair	9	90.39			
	De Havilland	7	139.00			
	Other	50	131.42			
	Total	229				
				11.357	5	0.045*

*P < 0.05

Table 50: Mann-Whitney *post hoc* significance test for the grouping variable manufacturer

Behavioural scale	Manufacturer	N	Mean rank	M-W U	Z	2-tail Asymp. Sig. (effect size)
Organisational Professionalism	Boeing	57	78.68			
	Airbus	95	75.19			
	Total	152				
				2583.5	-0.472	0.637
Individual Control of Training Outcomes	Boeing	57	85.73			
	Airbus	95	70.96			
	Total	152				
				2181.5	-2.014	0.044* (0.16)

*P < 0.05

Table 51: Kruskall-Wallis test for the grouping variables of initial (*ab initio*) training

Behavioural scale	Initial training	N	Mean rank	Chi-square	df	2-tail Asymp. Sig.
Organisational Professionalism	Military	81	127.39			
	Cadet	18	101.92			
	Self-sponsored (part-time)	64	118.45			
	Self-sponsored (full-time)	64	96.00			
	Total (2 missing)	227				
				9.083	3	0.028*
Intrinsic Motivation	Military	81	117.94			
	Cadet	18	96.86			
	Self-sponsored (part-time)	64	115.93			
	Self-sponsored (full-time)	64	111.90			
	Total (2 missing)	227				
				1.650	3	0.648
Individual Control of Training Outcomes	Military	81	109.93			
	Cadet	18	100.53			
	Self-sponsored (part time)	64	116.55			
	Self-sponsored (full-time)	64	120.39			
	Total (2 missing)	227				
				1.789	3	0.617

*P < 0.05

Table 51 and Table 52 suggest that a pilot's initial or *ab initio* flying training can significantly affect his or her perceptions of the advanced aircraft training climate at his or her airline with regard to the behavioural scale *Organisational Professionalism* [H(3) = 9.083, $p < 0.05$]. *Ab initio* training is defined as primary training, which a potential pilot undergoes when first entering the aviation industry (Moore *et al*., 2001).

The results indicate that there is a significant difference between the scores of those pilots who had military training (Mean Rank = 127.39) and the scores of pilots who had no military background (Mean Rank = 106.57) with regard to their views of *Organisational Professionalism* (U = 4828.5, $p < 0.05$).

Table 52 shows that, in terms of whether the candidate underwent a structured early training experience (that is an airline cadetship or military background) or an unstructured one (that is, participants who indicated that their primary training was concluded after self-sponsored or part-time methods), there was no impact on their perceptions of the advanced aircraft training climate with regard to the *Organisational Professionalism* scale.

It may be deduced from the results in Table 52, that having only a regimented, structured training background, as in the case of military trained pilots, may influence the candidate's perception of the professionalism associated with training for advanced aircraft. Although this information is useful, the practical significance of the differences in these categories should nevertheless be regarded as small, in terms of Cohen's (1988) criteria.

The discovery of the aforementioned effects between selected independent demographic variables and the latent factors of the main measurement construct was used as a basis for a further exploration of the possible phenomena that may exist in the dataset.

Table 52: Mann-Whitney *post hoc* significance test for the grouping variable nature of initial training

Behavioural scale	Nature of Initial training	N	Mean rank	M-W U	Z	2-tail Asymp. Sig. (effect size)
Organisational Professionalism	Structured training	99	122.76			
	Unstructured training	128	107.23			
	Total (2 missing)	227				
				5 469.0	-1.767	0.077
	Military-trained	81	127.39			
	Not military-trained	146	106.57			
	Total (2 missing)	227				
				4 828.5	-2.289	0.022* (0.15)

*P < 0.05

5.8 ASSOCIATIONAL STATISTICS: NON-PARAMETRIC MEASURES OF BIVARIATE RELATIONSHIPS

The general term explaining the concept of variable relationship is that, information about one variable is usually carried by the other variable in many instances (Cohen, Cohen, West & Alken, 2003). Exploiting the bivariate associational relationship between paired variables enhanced the exploration of the data set.

To assess the relationship or extent to which variables may be related and to determine the magnitude and subsequent direction of the possible relationship, a correlational analysis was conducted. Because the items were designed to comply with Likert's (1932) method, data were considered at least ordinal in nature. It was found that Pearson's coefficient would be unsuitable for the current study's data, as Pearson's formula is problematic where there are violations of normality or unequal variances. The difference in the non-parametric equivalent of the Pearson coefficient lies in the type of data that are used (Cohen & Lea, 2004; Comrey & Lee, 1992; Field, 2009). The non-parametric equivalent to Pearson's correlation coefficients is calculated by applying the formula to the ranks of the data, as opposed to applying it to the raw data values.

Kendall's tau-b was used as a measure of association between the variables chosen (Kendall & Stuart, 1963). Kendall's method was selected for its robustness in measuring the strength of the association between two ordinal or binary variables (Morgan *et al*., 2007). The Kendall tau-b (τ) statistic was also used because the computation allows for adjustments for ties (here, the geometric mean is used as an estimate of the relevant tied pairs). Kendall's tau-b is equivalent to Spearman's rho in terms of the underlying assumptions, and tau "has been emphasized recently as a substitute for *r* in various research contexts" (Walker, 2003:525). However, Spearman's rho and Kendall's tau are not identical in magnitude, since their underlying logic and computational formulae are relatively different (Kendall & Stuart, 1963; Gravetter & Wallnau, 2008). Relational strengths were thus found to be more conservative under the tau statistic.

When the associational process was replicated using Spearman's formula, statistically relevant strengths of association between variables were stronger to some extent (not reported here). Nonetheless, Kendall's formula was maintained for final computations and drawing conclusions in this thesis. The relationship between two measures using Kendall's tau formula is based on the goodness-of-fit of the least squares straight line; however, this correlation simply provides some evidence of a relationship and will therefore never prove causality *per se* (Tabachnick & Fidell, 2007). Cohen and Lea (2004) suggest a cut-off point of 0.30 (a medium effect size) for the practical significance of association.

The three main components of the construct under study were correlated with some core demographic data (see Table 53).

Table 53: Main demographic and factor correlations

Kendall's tau-b	Flight deck position	Interact_ Group	Adv. Aircraft Exp.	Age	Gender	Level of education	Level of computer literacy	Pilot unionisa-tion	Size of carrier	Instruc-tor rated	F1	F2	F3
Flt deck position	1.000	0.060	0.066	0.328**	0.128*	0.145*	0.000	0.176**	0.216**	0.057	0.003	0.075	0.068
Sig (2-tailed)	.	0.291	0.283	0.000	0.038	0.019	0.995	0.004	0.000	0.353	0.949	0.151	0.199
Interact_Group	0.060	1.000	0.810**	0.293**	0.060	-0.006	0.193**	-0.348**	-0.364**	0.069	0.082	0.053	-0.004
Sig (2-tailed)	0.291	.	0.000	0.000	0.325	0.918	0.002	0.000	0.000	0.262	0.104	0.307	0.941
Adv. Aircraft Exp.	0.066	0.810**	1.000	0.413**	0.058	-0.011	-0.279**	-0.442**	-0.452**	0.039	0.123*	-0.010	-0.100
Sig (2-tailed)	0.283	0.000	.	0.000	0.379	0.863	0.000	0.000	0.000	0.557	0.024	0.857	0.077
Age	0.328**	0.293**	0.413**	1.000	0.206**	0.122	-0.211**	-0.210**	-0.192**	0.035	0.143**	0.016	-0.050
Sig (2-tailed)	0.000	0.000	0.000	.	0.002	0.065	0.001	0.002	0.002	0.596	0.009	0.781	0.372
Gender	0.128*	0.060	0.058	0.206**	1.000	0.009	0.040	-0.002	0.045	0.048	0.072	0.097	0.038
Sig (2-tailed)	0.038	0.325	0.379	0.002	.	0.889	0.548	0.974	0.472	0.470	0.185	0.083	0.507
Level of education	0.145*	-0.006	-0.011	0.122	0.009	1.000	-0.011	0.033	-0.013	0.154*	0.031	0.046	0.037
Sig (2-tailed)	0.019	0.918	0.863	0.065	0.889	.	0.864	0.622	0.839	0.020	0.576	0.409	0.513
Level of com, literacy	0.000	0.193**	-0.279**	-0.211**	0.040	-0.011	1.000	0.133*	0.126*	0.054	-0.043	0.153**	0.216**
Sig (2-tailed)	0.995	0.002	0.000	0.001	0.548	0.864	.	0.044	0.047	0.414	0.433	0.006	0.000
Pilot unionisation	0.176**	-0.348**	-0.442**	-0.210**	-0.002	0.033	0.133*	1.000	0.854**	-0.025	-0.319**	-0.068	-0.010
Sig (2-tailed)	0.004	0.000	0.000	0.002	0.974	0.622	0.044	.	0.000	0.704	0.000	0.222	0.862
Size of carrier	0.216**	-0.364**	0.452**	-0.192**	0.045	-0.013	0.126*	0.854**	1.000	-0.012	-0.315**	-0.091	-0.027
Sig (2-tailed)	0.000	0.000	0.000	0.002	0.472	0.839	0.047	0.000	.	0.843	0.000	0.089	0.614

Kendall's tau-b	Flight deck position	Interact_ Group	Adv. Aircraft Exp.	Age	Gender	Level of education	Level of computer literacy	Pilot unionisa-tion	Size of carrier	Instruc-tor rated	F1	F2	F3
Instructor rated	0.057	0.069	0.039	0.035	0.048	0.154*	0.054	-0.025	-0.012	1.000	-0.023	0.050	0.111
Sig (2-tailed)	0.353	0.262	0.557	0.596	0.470	0.020	0.414	0.704	0.843	.	0.677	0.370	0.050
Organisational Professionalism (F1)	0.003	0.082	0.123*	0.143**	0.072	0.031	-0.043	-0.319**	-0.315**	-0.023	1.000	0.351**	0.448**
Sig (2-tailed)	0.949	0.104	0.024	0.009	0.185	0.576	0.433	0.000	0.000	0.677	.	0.000	0.000
Intrinsic Motivation (F2)	0.075	0.053	-0.010	0.016	0.097	0.046	0.153**	-0.068	-0.091	0.050	0.351**	1.000	0.396**
Sig (2-tailed)	0.151	0.307	0.857	0.781	0.083	0.409	0.006	0.222	0.089	0.370	0.000	.	0.000
Individual Control of Training Outcomes (F3)	0.068	-0.004	-0.100	-0.050	0.038	0.037	0.216**	-0.010	-0.027	0.111	0.448**	0.396**	1.000
Sig (2-tailed)	0.199	0.941	0.077	0.372	0.507	0.513	0.000	0.862	0.614	0.050	0.000	0.000	.

** $p < 0.001$;
* $p < 0.05$;
$\tau_s < 0.10$ suggests no effect;
$\tau_s \geq 0.10$ suggests a small effect;
$\tau_s \geq 0.30$ suggests a medium effect; and
$\tau_s \geq 0.50$ suggests a large effect

The correlation results depicted in Table 53 indicate that, for the present sample, the phenomena discussed below were noteworthy.

The position of the crewmember on the flight deck (either captain or co-pilot) is significantly correlated with age (τ_s = 0.328, $p < 0.001$, medium effect). Due to the very rigid seniority systems entrenched at the larger and unionised carriers in South Africa, a pilot is only eligible for command (a captain upgrade) after serving a prerequisite number of years in the organisation. The data therefore reflects this correlation. The results also show a small effect size in the relationship between a respondent's flight deck position and his or her level of education (τ_s = 0.328, $p < 0.001$). The data indicated that it is to be expected that, as a pilot gains experience and becomes more senior at an airline, the pilot's educational qualifications will improve.

The interaction effects between pilots' overall experience level on advanced aircraft and their perceived levels of computer literacy are significantly related to the ages of the pilots (τ_s = 0.293, $p < 0.001$, medium effect). This correlation suggests that levels of the experience-computer literacy interaction for this sample improved with the age of the pilots. It may therefore be relevant that older pilots who undergo advanced aircraft transition training have a vast amount of previous experience in advanced aircraft, together with better perceived computer literacy, and that this tends to improve their training experience. Similarly, it is intuitively logical to conclude that as people spend more years in an airline, they both age, and are likely to gain experience. However, levels of computer literacy do not display the same linear relationship. Therefore, older pilots may require higher levels of computer literacy or ability in order to report an improved perception of the advanced aircraft training they undergo.

The level of advanced aircraft flight experience of pilots in the sample was inversely related to their level of computer literacy (τ_s = -0.279, $p < 0.001$, medium effect), the degree to which the pilot group was unionised (τ_s = -0.442, $p < 0.001$, medium effect) and the size of the organisation (τ_s = -0.452, $p < 0.001$, medium effect). A surprising correlation in the data suggests that the more experience a pilot had in advanced aircraft, the less likely the pilot was to believe that he or she had good levels of computer literacy. This result may also be related to the effect of age on perceived levels of computer literacy, in that younger pilots tended to have more positive perceptions of their computer literacy ($\tau_s = -0.211$, $p < 0.001$, small to medium effect). It appears that the older pilots (over the age of 40 years) have had more experience on advanced aircraft ($\tau_s = 0.413$, $p < 0.001$), but less experience with commercial technology. This particular generation did not have as much exposure to computer-based technology while growing up as the younger generation has had (Moore, 2003). This effect is also related to a respondent's number of years of employment at the organisation.

Higher levels of computer literacy were significantly associated with greater pilot unionisation ($\tau_s = 0.133$, $p < 0.05$, small effect) and organisational size ($\tau_s = 0.126$, $p < 0.05$). These results suggest that the larger, unionised carriers employ pilots who perceive their levels of computer literacy to be relatively good. The results may also be related to the fact that one large unionised carrier in the sample issues personal laptop computers to its pilots. Such a pilot's familiarity with this item may then influence the pilot's perception of his or her computer literacy. Another explanation may be that, because pilots at highly unionised carriers command more earnings from complex negotiated agreements (Olney, 1996), they are more likely to be able to afford and enjoy the latest technologically advanced personal gadgets, such as laptops and tablet computers. Thus, these pilots are less likely to be averse to technology.

Greater unionisation of airline pilots is significantly related to high experience levels in advanced aircraft ($\tau_s = 0.442$, $p < 0.001$, medium effect). This was expected, as the large carriers generally hire the most experienced pilots in the industry. Unionised carriers are also more attractive to experienced pilots because of their strict seniority lists and benefits such as pensions and protection against a loss of their licences (Olney, 1996). Larger carriers in South Africa are known to operate the more advanced technology aircraft from the two main global aircraft manufacturers, Airbus and Boeing, which produce larger aircraft, flying longer distances. Pilots at the bigger, unionised carriers also fly higher frequency schedules, providing their pilots with an opportunity to gain more experience. Consequently, it was observed that in South Africa highly unionised carriers are significantly related to larger employers (τ_s

= 0.854, $p < 0.001$, large effect). These effects have important implications for smaller, non-unionised organisations. Such organisations are more likely to have inferior advanced aircraft training capability in terms of the perceptual behavioural scales, and therefore more effort should be made by these companies' management to ensure effective and efficient transfer of knowledge to their pilots, especially because inadequate training paradigms, structure and methodology can have an adverse impact on flight safety.

The latent factor *Organisational Professionalism* was positively related to a respondent's experience in advanced aircraft ($\tau_s = 0.123$, $p < 0.05$) and age ($\tau_s = 0.143$, $p < 0.001$). However, the effect size of this relationship was considered small. Similarly, the latent construct *Intrinsic Motivation* was positively associated with a pilot's level of computer literacy ($\tau_s = 0.153$, $p < 0.001$, small effect), which suggests that when these airline pilots perceive their computer literacy levels as improving, they may also have a greater interest in training for new technology. Furthermore, the third latent construct, *Individual Control of Training Outcomes*, was similarly correlated with pilots' perceptions of their computer literacy ($\tau_s = 0.216$, $p < 0.001$, small effect). This result indicates that as pilots begin to believe that their levels of computer literacy are relatively good, so too will their perceptions of their ability to be in control of, and to take charge of their advanced aircraft training outcomes.

The correlation matrix in Table 53 also shows that the three latent behavioural scales of the main measurement construct correlate with each other to a high degree ($p < 0.001$). This was expected because the construct was developed in terms of the systemic principle and factors derived from an oblique rotation. The tau inter-correlation coefficients for the factors ranged from 0.351 to 0.448.

Overall, the results of the non-parametric associational analysis should be interpreted within the present context and with a degree of caution. The non-parametric methods are considered more robust than parametric methods. Nonetheless, it was observed there were many low to medium effect sizes, according to Cohen's (1988) criteria, representing the overall effectiveness of significant associations. This indicates that many of the significant results can be dismissed from a practicality perspective. Again, however, effect sizes may also bear the brunt of a certain level of criticism for being subjectively broad in terms of unilaterally lessening the impact of a study, which Ellis (2010:230) dismisses as categorising significances according to “T-shirt sizes”. Nevertheless, it was necessary to understand why the data yielded such low to medium effect sizes. It was summarily hypothesised then, that one reason for this result might be from the impact of the relatively tiny and highly homogeneous sample frame and subsequent non-normality of the data set. These results pointed to areas for further exploration. Because of the lack of total clarity from this associational analysis, the aforementioned results were then used to guide further multivariate and regression statistical analyses, as discussed in the sections below.

5.9 NON-PARAMETRIC MULTIVARIATE ANALYSIS OF VARIANCE (MANOVA)

“Generalized linear models provide a unified theoretical and conceptual framework for many of the most commonly used statistical methods” (Dobson & Barnett, 2008:15). To examine the main and interactional effects of partially independent categorical variables on multiple dependent variables, a MANOVA was conducted by means of the general linear model (Anderson, 2001; Field, 2005). In this case, the dependent variables (factors) were conceptually related to a high degree. The risk of multicollinearity should be considered when correlations between the dependent variables are generally high, while conversely no correlation would imply that a multivariate analysis could be “pointless” (Leech *et al.*, 2005:177). A moderate median ranked correlation of 0.40 between the dependent variables was calculated, based on Kendall’s tau-b from an earlier associational exploration (τ, which makes adjustments for ties). This indicated that an analysis of a general linear model was indeed useful.

Based on a combination of the dependent variables, the general linear model procedure was used to compute a multivariate F. “The larger the value of F, the more likely it is that the null hypothesis (H_0) of no differences among the group means (locations) is false” (Anderson, 2001:34). The combination maximised the differentiation of the ordinal dependent variable groups. This procedure was followed to provide an analysis for *effects* on a linear combination of three dependent variables of multiple independent variables, or covariates. A MANOVA was computed to test the differences in the centroid or vector of medians of the multiple interval/ordinal dependent variables, for various categories of the independent variables. Because the Type I error rate of the standard MANOVA test statistics can be inflated, whereas their power attenuates when assumptions of normality and homogeneous covariance matrices are violated (Holmes, 2005), a non-parametric MANOVA was computed. Both Box's M-test (p = 0.001), which was used to assess the homogeneity of the variance-covariance matrices of the independent variables, and Levene's test of the equality of error variances in two of the three dependent variables were markedly violated (see Table 54).

Furthermore, assumptions on the dependent variable had to be considered as deviating from normality, because in a MANOVA, there is no single dependent variable as such, but rather a column matrix or vector of scores on each dependent variable.

Table 54: Tests for assumptions of normality and homogeneity

Levene's test of equality of error variances				
Ranked dependent factor or scale	F	df1	df2	Sig.
Organisational Professionalism (Factor 1)	1.462	51	177	0.037
Intrinsic Motivation (Factor 2)	1.938	51	177	0.001
Individual Control of Training Outcomes (Factor 3)	1.281	51	177	0.122
Box's M-test				
273.269	1.426	138	3699.387	0.001

In this situation (as portrayed in Table 54), a non-parametric or rank-order variant of the MANOVA was an appropriate option, as proposed by Zwick (1985). Original or raw observations were transformed by ranking the subjects on each of the dependent variables. The ranks were then subjected to a conventional MANOVA. However, the required test statistic in this case was equal to (N-1)*V*, where *V* is the Pillai-Bartlett statistic computed on the transformed data (Zwick, 1985). Five independent variables based on a previous theoretical premise and earlier associational analysis, together with appropriate strengths of association (phi), were selected and tested. A breakdown of the frequencies in terms of the independent categories is provided in Table 55.

To guide an examination and interpret the differences between the vectors of the mean ranked scores between groups, the following situation was subsequently investigated:
What are the phenomena that affect airline pilots who differ in

- *age* (40 years and younger, or over 40 years);
- *level of digital flight time experience* (where high experience meant 2001 hours or more on advanced aircraft, and low experience meant 2000 hours or less);
- *company status* (captain or co-pilot);
- *size of carrier* (employed at a large, medium or small company); and
- *level of computer literacy* (poor, average, above average, excellent)

on some linear combination of the three dependent factors (*Organisational Professionalism*, *Intrinsic Motivation* and *Individual Control of Training Outcomes*). In addition, interaction between the demographic levels in a distinguishing linear combination of the dependent variables was examined. The number of respondents in each category was collapsed and is depicted in Table 55.

Table 55: Frequency of between-subjects factors

Independent demographic grouping	Sub-grouping	Valid N
Age category	40 and under	119
	41 and over	110
Level of digital flight time experience	High digital	144
	Low digital	85
Company status	Captain	120
	Co-pilot	109
Size of carrier	Large	135
	Medium	48
	Small	46
Computer literacy	Poor	5
	Average	87
	Above average	92
	Excellent	45

The results of the non-parametric MANOVA of the five demographic variables in terms of the respondents' perceptions of the training climate associated with advanced aircraft are set out in Table 56.

Table 56: Omnibus Pillai-Bartlett multivariate test of significance

Effect		Value	F	Hypo-thesis df	Error df	Sig.	Partial eta squared	Observed power
Intercept	Pillai-Bartlett Trace	0.618	94.182	3	175	0.000	0.618	1.000
Age category	Pillai-Bartlett Trace	0.012	0.721	3	175	0.540	0.012	0.202
Digital flight experience	Pillai-Bartlett Trace	0.026	1.533	3	175	0.208	0.026	0.400
Company status	Pillai-Bartlett Trace	0.012	0.681	3	175	0.565	0.012	0.192
Size of carrier	Pillai-Bartlett Trace	0.194	6.311	3	352	0.000	0.097	0.999
Computer literacy	Pillai-Bartlett Trace	0.125	2.555	3	531	0.007	0.042	0.940
Digital flight experience* Computer literacy	Pillai-Bartlett Trace	0.149	3.090	3	531	0.001	0.050	0.976

The multivariate Pillai-Bartlett table tests the hypothesis that airline pilots, based on the selected demographics, do not differ significantly in terms of their overall perception of the advanced aircraft training climate. The result of the MANOVA depicted in Table 56 indicates that age of the respondents (F = 0.721; p = 0.540), digital flight experience (F = 1.533; p = 0.208) and whether the respondent was a captain or first officer (F = 0.681; p = 0.565) had no noticeable effect.

Then again, the *size of the carrier* in which the respondent is employed, the person's *level of computer literacy* and *digital flight experience*computer literacy* interaction have a substantive effect on the respondents' perceptions. For all three effects, the observed significance level for the Pillai-Bartlett test was at a 0.01 level of significance.

It appears that the size of the carrier (company) is possibly the most important independent variable in the model, combined with a high, observed power, indicating that the chance of failing to detect an effect that is present is less than a 1% (Arrindell & Van der Ender, 1985; Cohen, 1988). The Pillai-Bartlett trace was equal to 0.194, with an associated F (3, 229) = 6.311, $p < 0.001$ (Table 56). The squared eta of 0.097 indicates that the size of the carrier explained almost 10% of the variance in the specified model. The chi-square test statistic furthermore confirmed the inequality of the location of the median ranked scores of the three subgroups for the size of the company [χ^2 (3) = 1438.908, $p < 0.001$].

The effect of respondents' *digital flight experience*computer literacy* interaction (the multivariate Pillai-Bartlett's trace was 0.149; F [3, 229] = 3.090; $p < 0.001$) was significant, with a high observed power, providing only a 2% chance of failing to detect an effect which exists (Arrindell & Van der Ender, 1985; Cohen, 1988, Zwick, 1985). This indicates that the practical implications of subjects' experience in advanced aircraft with regard to their perception of the advanced aircraft training climate was highly dependent on their level of computer literacy.

The chi-square test statistic also confirmed the inequality of the localities of the median ranked scores for the following subgroups of the *digital flight experience*computer literacy* interaction [χ^2 (3) = 33.972, $p < 0.001$]:

- low advanced aircraft experience and poor, or average, or above average or excellent computer literacy (four subgroups); and
- high advanced aircraft experience and poor, or average, or above average or excellent computer literacy (four subgroups).

Overall, it appears that airline pilots' perceived levels of computer literacy had a significant effect on their experiences with the advanced aircraft training climate (multivariate Pillai-Bartlett's trace = 0.125; F [3, 229] = 2.555; $p < 0.01$). In addition, the chi-square test statistic also confirmed the inequality of the location of the median ranked scores for the four subgroups of computer literacy [χ^2 (3) = 28.5, $p < 0.001$].

5.9.1 Between-subjects effects

To determine exactly where the variations in the median ranked centroids were for each of the subgroups, across the three latent behavioural factors (*Organisational Professionalism*, *Intrinsic Motivation*, and *Individual Control of Training Outcomes*), it was necessary to proceed with an analysis of between-subjects effects. The examination was based on the results of the significance tests, and effect sizes were determined using Cohen's (1988) criterion of the partial eta squared (η^2). The recommendation for practical significance, based on this computation, is that there is a small effect size when $\eta^2 = 0.01$ (1%), a medium effect size when $\eta^2 = 0.06$ (6%) and a large effect size when $\eta^2 = 0.15$ (15%). The between-subjects effects was only examined for significant differences with regard to the size of the carrier, computer literacy and the interactive effect of digital flight experience combined with subjects' levels of computer literacy. For the size of the carrier, the results show that respondents' perceptions of the advanced aircraft training climate was significant ($p < 0.001$ with a medium effect) only for the *Organisational Professionalism* behavioural scale. The size of the carrier in which the pilot is employed appears to account for 13.9% of the variability in the pilot's perception of the professionalism in training for advanced aircraft. No statistically significant differences ($p = 0.724$) were noted between airline pilots' levels of computer literacy and their median ranked scores on the *Organisational Professionalism* scale. However, trainee's level of computer literacy had a small to medium effect ($\eta^2 = 0.047$ to 0.072) on the trainee's judgement on the *Intrinsic Motivation* and *Individual Control of Training Outcomes* scales respectively. Therefore, approximately 7.2% of the variance in the airline pilots' perceptions of their ability to control an outcome of training for advanced aircraft was related to their perceived level of computer literacy. These results show that information regarding computer skill and competence can be very useful in practical situations, such as airline pilot recruitment. To a recruiter, this may indicate a candidate's understanding, easiness and aptitude to operate advanced technology, which may in turn enhance a pilot's ability to learn successfully, and confidently operate the modern digital aircraft.

Table 57: Significance tests for between-subjects effects for Factors 1, 2 and 3

Source	Dependent variable	df	Mean square	F	Sig.	Partial eta squared	Observed power	Effect size (Cohen, 1988)
Corrected Model	*Organisational Professionalism*	51	8 661.756	2.746	0.000	0.442	1.000	Large
	Intrinsic Motivation	51	6 932.425	1.916	0.001	0.356	1.000	Large
	Individual Control of Training Outcomes	51	6 991.335	1.952	0.001	0.360	1.000	Large
Intercept	*Organisational Professionalism*	1	51 0716.861	161.885	0.000	0.478	1.000	Large
	Intrinsic Motivation	1	69 4995.446	192.101	0.000	0.520	1.000	Large
	Individual Control of Training Outcomes	1	71 9711.892	200.957	0.000	0.532	1.000	Large
Size of carrier	*Organisational Professionalism*	2	45 077.610	14.289	0.000	0.139	0.999	Medium
	Intrinsic Motivation	2	333.460	0.092	0.912	0.001	0.064	Insig.
	Individual Control of Training Outcomes	2	1 827.355	0.510	0.601	0.006	0.133	Insig.
Computer literacy	*Organisational Professionalism*	3	1 391.231	0.441	0.724	0.007	0.137	Insig.
	Intrinsic Motivation	3	10 429.173	2.883	.037	0.047	0.681	Small
	Individual Control of Training Outcomes	3	16 458.487	4.596	0.004	0.072	0.884	Medium
Level of digital flight time experience * Computer literacy	*Organisational Professionalism*	3	11 045.181	3.501	0.017	0.056	0.773	Small
	Intrinsic Motivation	3	19 322.494	5.341	0.002	0.083	0.929	Medium
	Individual Control of Training Outcomes	3	26 389.302	7.368	0.000	0.111	0.984	Medium

Finally, Table 57 shows that the level of *interaction* between Digital flight time experience and Computer literacy, significantly affects pilots' perceptions of the training climate associated with advanced aircraft training. The analysis of between-subjects effects show a significant relationship (p = 0.05 to 0.001) between Digital flight experience*Computer literacy interaction for all three latent factors in the measurement construct with a small to medium effect, according to Cohen's (1988) criterion.

These results show that the respondent's perceptions on the *Organisational Professionalism*, *Intrinsic Motivation* and *Individual Control of Training Outcomes* behavioural scales depend strongly on the joint effect of the participant's advanced aircraft flight experience, combined with his or her levels of computer competence. In other words, the two variables independently may have very little impact on perceptions of the training climate associated with the advanced aircraft; however, taken in unison, they appear to interact in such a way that this joint effect significantly separates categorical groupings on the various perception scales.

5.9.2 Non-parametric comparative *post hoc* tests for independent samples (Mann-Whitney) based on the GLM results

An omnibus F-test indicated only that the centroids of the median ranked scores were not co-located (Tabachnick & Fidell, 2007). Therefore, a series of *post hoc* non-parametric comparisons were carried out between bivariate subgroups from each category to ascertain the exact location of significant differences. Because the distributions of the scores on the dependent variables were regarded as non-normal, the Mann-Whitney test was once more used to determine any significant differences. The method is ideal for unequal and small sample sizes (Babbie, 2010), as in the case of the computer literacy subgroups. The z-value provided by the test is an indication of whether the two subgroups come from the same distribution. However, other common interpretations of the Mann-Whitney non-parametric procedure are that the test actually checks the median equality of the two samples (Field, 2005). Table 58 depicts the calculated z-values and effect size for the three categories in terms of the size of the carrier, and four categories related to the respondent's level of computer literacy. Table 59 in turn depicts the results for the calculated z-values and effect size for the six categories of Digital flight time experience* Level of computer literacy.

5.9.3 Size of the carrier

Results from the non-parametric test (see Table 58) indicate that there is a significant difference between the perceptions of pilots employed at either a small organisation or a large organisation ($p < 0.01$). When one compares the large airline pilots' scores to those of the pilots from the group of smaller airline pilots, the scores of the pilots from the larger carriers are significantly higher only on the *Organisational Professionalism* scale (at a macro and intermediate level of the main measurement construct; $Z = -2.941$, $p < 0.01$, with a medium effect size, $r = 0.396$). The results indicate that, it is likely that pilots employed at the relatively larger airlines will have a better perception of the training and organisational structures that are in place with regard to the training they experience for advanced aircraft. According to Child (1973), "Size of organization has often been cited as the attribute having the greatest single influence on the extent to which organizations develop bureaucratic forms of organization structure". In addition, larger enterprises have access to far more resources and greater budgets, thus creating an impression of professionalism. Another hypothesised reason for this phenomenon may be that because there are more opportunities for trainees to come into contact with flight instructors and management outside of training events at smaller airline companies, it may begin to create a casual atmosphere for learning; by virtue of over-familiarity, perceptions of professionalism can be reduced (Katz & Khan, 1966).

The medium effect size of the difference between larger and smaller carriers implies that it is of practical importance to understand the phenomenon of organisational professionalism and its subsequent impact in designing and developing training interventions and methodologies. In addition restructuring specific areas of the training may be necessary to enhance perceptions of training professionalism. For relatively small operators, it may be prudent to ensure that the training centre is well organised in terms of available learning resources, and that pertinent information about training issues (such as programmes, the names of instructors and timetables) are communicated effectively and timeously.

Table 58: Non-parametric comparison of mean rank scores by size of carrier and level of computer literacy

Latent behavioural scale	Variable	N	Mean rank score	U	Z	Asym. Sig. two-tailed	Effect size *r*	
Size of carrier								
Organisational Professionalism	Large Small Total	135 46 181	97.68 71.39	2203.00	-2.941	0.003*	0.396	Medium
Organisational Professionalism	Medium Large Total	48 135 183	78.81 96.69	2607.00	-2.009	0.440	-	-
Computer Literacy								
Intrinsic Motivation	Poor Average Total	5 87 92	16.20 48.24	66.00	-2.617	0.009*	0.273	Small
Individual Control of Training Outcomes	Poor Average Total	5 87 92	24.50 47.76	107.50	-1.905	0.057	-	-
Intrinsic Motivation	Poor Above Average Total	5 92 97	14.70 50.86	58.50	-2.810	0.005*	0.285	Small
Individual Control of Training Outcomes	Poor Above Average Total	5 92 97	18.00 50.68	75.00	-2.547	0.011*	0.259	Small
Intrinsic Motivation	Poor Excellent Total	5 45 50	8.00 27.44	25.00	-2.848	0.004*	0.403	Medium
Individual Control of Training Outcomes	Poor Excellent Total	5 45 50	8.90 27.34	29.50	-2.706	0.007*	0.383	Medium
Intrinsic Motivation	Average Above Average Total	87 92 179	82.66 96.94	3363.50	-1.849	0.064	-	-
Individual Control of Training Outcomes	Average Above Average Total	87 92 179	79.39 100.03	3079.00	-2.678	0.007*	0.200	Small
Intrinsic Motivation	Average Excellent Total	87 45 132	61.40 76.36	1514.00	-2.137	0.033*	0.186	Small
Individual Control of Training Outcomes	Average Excellent Total	87 45 132	58.36 82.24	1249.00	-3.418	0.001*	0.297	Small
Intrinsic Motivation	Above Average Excellent Total	92 45 137	67.95 71.14	1973.50	-0.445	0.657	-	-
Individual Control of Training Outcomes	Above Average Excellent Total	92 45 137	65.53 76.09	1751.00	-1.473	0.141	-	-

• P < 0.05

5.9.4 Computer literacy

Table 58 shows that there are statistically significant differences ($p \leq 0.05$ to $p < 0.01$) between the pilots who regarded their level of computer literacy as either poor, average, above average or excellent, on most of the latent scales of the research construct. The more prominent differences appeared between participants who regarded their competence in computers as poor to those who felt that their competence was excellent.

Airline pilots who regarded their level of computer literacy as poor were more negative about their training experiences for advanced aircraft, with regard to the microelements of the construct (that is, at an individual level of analysis). In other words, these participants felt less motivated to learn about advanced aircraft ($Z = -2.848$; $p < 0.01$, $r = 0.403$) and felt a sense of a loss in control over the outcomes of such training ($Z = -2.706$; $p < 0.01$, $r = 0.383$).

Furthermore, it appears from the results that when the respondents reported an improvement in their computer literacy skills, their perception of their ability to take charge of their learning for advanced technology aircraft also improved at a statistically significant level. Less significant differences appeared, however, at the upper echelon of the computer literacy band (that is, differences between above average and excellent computer skills). This may be regarded as a ceiling effect.

5.9.1 Digital flight time experience* Level of computer literacy

The aforementioned results indicated that there were significant differences between the effects of the interaction between pilots' experience in advanced technology aircraft combined with their level of computer literacy across most of the demographic categories and the latent scales. This effect warranted further investigation.

According to the results depicted in Table 57 the respondents' perceptions on the latent behavioural scales *Organisational Professionalism*, *Intrinsic Motivation* and

Individual Control of Training Outcomes were significantly ($p = 0.05$ to $p < 0.01$) affected by this interaction, with a small to medium effect size in terms of Cohen's (1988) criterion.

A closer examination of the effect of the interaction between pilots' digital flight experience combined with their level of computer literacy in Table 59 revealed that the impact of subjects' advanced aircraft experience levels on their perceptions of the research construct depended more on their perceived computer literacy levels than on any other variable (such as size of carrier or age). In general it seems that among pilots with a low level of digital flying experience, those who felt that they had a high computer literacy tended to score higher on the three behavioural scales ($p = 0.05$ to $p < 0.01$). The results furthermore suggest that airline pilots with high levels of computer literacy and high digital flight time experience were statistically ($p < 0.05$) more positive about the overall advanced aircraft training climate than the respondents in all the other categories. In other words, it appears that computer literacy is a significant intervening variable on the potential impact that actual digital flight experience has on the training attitude of pilots. It may be posited at this point that low perceived computer literacy is linked to technological averseness, and that higher perceived computer literacy levels may substitute for digital flight experience.

The overall results of the *post hoc* analyses suggest that the combined effect of a pilot's level of computer literacy and experience levels on advanced aircraft have an important role to play in creating an understanding of the phenomena associated with pilots' perceptions of the training climate. It appears that high levels of computer literacy compensate for a lack of experience on the aircraft, in respect of the pilots' confidence and the way respondents feel and experience their training for new technology aircraft. Although the effect size was small in all significant differences of the interaction effect categories, the finding may be an important contribution to the current theory (see Table 59).

Table 59: Non-parametric comparison of the mean rank scores by the level of Digital flight time experience*Computer literacy

Latent perception scale	Level of digital flight time experience* Computer literacy	N	Mean rank score	U	Z	Sig.	Effect size *r*	
Organisational Professionalism (F1)	Low Experience*Low Computer Literacy Low Experience*High Computer Literacy Total	19 66 85	32.05 46.15	419.0	-2.195	0.028	0.238	Small
Intrinsic Motivation (F2)	Low Experience*Low Computer Literacy Low Experience*High Computer Literacy Total	19 66 85	31.29 46.37	404.5	-2.357	0.018	0.256	Small
Individual Control of Training Outcomes (F3)	Low Experience*Low Computer Literacy Low Experience*High Computer Literacy Total	19 66 85	27.18 47.55	326.5	-3.194	0.001	0.346	Medium
Organisational Professionalism (F1)	Low Experience*Low Computer Literacy High Experience*Low Computer Literacy Total	19 73 92	31.13 50.50	401.5	-2.818	0.005	0.294	Small
Intrinsic Motivation (F2)	Low Experience*Low Computer Literacy High Experience*Low Computer Literacy Total	19 73 92	38.00 48.71	532.0	-1.562	0.118	-	-
Individual Control of Training Outcomes (F3)	Low Experience*Low Computer Literacy High Experience*Low Computer Literacy Total	19 73 92	39.32 48.37	557.0	-1.324	0.186	-	-
Organisational Professionalism (F1)	Low Experience*Low Computer Literacy High Experience*High Computer Literacy Total	19 71 90	32.82 48.89	433.5	-2.384	0.017	0.251	Small

Latent perception scale	Level of digital flight time experience* Computer literacy	N	Mean rank score	U	Z	Sig.	Effect size *r*	
Intrinsic Motivation (F2)	Low Experience*Low Computer Literacy High Experience*High Computer Literacy Total	19 71 90	32.97 48.85	436.5	-2.365	0.018	0.249	Small
Individual Control of Training Outcomes (F3)	Low Experience*Low Computer Literacy High Experience*High Computer Literacy Total	19 71 90	33.11 48.82	439.0	-2.342	0.019	0.247	Small
Organisational Professionalism (F1)	Low Experience* High Computer Literacy High Experience* Low Computer Literacy Total	66 73 139	63.39 75.97	1 973.0	-1.840	0.066	-	-
Intrinsic Motivation (F2)	Low Experience* High Computer Literacy High Experience* Low Computer Literacy Total	66 73 139	76.86 63.80	1 956.5	-1.916	0.055	-	-
Individual Control of Training Outcomes (F3)	Low Experience* High Computer Literacy High Experience* Low Computer Literacy Total	66 73 139	82.48 58.72	1 585.5	-3.492	0.000	0.296	Small
Organisational Professionalism (F1)	Low Experience* High Computer Literacy High Experience* High Computer Literacy Total	66 71 137	67.17 70.70	2 222.0	-0.521	0.602	-	-
Intrinsic Motivation (F2)	Low Experience* High Computer Literacy High Experience* High Computer Literacy Total	66 71 137	69.52 68.52	2 309.0	-0.147	0.883	-	-

Latent perception scale	Level of digital flight time experience* Computer literacy	N	Mean rank score	U	Z	Sig.	Effect size *r*	
Individual Control of Training Outcomes (F3)	Low Experience* High Computer Literacy High Experience* High Computer Literacy Total	66 71 137	75.06 63.37	1 943.0	-1.736	0.083	-	
Organisational Professionalism (F1)	High Experience* Low Computer Literacy High Experience* High Computer Literacy Total	73 71 144	77.95 66.89	2 193.5	-1.591	0.112	-	-
Intrinsic Motivation (F2)	High Experience* Low Computer Literacy High Experience* High Computer Literacy Total	73 71 144	66.64 78.53	2 163.5	-1.717	0.086	-	-
Individual Control of Training Outcomes (F3)	High Experience* Low Computer Literacy High Experience* High Computer Literacy Total	73 71 144	65.51 79.68	2 081.5	-2.049	0040	0.171	Small

5.10 BACKWARD STEPWISE LOGISTIC REGRESSION

Logistic regression was used as a predictive analysis to classify subjects with positive perceptions regarding their training for advanced technology aircraft. One of the main advantages of using the logistic method, as opposed to a linear one, is that there are fewer requirements in terms of the assumptions in a logistic regression than for a linear regression (Cohen et al., 2003). In logistic regression, "the predictors do not have to be normally distributed, linearly related, or of equal variance within each group" (Tabachnick & Fidell, 2007:437). The logistic process was also used for its inherent robustness when dealing with dichotomous or ordinal dependent variables (Leech et al., 2005). Ho (2006) asserts that the logistic regression method is ideal for the analysis of perception instruments that yield dichotomous scores.

The model in this particular exploration was used primarily to uncover further important phenomena in the data set. The statistical analyses thus far had revealed some important outcomes; however, uncertainty remained about the relative predictive power of the independent variables. Some authors point out that one of the drawbacks in conducting a stepwise logistic regression is that it may model noise into the equation (Cohen *et al*., 2003; Field, 2005; Goldstein & Wood, 1989). Such noise may result in further uncertainty in conclusions about the data set. Hence, it was important to discover the predictive power of variables from a regression analysis in order to add to the theoretical knowledge on the topic of interest.

Logistic regression was deemed suitable for this final examination of the research construct, because the dependent variable to be studied was a dichotomy and the independent variables were of varying types. The exploratory method involved an analysis, which began with a full or saturated model; variables were then eliminated from the model in an iterative process.

The aim in this stage of the study was to develop a model that could, to some extent, predict whether pilots would perceive the technologically advanced aircraft training climate as favourable or unfavourable, based hitherto on what is currently known about the latent structure of the dataset. Dummy variables of "1" or "0" were allocated to cases where the average perception on the three-factor model exceeded 5.0

(because of the right skewness in the final dataset) or not, as the case may be. This provided a dichotomous measure, *Favourability,* of a categorical outcome that indicated level of airline pilots' comfort in the advanced aircraft training climate in terms of their overall perception. A backward stepwise logistic method was used to determine the contribution each variable would make to the regression equation because "sequential logistic regression should be part of a cross-validation strategy to investigate the extent to which sample results may be more broadly generalized" (Tabachnick & Fidell, 2007:456). The equation for this exploratory method starts with all the selected independent variables first entered and then deleted after evaluation (Field, 2005; Ho, 2006). The following nine categorical or ordinal variables were included as predictors in the original model:

- The interaction effect between a pilot's level of flight experience in advanced aircraft with that of his or her level of computer literacy, where

  1 = Low Experience*Low Computer Literacy;

  2 = Low Experience*High Computer Literacy;

  3 = High Experience*Low Computer Literacy; and

  4 = High Experience*High Computer Literacy.

- A bivariate age grouping categorised in terms of respondents who indicated that they were either 40 years of age or younger and those who were over 40 years old, where

  0 = 40 and younger; and

  1 = 41 and older.

- Practical flight experience in advanced aircraft, where pilots with over 2 000 hours flight time on these aircraft were considered as having high experience, and where

  1 = Low Experience; and

  2 = High Experience.

- Route training, which indicated how much a pilot enjoyed training on the actual aircraft, where

  1 = Never Enjoy;

2 = Sometimes Enjoy; and

3 = Always Enjoy.

- Simulator training, which indicated how much the pilot enjoyed training experience in the flight simulator, where

  1 = Never Enjoy

  2 = Sometimes Enjoy; and

  3 = Always Enjoy.

- A variable which indicated the size of the carrier for which the respondent was employed, and where the size of the enterprise is represented as

  1 = Large;

  2 = Medium; and

  3 = Small.

- Pilot unionisation, which refers to the degree to which the pilot group was unified in terms of belonging to the ALPA-SA – the organisation was either unionised, or not unionised, where

  1 = Unionised; and

  2 = Not-unionised.

- The level of computer literacy, which indicated the extent to which the candidate felt that his or her computer skills were either poor or good, where

  1 = Low Literacy; and

  2 = High Literacy.

- A company status/position category, which divided the sample into two specific groupings, either captain or co-pilot, where

  0 = Co-pilot; and

  1 = Captain.

The backward stepwise regression analysis was completed after five steps. A model containing four predictors subsequently emerged. In addition, five of the aforementioned predictor variables were removed iteratively, namely pilot unionisation, age group, the size of carrier, the company status/position of the pilot and the pilot's actual perceived level of computer literacy. The four variables that successfully predicted a subject's perception of the *Favourability* associated with the advanced aircraft training climate were:

i) the interaction effect between a pilot's level of flight experience in advanced aircraft and his or her level of computer literacy:
ii) practical flight experience in advanced aircraft;
iii) preference regarding training in the flight simulator; and
iv) preference regarding route training in the actual aircraft.

Table 60 shows that the overall percentage of cases for which the dependent variable was correctly predicted by the model was 63.8%. The results also show that the model was correct in predicting perceptions of a *favourable* climate 100% of the time (that is, high positive predictive validity). However, the model could not successfully predict respondents' perceptions of an unfavourable training climate.

The effect size and practical significance of a regression analysis is generally provided for by examining the odds ratio and, less commonly, by computing the differences between Nagelkerke's pseudo R-Square values ($R^2\Delta$) (Cohen *et al.*, 2003). A cut-off value of $R^2\Delta = 0.02$ was used, as suggested by Schaap (2011), because the pseudo R^2 is not technically a goodness-of-fit index, and cannot explain the proportion of the variance. Therefore, the result based on computation of $R^2\Delta$ for this study was used with caution in interpreting practical significance

Nagelkerke's $R^2\Delta$ was computed by first calculating the value of the pseudo R^2 at the initial step and thereafter finding the difference at each subsequent step. The effect size of the model at each subsequent step was less than 0.02 and could therefore be regarded as not practically significant in terms of this criterion. Nonetheless, overall, the final model was regarded as efficient (see Table 60).

Table 60: Classification table and model summary

Classification		Predicted		
Observed		Favourability of training climate		
		Climate unfavourable	Climate favourable	Percentage correct
Favourable training climate	Climate unfavourable	0	83	(none correctly identified) 0.0
	Climate favourable	0	146	100.0
	Overall percentage			63.8
Model summary				
Step	-2 Log likelihood	Cox & Snell R^2	Nagelkerke's R^2	Nagelkerke's $R^2\Delta$
1	266.955	0.134	0.184	-
2	267.024	0.134	0.183	0.001
3	267.534	0.132	0.181	0.002
4	268.678	0.127	0.175	0.006
5	270.021	0.122	0.168	0.007
Step	Hosmer and Lemeshow test chi-square	Df	Sig. (2-tailed)	
1	13.884	8	0.085	
2	8.399	8	0.396	
3	13.720	8	0.089	
4	11.272	8	0.187	
5	2.365	7	0.937	

The efficiency of the resulting model is endorsed by the non-significance in the result of the Hosmer and Lemeshow test chi-square statistic in the final step (χ^2 [7, N=229] = 2.365, p = 0.937). These results confirmed that the final variables in the model predicted the observed data relatively effectively. According to Field (2005), the proportion of variance in the outcome variable associated with each of the predictor variables can be given by R^2. However, because the dependent variable in this regression model was dichotomous or categorical in nature, only approximations of R^2 were possible in SPSS Version 17.0; hence the choice of Nagelkerke's pseudo R^2 to gauge effect. Additionally, both the Cox and Snell $R^{2,}$ together with the Nagelkerke

R^2 values, were used to determine that the final model could account for approximately 12% to 17% of the variability in the criterion variable.

Table 60 shows that moderate changes occurred in the -2 log-likelihood values between the constant only model, and the first and last step, which is a good indication that the final model had improved predictive power. A nominal regression of the final four predictor variables in the model then produced a comparison in which the -2 log-likelihood values of the intercept only (95.338) and final model (65.456) indicated that the change in the amount of predictive power provided in the final solution was statistically significant [χ^2 (4) = 29.883, $p < 0.0001$].

Moreover, McFadden's p^2 [1-log likelihood (final)/log likelihood (constant)] = 0.313 was computed as an indication of a measure of the strength of association between the predictor variables and the model. According to Tabachnick and Fidell (2007:460), McFadden's p^2 is expected to be "lower" than the traditional R^2 as a measure of effect size, and values between 0.20 and 0.40 are considered highly satisfactory. In terms of McFadden's p^2 (as opposed to Nagelkerke's $R^2\Delta$), the analysis suggests that the size of the final logistic model is large and of practical importance.

Table 61: Final logistic regression model

Predictors in the equation (X_j)	B	S.E.	Wald Chi-square ($B^2/S.E.^2$)	Df	Sig.	Odds Ratio (e^B)	95% C.I. for odds ratio	
							Lower	Upper
Interaction effect	0.630	0.310	4.126	1	0.042	1.878	1.022	3.448
Advanced aircraft experience	-1.064	0.613	3.011	1	0.083	0.345	0.104	1.148
Enjoy route training	0.485	0.289	2.814	1	0.093	1.624	0.922	2.861
Enjoy simulator training	0.806	0.267	9.138	1	0.003	2.238	1.327	3.773
Constant	-2.603	0.912	8.142	1	0.004	0.074		

In the final model, the dependent variable, *Climate Favourability*, is on the logit scale. The computation results in Table 61 show that the probability that a respondent would perceive the advanced aircraft training climate as favourable can be given by the following three logistic regression equations (Tabachnick & Fidell, 2007:438):

- Logit = ln (p/1-p) = -2.603 + 0.63 * (interaction effect) – 1.064 * (advanced aircraft experience) + 0.485 * (route training) + 0.806 * (simulator training) … *[Equation 1]*
- Prob (Favourable perception) = $(e^{-2.603 + 0.63 X_1 - 1.064 X_2 + 0.485 X_3 + 0.806 X_4})/(1 + e^{-2.603 + 0.63 X_1 - 1.064 X_2 + 0.485 X_3 + 0.806 X_4})$ … *[Equation 2]*
- Log odds ratio = p/(1-p) … *[Equation 3]*

The model produced by the logistic regression is non-linear. In the current case, Table 61 shows that experience in an advanced aircraft and preference for route training did not improve the predictive efficiency of the model. The Wald chi-square statistics were non-significant for these two variables ($p > 0.05$), whereas the chi-square value for the interaction effect of advanced aircraft experience and computer literacy as well as for enjoying simulator training was significant at a 0.05 level. Nonetheless, Hosmer and Lemeshow (2000), also cited in Tabachnick and Fidell (2007:456), recommend that for “a criterion for inclusion of a variable that is less stringent than 0.05 …something in the range of 0.15 or 0.20 is more appropriate to ensure entry of variables with coefficients different from zero”.

Thus, given that the other predictors remain in the model, it would appear that removing the interaction effect or information about preference for simulator training would result in significantly poorer predictive efficiency of the model (because these variables have greater significance). Overall, it can therefore be concluded that the predictive efficiency of a four-predictor model is not noticeably greater than that of a two-predictor model, which would then include only the variables *Interaction Effect* and *Preference for Simulator Training*.

The estimated logistic regression coefficient for the interaction effect between flight experience in advanced aircraft * level of computer literacy was 0.63 and the exponential of this value is $e^{0.63}=1.878$. This indicates that for a one-unit increase in the interaction effect, the odds in favour of a trainee's perceiving the training climate as favourable are estimated to increase by a multiplicative factor of 1.878. In addition, (p/1-p) = 1.878 implies that knowledge of the interaction effect can improve the probability of correctly predicting that a pilot will perceive the advanced aircraft training climate as either favourable or unfavourable by 65%.

The estimated logistic regression coefficient for pilots' preference for simulator training was 0.806 and the exponential for this value is $e^{0.806}=2.238$. This indicates that for a one-unit increase in a pilot's positive preference for simulator training, the odds in favour of the trainee's perceiving the training climate as favourable are estimated to increase by a multiplicative factor of 2.238. Furthermore, (p/1-p)=2.238 implies that having knowledge of a pilot's positive preference for, or an enjoyment of simulator training, can improve the probability of correctly predicting whether he or she will perceive the advanced aircraft training climate as favourable by 69% (that is, p = 0.69).

To test the predictive power of the model, Equation 2 was then applied to the following extreme scenarios:

- The computation of the probability of a trainee's perceiving the climate as favourable given that the pilot has a low advanced aircraft experience*low computer literacy combination, low advanced aircraft flight experience, and reports never enjoying route or simulator training:

$$\text{Prob(Fav)}=(e^{-2.603 + 0.63 (1) - 1.064 (1) + 0.485(1)+ 0.806 (1)})/(1+ e^{-2.603 + 0.63 (1) - 1.064 (1) + 0.485(1)+ 0.806 (1)})$$

$$=e^{-1.746}/1+ e^{-1.746}$$

$$=0.174/1.174 = 0.148$$

  Therefore, the model predicts a very low probability (14.8%) that the candidate, given the aforementioned criterion, will perceive the advanced aircraft training climate as favourable.

- The computation of the probability of a favourable climate perception by a trainee who has a high advanced aircraft experience*high computer literacy combination, and high advanced aircraft flight experience, and who always enjoys route or simulator training:

$$\text{Prob(Fav)}=(e^{-2.603 + 0.63 (4) - 1.064 (2) + 0.485(3)+ 0.806 (3)})/(1+ e^{-2.603 + 0.63 (4) - 1.064 (2) + 0.485(3)+ 0.806 (3)})$$
$$=e^{1.662}/1+ e^{1.662}$$
$$=5.270/6.270 = 0.840$$

Therefore, the model predicts a very high probability (84%) that the candidate, given the aforementioned criterion, will perceive an advanced aircraft training climate as favourable.

The illustration of the usefulness of the logistic model as a predictor of climate favourability is demonstrated in the above example. Predictive efficiency thus appears to be very good for the two extreme cases.

The probability curves based on the derived probability formula are depicted in Figures 28 and 29. The plots were computed for the two most important independent variables found by the regression model:

- the *Interaction Effect*, where the variable is based on a combination of an airline pilot's number of hours experience in advanced aircraft and perceived level of computer literacy; and
- a preference *for Simulator Training*, a variable that refers to the airline pilot trainee's enjoyment of flight simulator training.

Figure 28: Probability plot for the interaction effect between experience and computer literacy

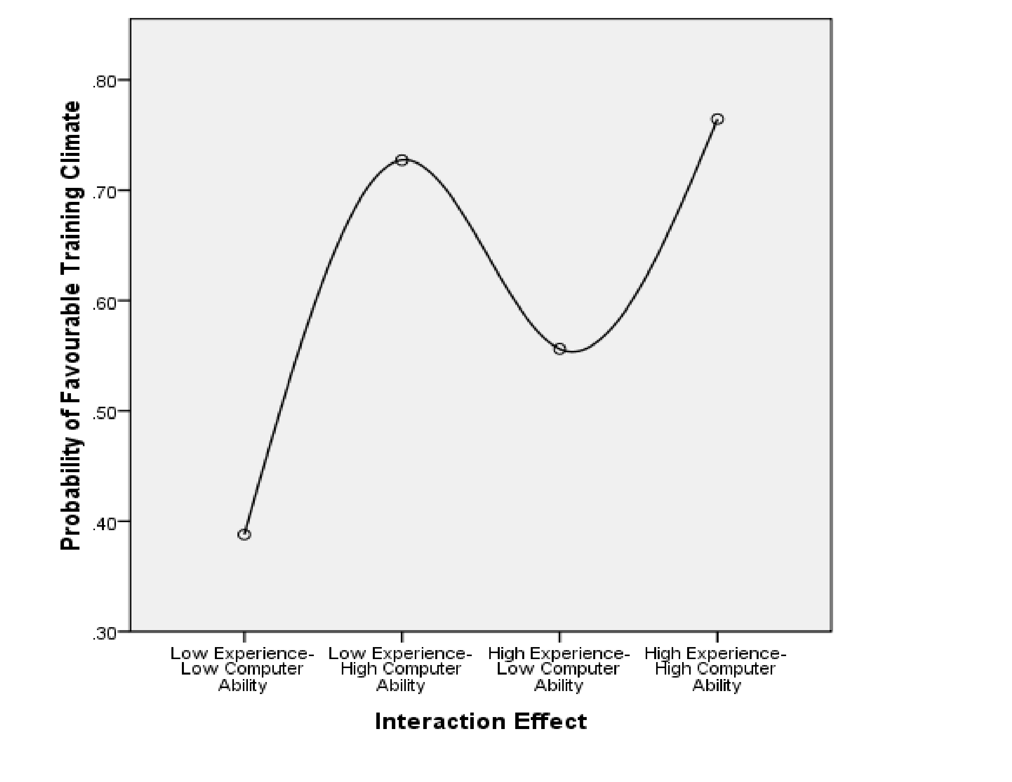


Figure 28 clearly shows the importance of a pilot's perceived computer literacy in the interaction effect with the candidate's advanced aircraft flight experience. The noticeable impact of this interaction on the probability curve is evident in the centre of the plot. The probability that a trainee will perceive the advanced aircraft training climate as favourable diminishes significantly when his or her levels of computer literacy drop, even when the person's levels of flight experience is higher (that is, the high probability of a favourable climate decreases between low experience*high computer literacy and high experience*low computer literacy).

The second important predictor in the regression model is plotted in Figure 29.

Figure 29: Probability plot for Preference for Simulator Training

The graph (Figure 29) depicts an almost linear relationship between a pilot trainee's preference or enjoyment of flight simulator training and the probability that the subject will perceive the training climate as either favourable or unfavourable. It appears from Figure 29 that understanding the levels of enjoyment a subject may experience with regard to simulator training has a significant impact on predictive power.

The confidence intervals of the odds ratios for all the remaining singular predictors for the final model were consistently positive. According to Field (2009), findings based on the results of reporting positive confidence interval values present a possibility to generalise a regression model to the broader population. However, the small values of Nagelkerke's $R^2\Delta$ in the current analysis imply that the practical significance of the model should be considered in the context of the shortcomings of the study. Moreover, the relatively high standard errors (SE) of each predictor variable point to the possibility that the model may be unstable due to interference or noise (Tabachnick & Fidell, 2007).

The high homogeneity of the sample frame in terms of their education, experience and training profoundly skewed the data that dominate in the final model, creating interference in the predictors. Nonetheless, the results and conclusions drawn from the regression analysis provide pertinent information for assessing the suitability of candidates for advanced aircraft training. This study supports the conclusion that knowledge of both the interaction effect of pilots' experience in advanced aircraft combined with a pilot's computer literacy or competence, and the knowledge of the levels of enjoyment a pilot derives from simulator training, are highly useful predictors of a trainee's perception of the advanced aircraft training climate. Appropriate interventions based on the understanding of this knowledge may in turn affect the level of success of the education methods and the interventions that are chosen, and thus the eventual success of training outcomes.

5.11 SUMMARY

The focus in this chapter was on interpreting and reporting the results of the phenomena associated with the latent factors of the theoretical construct. The results of a thorough exploration based on principal axis factoring and oblique rotation produced more factors than had been anticipated. Horn's parallel method, based on the algorithm provided by O'Connor, resulted in a three-factor final solution. It also eliminated the Heywood anomaly in the data. The latent behavioural scales were labelled *Organisational Professionalism*, *Intrinsic Motivation* and *Individual Control of Training Outcomes*. The first factor is related to both the macro (organisational) and intermediate (instructor-trainee) levels of analysis. The remaining two latent factors both entail variables related to the micro or individual level of analysis, suggesting that the trainee, in this case, is a focal point in the phenomena associated with training for operating advanced aircraft. The various analyses conducted suggest that the measurement construct under investigation has a relatively stable factor solution.

A reliability analysis of the final item cohort produced highly satisfactory Cronbach's coefficient alphas of between 0.70 and 0.95. It was also shown that the scale demonstrated very good discriminatory properties, given its ability to distinguish effectively between high and low scorers, suggesting excellent scale reliability and good overall development.

Exploration of the data set revealed that the scores were non-normal; therefore, subsequent analyses were based on using the appropriate family of non-parametric statistics. To establish a foundation for a further exploration of phenomena present in the data, a basic overall non-parametric comparative assessment was conducted, using the Kruskall-Wallis test statistic and supplemented for specific group comparisons by the Mann-Whitney test statistic.

For associational evaluation of the data, the non-parametric equivalent of Pearson's R, Kendall's tau-b, was deemed most appropriate for the exploration of correlations. The robustness of Kendall's tau-b assisted in offsetting the inherent problems associated with the non-normality and subsequent violations of assumptions in the current data set. The associational computation of conceptually significant variables allowed for a further statistically complex examination of the underlying phenomena present in the data. The results with regard to statistically significant correlations guided the researcher in the appropriate direction.

Subsequently, a multivariate analysis of variance was conducted on the ranked order of the scores on digital flight deck experience, a pilot's company status, the size of the carrier, pilots' levels of computer literacy and an interaction effect between experience in advanced aircraft and computer literacy. The results were then scrutinised and thoroughly discussed in a *post hoc* analysis.

In order to determine the predictability of the independent demographic data and to develop a structured explanatory model, a stepwise logistic regression analysis was carried out. The results produced a highly satisfactory model with four predictor variables. It was found that the derived logistic equation could satisfactorily compute, with reasonable probability (0.148 at the low end to 0.840 at the high end), whether a subject would hold an unfavourable or a favourable perception of the training climate.

Overall, the results indicate that, although the nature of the data was differentially skewed, the use of appropriate (more robust) statistics provided satisfactory answers and revealed distinct and important phenomena, thus making a significant contribution to the current theory base on the topic. Furthermore, the methodology followed exposed both a relatively stable underlying factorial structure and a prediction model that will be useful to airline organisations engaged in training advanced aircraft pilots.

One of the main findings from the data analyses is that the interaction between a pilot’s level of advanced aircraft experience in combination with his or her computer literacy, competence and/or abilities, has a significant impact on the pilot’s perception of the advanced aircraft training climate. In addition, it was found that the use of flight simulators plays an extremely important role in training perceptions. Further study linking synthetic flight training (simulation) to real-life aircraft training and operational behaviour is recommended in order to enhance the current theoretical knowledge base.

Against the background of the aforementioned research findings, a discussion of the results and recommendations follow in the final chapter.