Chapter Five: Methodology

The following methodology chapter will delineate the quantitative method used to address the aims of the present study. It will begin with a description of the sample and sampling strategy, followed by a step-by-step outline of the data gathering process and the measures that were used. Ethical considerations will then be presented, followed by a discussion of the statistical procedures employed.

5.1 Sample

The sample consisted of 74 male undergraduate students from the School of Actuarial Science and Statistics at the University of the Witwatersrand (Wits). As the sampling strategy was non-probability, convenience sampling, it was not possible to estimate the probability of each person having an equal chance of being included in the sample (Niele & Liebert, 1986). Non-probability sampling also places a limit on the generalisability of the findings of the present research with respect to both population validity and ecological validity (Rosnow & Rosenthal, 1996). This is because samples of undergraduate students are likely to have limited diversity regarding variables such as age, culture, marital status and career (Prinsloo, 1992). Moreover, while ethically necessary, the use of volunteers in the present study may also limit the generalisability of the findings due to the possibility of volunteer bias (Rosnow & Rosenthal, 1996).

While the present research was cognisant of the concerns raised previously, the choice of sample was advantageous for various reasons, including cost, time and convenience. It was also hoped that the relatively high number of men in the School of Actuarial Science and Statistics would have a positive effect on the sample size, and hence the statistical measures that could be used. The sample also had additional benefits to the present study in view of the theoretical framework upon which it was based. For example, as a particular second year class was targeted, it was likely that participants would be of similar age, and hence in similar stages of psychological development. Given that the typical age of the students in the present sample was 19-20 years, it was likely that the majority of participants would be nearing the end of adolescence and entering young adulthood. In Erikson's theory outlined in the literature review, this implied that the participants would have had to negotiate the

identity crisis and have come to at least a basic resolution. This has strong implications for the integration of the ego, as a clearer conceptualisation of one's identity has been positively associated with greater ego strength (Markstrom et al., 1997). While this negotiation is likely to imply that a firmer gender identity has been established, it should be noted that the individual may still be experimenting with their sexual orientation (Tyson & Tyson, 1990). Moreover, a large proportion of the data on the Adonis Complex, which feeds into the body image variable in the present research, has also been generated from data from American college students in similar age groups to the present sample (c.f, Liet, Gray and Pope, 2001; Pope et al. 2000). The sample may consequently be a valuable yardstick for corroborating the findings from such research. It should, however, be noted that while many individuals may be dissatisfied with their bodies, disturbance is far less common, and is typically found in more clinical populations, or in persons who are acutely interested in their body shape, such as weightlifters and body builders. As such, while disturbance was included in the analysis, the present sample was likely more appropriate to assess body image dissatisfaction rather than disturbance.

5.1.1 Descriptive information of the present sample

The following information was derived from the demographic section of the questionnaire used in the research. This section inquired into the respondents' age, year of study, height, weight, physical activities enjoyed and sexual orientation. Frequency distributions demonstrated that the participants' ages ranged from 18 to 22, with a mean age of 19.7 (s.d = 0.1). Thirty-eight students (51%) were 19 years of age, nineteen students (26%) were 20 years of age, and a further twelve students (16%) were 21 years of age. Two of the remaining five respondents were 18 years of age, while three were 22 years of age. In terms of year of study, 65 students (88%) were in their second year, while the remaining 9 students were in their third year (12%). With respect to sexual orientation, 65 of the students maintained that they were heterosexual (88%), seven stated they were homosexual (9%), while the remaining three students claimed they were bisexual (3%). This represents a surprisingly high proportion of non-heterosexual students in the sample, with a combined bisexual and homosexual proportion of 12%.

The following table presents the proportion of sexual orientation relative to each dimension of gender identity as measured by the BSRI. Details of the way in which these classifications are determined is discussed in section 5.3.3 of the present chapter.

Table 5.1

Proportion of sexual orientation relative to each dimension of gender identity

	Proportion of each sexual orientation		**Total**
Gender identity	**Heterosexual men** (*65 participants)*	**Gay/ Bisexual men** (*9 participants)*	**N**
Sex-typed *(men who are masculine)*	9 (12%)	2 (3%)	11
Cross-typed *(men who are feminine)*	12 (16%)	3 (4%)	15
Androgynous	21 (28%)	2 (3%)	23
Undifferentiated	23 (31%)	2 (3%)	25

While this was not of primary concern, the results showed that 48 of the 74 participants (65%) were either undifferentiated (34%) or androgynous (31%). This is a high proportion of individuals who appear to take on integrated sets of traits as compared to the 35% who were either sex typed (11 participants) or cross-typed (15 participants). This appears to attest to a growing sense of more integrative gender identities being adopted.

The category of physical activities enjoyed was open-ended, and was used to gauge whether the students enjoyed physical activities, and if so, which were the most common. The results showed that while 7 respondents did not partake in any physical activities, the remaining 67 were involved in 25 different physical activities. In view of the diverse range of activities found, only the most recurring activities will be discussed. It should be noted that this was a multiple response item and so respondents could score in more than one sporting category. The most common

activities were going to gym (28%), running (24%), and squash (22%). A small proportion partook in more aggressive sports such as karate (10%) and rugby (1%), while slightly more enjoyed cricket (15%) and soccer (19% respondents).

Finally, the height and weight variables were used to generate Body Mass Index (BMI) scores, which were later incorporated in the correlation analyses.
The BMI is calculated using the following formula:

$$\frac{\text{Weight (in kilograms)}}{\text{Height (in metres)} \cdot \text{Height (in metres)}}$$

Table 5.2
Results of the range and frequency of BMI scores

Range of BMI scores	Frequency	Cumulative percentage
< 20	10	14%
20 – 25	46	78%
25 – 30	15	100%

The results suggested that while the majority of the respondents (64%) were within the acceptable BMI ranges, 14% were underweight and 22% were theoretically overweight. The BMI scores ranged from 18.5 to 29.4, with a mean and median BMI score of 23. The mode was however slightly lower at 20.

5.2 Data Gathering Procedure

The data was collected by means of three self-report paper-pencil measures administered at one sitting to a group of University students. The procedure followed for collecting the data is described below.

5.2.1 The pilot study

Before the questionnaire was administered, a pilot study was conducted on nine men between the ages of 23-25. While this age range was slightly higher than the students that were later targeted, all nine men had completed their own degrees within the past year. The pilot was used as an opportunity to see whether items were understandable, whether there were any spelling errors, if the instructions were clear, and if any

changes should be made. The participants were asked to complete the questionnaire and report how they felt regarding the above considerations. They could choose to either write their comments on the questionnaire itself, or give verbal feedback which the researcher took note of. It was also noted from the pilot that the questionnaire took an average of 12 – 15 minutes to be filled out. The pilot was particularly useful for feedback on the Adonis Complex Questionnaire-Revised (ACQ-R) as it had no previous reliability or validity information. These items were not reported as invasive or otherwise problematic, though one respondent asked if the questionnaire was specifically targeted at students, as one of the items (ACQ-R5) asks about body image concerns interfering with studies. While a missing word was picked up, no other changes were required.

5.2.2 Data collection

In order to conduct the study, written permission was required from the Head of School of Actuarial Science and Statistics. Once this was received, ethical clearance was obtained. The researcher then contacted the relevant lecturer to set up a meeting to discuss the nature of the research and the possibility of using lecture time to conduct the study. The researcher explained that a study in partial fulfilment of a Master of Arts degree was being conducted on men only, and that participation would entail completing a questionnaire which would take no more than 20 minutes. Having agreed, the lecturer suggested that a particular second year class be used in view of the number of males registered for the course. The use of this particular class was also more convenient for the lecturer himself, and so a mutually convenient date and time was confirmed. The data was collected from this one sitting.

On the day of the data collection, the researcher entered the lecture hall 25 minutes before it was scheduled to end, as agreed with the lecturer. The lecturer then ended the class, and requested that the female students leave the lecture venue. It was deemed more appropriate to address the men separately so as to avoid too much disruption. The researcher then addressed the male students and invited them to participate in a study in partial fulfilment of a Master of Arts degree. Students that were willing to participate were then asked to stagger themselves so that there was at least one free seat between them and the next person to ensure confidentiality. They were then given an envelope which contained the questionnaire, and were asked to

seal their questionnaire in the envelope provided (whether completed or not) prior to placing it in the box at the front of the lecture hall as they exited. As participants were also given the opportunity to take the questionnaire home and return it later, the information sheet provided details of a collection point.

After the data collection, the researcher proceeded to enter the responses into the computer programme and conducted the appropriate statistical analysis. The descriptive statistics were analysed first, followed by a reliability analysis of each scale. Thereafter, tests of the assumptions for the desired correlation and multiple regression techniques were carried out, followed by the actual analyses.

5.3 Measures

The following discussion presents an overview of the different measures used, and where appropriate, it includes information on the psychometric properties of the tests, such as their reliability and validity. "Reliability has been defined as the extent to which [a] test is effectively measuring anything at all, and validity as the extent to which the test is measuring what it is purported [or claimed] to measure" (Rust & Golomok, 1989, p. 89). Reliability information is consequently important as it provides evidence of the degree to which the test assigns scores in a consistent fashion (Murphy & Davidshofer, 2001), while validity concerns "what the test measures and how well it does so" (Anastasi & Urbina, 1997, p. 113). While dependant on what the instrument is used for, reliability estimates of .80 or more are regarded as moderate to high reliability (Murphy & Davidshofer, 2001). Assessment of validity involves, for example, examining the test content to determine its appropriateness as well as finding correlations between related and unrelated constructs (Murphy & Davidshofer, 2001). Given that two of the scales were modified for use in the present research, the rationale and details of the procedure to effect these changes are explicitly outlined later in this chapter.

The format of the questionnaire (in order of appearance) was an information sheet inviting participation in the study (see Appendix A), a demographic section (see Appendix B), the Adonis Complex Questionnaire-Revised (ACQ-R) (see Appendix C), the Bem Sex Role Inventory (BSRI) (see Appendix D), and the Psychosocial Inventory of Ego Strength (PIES) (see Appendix E). A final demographic question on

sexual orientation was placed at the end of the questionnaire at the recommendation of the Human Research Ethics Committee (Non-Medical). A short set of instructions appeared prior to each measure.

5.3.1 Demographic questionnaire

This was used to obtain a descriptive profile of the participants. It asked the participants' age, year of study, height, weight, physical activities enjoyed and sexual orientation. The findings were discussed in section 5.1.1.

5.3.2 The Adonis Complex Questionnaire-Revised (ACQ-R)

As stated in the literature review, Pope et al. (2000), developed different measures to assess traits of the Adonis complex and muscle dysmorphia respectively. While the Adonis Complex Questionnaire (ACQ) was believed to measure body image concerns and the extent to which they affect the individual's daily life, the Muscle Dysmorphia Questionnaire (MDQ) was suggested to look at the extent to which the person is obsessively preoccupied with the muscularity of their body (Pope et al., 2000). The ACQ is a 13 item scale where respondents choose between one of three available options ranging from "rarely or not at all" to "frequently," depending on their level of agreement. The total score ranges from 0 to 39, where higher scores indicate more problems associated with the Adonis Complex (Pope et al., 2000). The MDQ is a 15 item measure where respondents answer either "yes" or "no" to the questions. Men with at least some degree of muscle dysmorphia will typically answer "yes" to at least four or five of the questions (Pope et al., 2000). To the present researcher's knowledge, neither scale has had reliability or validity information reported.

While it was originally thought best to administer both the ACQ and the MDQ to fully engage with male body image concerns, closer inspection showed that there was substantial overlap between the two questionnaires. To avoid redundancy, the questionnaires were combined to form one measure. Items were chosen on the basis of their ability to provide suitable information on both general aspects of the Adonis complex and more specific muscularity concerns. Nine items were taken from the original ACQ while the remaining eight were taken from the MDQ. Eleven items were not used as they were repetitions of a question already included. While two items were rephrased to avoid poor wording and ambiguity, the original meaning of

the items was not changed. For example, item 2 of the original ACQ was phrased "how often are you distressed by your appearance concerns (that is, feeling upset, anxious, or depressed)?" In the ACQ-R, the item is phrased as "how often are you anxious or upset regarding your general physical appearance?"

The combined version of the measures was called the Adonis Complex Questionnaire-Revised (ACQ-R) and contained 17 items. The ACQ-R was administered in the pilot study and was found to be unproblematic. The scoring for the ACQ-R is in line with the original ACQ as respondents need to choose from one of three options ranging from "rarely or not at all" to "frequently." While the same principle is used, a different set of options follows ACQ-R1, ACQ-R2 and ACQ-R11 in view of the specific information asked. Higher scores on the ACQ-R indicate greater body image dissatisfaction and more preoccupation with muscularity.

As the original ACQ and MDQ had no previous psychometric information, it was deemed necessary to run a series of statistical procedures to ensure the ACQ-R was a reliable measure of body image concerns. Initial reliability of the 17 item ACQ-R showed a raw Cronbach Coefficient alpha of .84, with the standardised coefficient alpha of slightly less at .82. Both correlations represent acceptable reliability. However, given that there were adjustments to the scale, a factor analysis was used to further assess its properties. This factor analysis used a principal component analysis to extract the initial factor solution, which was then rotated using a varimax rotation. Principal component analysis is defined as "a technique that makes use of the covariance or correlation matrix of a large set of observed (metric) variables to obtain a smaller set of new variables that have the desired properties" (Pietersen & Damianov, 1998, p. N-1).

In order to conduct a sensible principal component analysis, it is necessary to have significant correlations amongst the variables (Kline, 1994). As such, a correlation matrix of the observed variables was generated, and suggested significant correlations amongst the majority of the items, ranging from the.40 to .84. However, ACQ-R10 correlated poorly with all the other items, and was later removed. Kaiser's measure of sampling adequacy (MSA) was also calculated to assess the degree of correlations amongst variables. The overall MSA of 0.46 was concerning as it stood below the

recommended 0.5, but it was nevertheless deemed appropriate to continue with the procedure given the strong inter-item correlations found previously.

Three different guidelines were used to determine the number of factors to extract. Firstly, an eigenvalue correlation matrix was calculated. The eigenvalue is essentially a representation of the proportion of variance explained by each factor, where "the larger the eigenvalue the more variance is explained by the factor" (Kline, 1994, p. 30). A common rule of thumb is that the number of factors that should be extracted corresponds to the number of eigenvalues greater than one. With regard to the table reproduced below, this would suggest that a six-factor solution would be appropriate to the present research, but this would likely overestimate the number of relevant factors (Kline, 1994). An alternative guideline is to assess the *proportion of variance explained* inferred from the fourth column of the table below. This guideline suggested a two-factor solution, as factor one could explain 30.4% of the original variance, while factor two accounted for a further 13.8%

Table 5.3

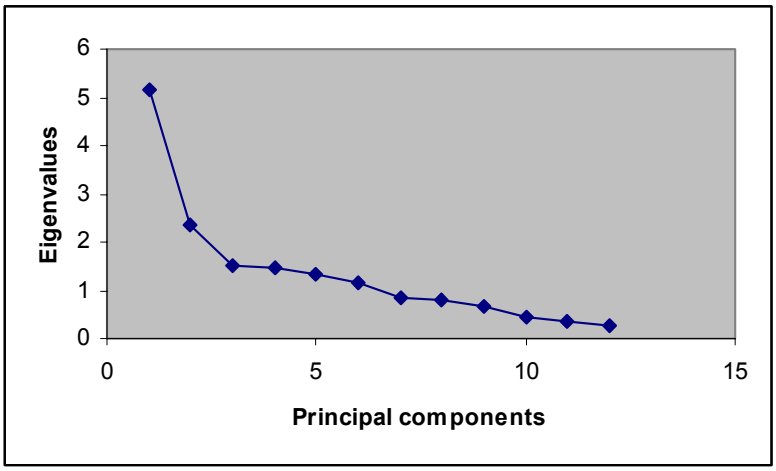
Eigenvalues of the correlation matrix for the ACQ-R

	Eigenvalue	Difference	Proportion	Cumulative
1	5.17435201	2.83623043	0.3044	0.3044
2	2.33812158	0.81439794	0.1375	0.4419
3	1.52372364	0.03489884	0.0896	0.5315
4	1.48882480	0.16717525	0.0876	0.6191
5	1.32164955	0.18407220	0.0777	0.6969
6	1.13757735	0.27951975	0.0669	0.7638
7	0.85805760	0.07136399	0.0505	0.8143
8	0.78669361	0.12278741	0.0463	0.8605
9	0.66390620	0.20502397	0.0391	0.8996
10	0.45888224	0.11305931	0.0270	0.9266
11	0.34582293	0.06197215	0.0203	0.9469
12	0.28385078	0.04279828	0.0167	0.9636
13	0.24105250	0.09545178	0.0142	0.9778
14	0.14560072	0.03589669	0.0086	0.9864
15	0.10970404	0.01766270	0.0065	0.9928
16	0.09204134	0.06190224	0.0054	0.9982
17	0.03013910		0.0018	1.0000

The third guideline is a scree plot of the eigenvalues and principal components, which is reproduced in Figure 5.1 below. According to Kline (1994), this is one of the best

solutions to selecting the number of factors to extract. In the scree test, a graph is made of the principal components and eigenvalues, where "the cutoff point for factor rotation is where the line changes slope" (Kline, 1994, p. 75). The scree plot also suggested a two or three factor solution. While some discrepancy can be seen in the findings of the three guidelines, it was deemed appropriate to use a two-factor solution on balance.

Figure 5.1
Scree plot of eigenvalues and principal components for the ACQ-R



Before the factors could be interpreted, a factor rotation was performed using varimax. "Rotating factors changes the factor loadings and thus the meanings of the factors, but the different factor analytic solutions are mathematically equivalent in that they explain the same amount of variance in each variable and thus in the matrix as a whole" (Kline, 1994, p. 62). The varimax rotation is orthogonal, which ensures that the rotated factors remain uncorrelated, and aims to "maximise the sum of *variances* of squared loadings in the columns of the factor matrix. This produces in each column (which is, of course, a factor) loadings which are either high or near zero" (Kline, 1994, p. 68, original emphasis). As such each ACQ-R items should load significantly on only one factor. Factor loadings may be regarded as correlations and interpreted accordingly. The results are presented in Table 5.4.

Table 5.4

Rotated factor pattern for the ACQ-R

	Factor one	Factor two
ACQ-R1	**0.45019**	**0.53720**
ACQ-R2	**0.53014**	**0.42580**
ACQ-R3	**0.75713**	0.35709
ACQ-R4	**0.84009**	0.30428
ACQ-R5	**0.52604**	0.20210
ACQ-R6	**0.60658**	0.01305
ACQ-R7	0.10939	**0.77586**
ACQ-R8	0.11317	**0.68565**
ACQ-R9	0.21467	**0.83029**
ACQ-R10	0.13261	-0.15606
ACQ-R11	0.09810	**0.56875**
ACQ-R12	**0.41382**	0.14801
ACQ-R13	**0.73641**	-0.08011
ACQ-R14	-0.11044	**0.44013**
ACQ-R15	0.04923	0.32419
ACQ-R16	**0.85746**	-0.04662
ACQ-R17	**0.58326**	-0.18395

From the results of the factor pattern and the rotated factor pattern, a two factor solution was established. ACQ-R1 and ACQ-R2 loaded on both factors, while ACQ-R10 did not load on either factor. As such, all three items were discarded from further analyses.

Items that loaded highly on factor one were ACQ-R3, ACQ-R4, ACQ-R5, ACQ-R6, ACQ-R12, ACQ-R13, ACQ-R16, ACQ-R17. The central theme underlying this factor appeared to be dissatisfaction and concern with body image. For example, ACQ-R3 and ACQ-R4 ask directly about whether the person is anxious or worried about their general physical appearance (ACQ-R3) or their muscularity (ACQ-R4), leading to ACQ-R6 which asks explicitly whether body image concerns undermine their relationships with other people. This is followed in ACQ-R12 which asks if the individual avoids places where his body will be exposed, implying a concern with appearance and feeling self-conscious. Finally, ACQ-R13, ACQ-R16 and ACQ-R17 look at body image anxiety directly related to muscularity, including wearing clothes that make one more muscular (ACQ-R13), comparing one's muscularity with other men (ACQ-R16), or feeling envious about another man's muscularity (ACQ-R17). These again could imply either satisfaction or dissatisfaction with body image,

specifically with the bodies' muscularity. In light of the theme of anxiety and possible unhappiness with the body, this factor was believed to be a measure of body image dissatisfaction, and was designated as ACQ-R_Image in later analyses.

Items that most heavily loaded on factor two were ACQ-R7, ACQ-R8, ACQ-R9, ACQ-R11, ACQ-R14, ACQ-R15, each of which were linked by the theme of actions taken to redress body image concerns. Indeed, two items asked about the person's eating and nutritional habits (ACQ-R7 and ACQ-R8), ACQ-R9 asked whether the individual supplemented their diet with legal drugs, while ACQ-R11 asked about how much money is spent on items designed to improve appearance. Finally ACQ-R14 asked how often the person measured aspects of their body, such as their muscles, and ACQ-R15 asked if the person had continued training even through injury. While these questions may seem different, they all involve an impetus on the part of the individual to perform actions that would lead to improved appearance. These concerns are different to the anxiety and concern in the body image dissatisfaction of factor one, as here the items seem to suggest actual behaviours performed to address body image dissatisfaction. As such, this label was thought more closely associated with disturbances of body image, and will hereafter be referred to as ACQ-R_Dysmorphia.

The distinction between the one scale measuring dissatisfaction with body image and the other a disturbed body image resonates with the work of Thompson (1995) who advocates the utility and necessity of such a distinction for assessing "normal" as opposed to clinical types of concerns. Having found two separate dimensions represented in the original 17-item ACQ-R, separate reliabilities were re-run on the two subscales, the results of which are presented in the following chapter.

5.3.3 The Bem Sex Role Inventory (BSRI)

The BSRI is made up of 60 items representing personality traits. 20 of these items are believed to be socially desirable for men (comprising the *Masculinity scale*), a further 20 items are believed to be socially desirable for women (comprising the *Femininity scale*), while the remaining 20 traits act as neutral or filler items. The measure works on a 7-point Likert type scale ranging from "never or almost never true" to "always or almost always true." The scale yields a *Masculine score* for the masculine items and a *Feminine score* for the feminine items (Bem, 1974; 1975). It is also possible to

calculate an *Androgyny score*. The masculinity variable in this research was operationalised using the score obtained on the Masculinity scale of the BSRI, where higher scores represent more endorsement of stereotypical masculine traits. However, while this research was primarily concerned with masculinity, it was argued in the literature that masculinity can exist only in relation to femininity. As such, a femininity score was calculated from the Femininity scale to represent the extent to which the individual identified with stereotypically feminine traits. Higher scores on femininity represent higher endorsement of the feminine traits.

The BSRI is conventionally used to classify people into one of four categories, namely masculine, feminine, androgynous or undifferentiated using a median split method. High on the Masculinity scale and low on the Femininity scale will suggest that an individual is *masculine*, while low on the Masculinity scale and high on the Femininity scale is suggestive of a *feminine* individual (Bem, 1974). One who is *androgynous* is both high on the Femininity and Masculinity scale, while low on both scales suggests an *undifferentiated* individual (Bem, 1976). This classification was not directly applied to this research, as dividing the students amongst categories would have excluded participants from some analyses. Rather, each person was assigned a masculinity, femininity and androgyny score. Details of the way in which androgyny was calculated is found in section 5.3.3.6. As such, when the variable gender identity is referred to in the present study, it can be taken to designate one's sense of self as masculine, feminine or androgynous.

5.3.3.1 Psychometric analyses of the BSRI

During its construction, the BSRI underwent several psychometric analyses in samples of students from Foothill College and Stanford University. The results showed good internal consistency reliability in the Foothill sample, with *Masculinity* = .86; *Femininity* = .82 and *Social Desirability* = .70. Similarly the Stanford sample also showed satisfactory results with coefficients of .86 for *Masculinity*; .80 for *Femininity* and .75 for *Social Desirability* (Bem, 1974). High test-retest reliability was also found from the scores of the first and second administration of the Stanford sample (*Masculinity*: r = .90; *Femininity*: r = .90; *Androgyny*: r = .93; *Social Desirability*: r = .89) (Bem, 1974). While this suggests that the BSRI has a sound reliability record, the Masculinity and Femininity scales were re-analysed in the

context of this study. In this regard, initial Cronbach alpha coefficient for the Masculinity scale was .77, with the standardised coefficient slightly higher at .79. The Femininity scale had a slightly higher reliability with a raw Cronbach Alpha Coefficient of .79 and a standardised coefficient of .82. As the Androgyny score was calculated as a function of the masculinity and femininity items, a reliability check could not be obtained. While both correlations represented acceptable reliability, for reasons considered below, the BSRI underwent further psychometric analysis in the present research, and final reliability estimates for the scales are found in the results section.

While the reliability of the BSRI is typically high, the validity of the scale is more difficult to assess, as the traits were derived empirically rather than from theory. This led Pedhazur and Tetenbaum (1979) to criticise the BSRI on its limited ability to show construct validity, as the process of construct explication cannot be effected unless one has predefined understandings of masculinity and femininity that can be compared to other related or unrelated constructs. Bem (1979) replied to this critique as follows:

> "*The theory underlying the BSRI asserts that sex-typed individuals will conform to whatever definitions of femininity and masculinity the culture happens to provide. The theory deliberately does not specify the particular content of these definitions, however, because these will vary from culture to culture*"
> (p. 1049)

The literature has indeed suggested a growing sense that masculinity and femininity can take multiple forms that are likely to shift over time. As such, in order to determine whether the Masculinity and Femininity scales were still reliable and valid measures of these constructs, it was deemed appropriate to conduct a factor analysis to confirm the properties and validity of the scale. As with the ACQ-R, this factor analysis used a principal component analysis to extract the initial factor solution, which was then rotated using a varimax rotation. As a similar procedure was followed when investigating the ACQ-R, the theoretical basis of principal component analysis will not be repeated to avoid redundancy.

The same three guidelines discussed in section 5.3.2 were used to determine the number of factors to extract. While there are 60 items on the BSRI, the summary of

the eigenvalue correlation matrix presented in Table 5.5 below contains only the first 18 eigenvalues due to space constraints. In view of the general rule of thumb expressed earlier, this result suggested a 16-factor solution, though again this was very likely an overestimate of the number of relevant factors. However, the table may also sensibly suggest a three or four factor solution in view of the proportion of variance explained by a three-factor solution (39.5%) or a four-factor solution (47%).

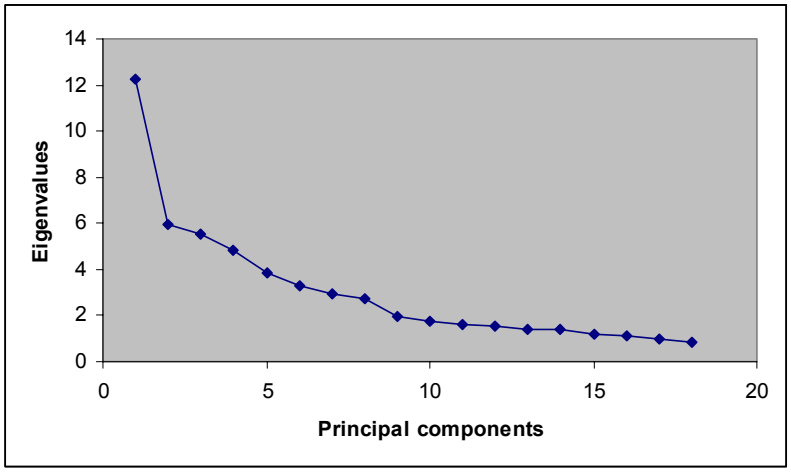
Table 5.5

Eigenvalues of the correlation matrix for the BSRI

	Eigenvalue	Difference	Proportion	Cumulative
1	12.2628390	6.3292393	0.2044	0.2044
2	5.9335997	0.4385526	0.0989	0.3033
3	5.4950470	0.6989070	0.0916	0.3949
4	4.7961400	0.9268766	0.0799	0.4748
5	3.8692635	0.5692803	0.0645	0.5393
6	3.2999832	0.3318609	0.0550	0.5943
7	2.9681222	0.2601797	0.0495	0.6437
8	2.7079425	0.7253094	0.0451	0.6889
9	1.9826331	0.2181050	0.0330	0.7219
10	1.7645281	0.1206002	0.0294	0.7513
11	1.6439279	0.1024160	0.0274	0.7787
12	1.5415119	0.1223819	0.0257	0.8044
13	1.4191300	0.0380116	0.0237	0.8281
14	1.3811184	0.2167025	0.0230	0.8511
15	1.1644159	0.0431886	0.0194	0.8705
16	1.1212272	0.1286350	0.0187	0.8892
17	0.9925923	0.1182355	0.0165	0.9057
18	0.8743567	0.1171533	0.0146	0.9203

The scree plot reproduced in Figure 5.2 suggested a two or three-factor solution. On balance, it was decided that a three-factor solution would be most appropriate.

Figure 5.2
Scree plot of eigenvalues and principal components for the BSRI



Before the factors could be interpreted, a factor rotation was performed using varimax. For ease of reference, the traits that each item refers to are included in the table. The results are as follows:

Table 5.6

Rotated factor pattern for the BSRI

	Characteristic	Factor one	Factor two	Factor three
BEM1	Self-reliant	-0.15005	**0.63131**	-0.39037
BEM2	Yielding	0.27285	**0.44650**	-0.13624
BEM3	Helpful	**0.76112**	0.14972	-0.08477
BEM4	Defend own beliefs	-0.06687	**0.51843**	0.08064
BEM5	Cheerful	0.15307	**0.39176**	-0.05945
BEM6	Moody	0.24697	-0.18830	**0.34198**
BEM7	Independent	0.10412	**0.24493**	0.02967
BEM8	Shy	0.36645	-0.20397	**-0.46097**
BEM9	Conscientious	**0.80849**	-0.15730	0.02849
BEM10	Athletic	0.26127	0.19764	**0.49910**
BEM11	Affectionate	**0.48024**	0.16794	-0.01517
BEM12	Theatrical	0.15144	**0.22218**	0.20007
BEM13	Assertive	0.37394	**0.40051**	-0.16616
BEM14	Flatterable	0.05617	0.19491	**-0.55401**
BEM15	Happy	0.17451	**0.51770**	-0.15445

BEM16	Strong Personality	0.26265	**0.61676**	-0.33436
BEM17	Loyal	**0.50627**	0.19417	-0.31337
BEM18	Unpredictable	-0.30436	**0.62547**	0.38650
BEM19	Forceful	0.08762	**0.59034**	-0.19601
BEM20	Feminine	0.24459	-0.00075	**0.27508**
BEM21	Reliable	0.48972	0.00878	**-0.65814**
BEM22	Analytical	0.32180	**0.41157**	0.15696
BEM23	Sympathetic	**0.68430**	0.14769	0.25263
BEM24	Jealous	0.19108	-0.30663	**0.39159**
BEM25	Have leadership abilities	0.18474	**0.75050**	0.14780
BEM26	Sensitive to the needs of others	**0.75622**	0.41689	0.19288
BEM27	Truthful	**0.73203**	-0.05313	0.28461
BEM28	Willing to take a stand	0.03143	**0.61300**	0.11094
BEM29	Understanding	**0.59535**	0.45075	0.04117
BEM30	Secretive	0.11174	**0.24992**	0.18654
BEM31	Makes decisions easily	**-0.55983**	0.21383	-0.10456
BEM32	Compassionate	**0.62191**	0.48158	0.07405
BEM33	Sincere	**0.72561**	0.21255	-0.05717
BEM34	Self-sufficient	-0.14660	**0.43638**	-0.24482
BEM35	Eager to soothe hurt feelings	**0.53106**	0.06196	0.47259
BEM36	Conceited	0.00608	-0.15282	**0.54296**
BEM37	Dominant	-0.08599	**0.61239**	0.08656
BEM38	Soft-spoken	**0.20608**	-0.09009	-0.13257
BEM39	Likeable	0.35339	**0.51302**	-0.16672
BEM40	Masculine	0.30423	**0.38696**	0.29370
BEM41	Warm	**0.73102**	0.17444	0.07189
BEM42	Solemn	0.03563	0.01706	**0.42852**
BEM43	Willing to take a stand	0.07260	**0.54795**	0.08889
BEM44	Tender	**0.73986**	0.25919	-0.02300
BEM45	Friendly	**0.63253**	0.05407	0.08888
BEM46	Aggressive	**-0.49110**	0.26979	0.24692
BEM47	Gullible	0.17867	0.07695	**0.57575**
BEM48	Inefficient	-0.46738	-0.13167	**0.52449**
BEM49	Acts as a leader	0.18954	**0.54002**	-0.02556
BEM50	Childlike	-0.01408	0.24196	**0.46213**

BEM51	Adaptable	-0.09430	**0.39873**	-0.29783
BEM52	Individualistic	-0.14062	**0.55572**	-0.06360
BEM53	No harsh language	0.13980	-0.03092	**0.24674**
BEM54	Unsystematic	-0.11969	0.11997	**0.76441**
BEM55	Competitive	**0.20493**	-0.04307	0.08967
BEM56	Loves children	**0.59372**	0.38224	0.31733
BEM57	Tactful	**0.35250**	0.25059	-0.24705
BEM58	Ambitious	0.35910	**0.51077**	0.17797
BEM59	Gentle	**0.78807**	0.32019	-0.11017
BEM60	Conventional	0.31575	0.22646	**-0.60035**

From the results of the rotated factor pattern, it can be seen that BEM7 (*independent*), BEM12 (*theatrical*), BEM20 (*feminine*), BEM24 (*jealous*), BEM30 (*secretive*), BEM38 (*soft-spoken*), BEM53 (*no harsh language*), BEM55 (c*ompetitive*), and BEM57 (*tactful*) did not load highly on a particular factor and so were discarded from further analyses. BEM13 (*assertive*), BEM51 (*adaptable*), and BEM40 (*masculine*) had fairly low loadings on factor two (approximately.4), but were retained in the analysis as they represent traits consistent with Bem's understanding of masculinity. For ease of engaging with the item loadings, the discussion of each factor includes a series of tables indicating the way in which the findings from the present study were both similar and different to the original BSRI.

5.3.3.2 Factor one

The 18 items that loaded significantly on this factor seemed to resonate most strongly with stereotypically feminine traits. Indeed, 12 of the items in this factor are found on the original BSRI Femininity scale, as seen in the table below. The additional items such as *helpful*, *conscientious*, *tactful* and *friendly* are not adverse to the BSRI feminine traits such as *sympathetic*, *understanding*, *warm* and *compassionate*. Even the items *makes decisions easily* and *aggressive* which should have been on the Masculine scale still made intuitive sense loading on this factor as the correlation was negative, implying that the characteristics represent the antithesis of femininity. For the characteristics that did not load on Femininity but should have according to the BSRI, with the exception of *cheerfulness*, the findings corroborated Antill and Russell's (1982) study. Interestingly, the items that no longer loaded on Femininity

appeared to highlight more stereotypical versions of women as subservient, which in the new Femininity scale appears to have been discarded. Consequently the characteristics that made up factor one formed the items that were used as the measure of femininity in the present research.

Table 5.7

Items loading on factor one - Femininity

	Items loading on factor one: ***Femininity***	*Items loading on factor one that* **ARE** *part of BSRI Femininity*	*Items loading on factor one that* **ARE NOT** *part of BSRI Femininity*	*BSRI feminine items that* **DID NOT** *load on factor one*
1	Helpful		Helpful	Cheerful
2	Conscientious		Conscientious	Childlike
3	Affectionate	Affectionate		Does not use harsh language
4	Loyal	Loyal		Flatterable
5	Sympathetic	Sympathetic		Gullible
6	Sensitive to needs of others	Sensitive to needs of others		Shy
7	Truthful		Truthful	Yielding
8	Understanding	Understanding		
9	Make decisions easily (Negative)		Make decisions easily (Negative)	
10	Compassionate	Compassionate		
11	Sincere		Sincere	
12	Eager to soothe hurt feelings	Eager to soothe hurt feelings		
13	Warm	Warm		
14	Tender	Tender		
15	Friendly		Friendly	
16	Aggressive (Negative)		Aggressive (Negative)	
17	Love children	Love children		
18	Gentle	Gentle		
			Tactful	

5.3.3.3 Factor two

While factor 1 had fewer of the Feminine items than expected, factor two clearly measured the masculinity component of the BSRI. Indeed, 15 of the original 20 Masculine BSRI items loaded on this factor. Interestingly, there were also several additional positive characteristics that loaded on masculinity which were originally associated with neutral items, including *happy*, *cheerful*, and *likeable*. This implied

that masculinity is perceived as having a more positive disposition. However, a surprising item loading on this factor was *yielding*. This trait would seem to be the antithesis of masculinity in view of the other items such as *forceful*, *dominant* and *willing to take a stand*. While it is possible that students were not familiar with the word, it should be noted that this corroborates the Antill and Russell (1982) study which also showed *yielding* to load on masculinity. Similarly, athleticism, which was thought to be a masculine trait in the BSRI, loaded instead on factor-three. This may be symbolic of athleticism no longer being regarded as a gender specific trait.

Table 5.8
Items loading on factor two – Masculinity

	Items loading on factor two : ***Masculinity***	*Items loading on factor two that* **ARE** *part of BSRI Masculinity*	*Items loading on factor two that* **ARE NOT** *part of BSRI Masculinity scale*	*BSRI masculine items that* **DID NOT** *load on factor two*
1	Self-reliant	Self-reliant		Aggressive
2	Yielding		Yielding	Make decisions easily
3	Defend own beliefs	Defend own beliefs		Athletic
4	Cheerful		Cheerful	
5	Assertive	Assertive		
6	Happy		Happy	
7	Strong personality	Strong personality		
8	Unpredictable		Unpredictable	
9	Forceful	Forceful		
10	Analytical	Analytical		
11	Have leadership abilities	Have leadership abilities		
12	Willing to take risks	Willing to take risks		
13	Self-sufficient	Self-sufficient		
14	Dominant	Dominant		
15	Likeable		Likeable	
16	Masculine	Masculine		
17	Willing to take a stand	Willing to take a stand		
18	Act as a leader	Act as a leader		
19	Adaptable		Adaptable	
20	Individualistic	Individualistic		
21	Ambitious	Ambitious		

5.3.3.4 Factor three

The remaining items were understood to represent the neutral items given the diverse range of traits captured that together do not seem conventionally feminine or masculine. Of the 20 original neutral items in the BSRI scale, only 7 loaded significantly on this factor. With the exceptions of *athleticism,* and the negative correlation of *shy*, the remainder of the traits on this factor seemed less desirable than those found on the other two factors. For example, an individual who is moody, unreliable, conceited, inefficient and unsystematic is likely the antithesis of what is socially desirable. As such, it is possible that rather than measuring the social desirable traits that were deemed to be neutral by Bem (1974), this factor may in fact represent a measure of socially undesirable traits for both men and women.

Table 5.9
Items loading on factor three – Neutral items

	Items loading on factor 3: ***Neutral***	*Items loading on factor 3 that* **ARE** *part of BSRI Neutral*	*Items loading on factor 3* **NOT** *part of BSRI Neutral scale*	*BSRI neutral items that* **DID NOT** *load on factor 3*
1	Moody	Moody		Adaptable
2	Shy (negative)		Shy (negative)	Conscientious
3	Athletic		Athletic	Friendly
4	Flatterable (negative)		Flatterable (negative)	Happy
5	Reliable (negative)	Reliable (negative)		Helpful
6	Conceited	Conceited		Likeable
7	Solemn	Solemn		Unpredictable
8	Gullible		Gullible	
9	Inefficient	Inefficient		Likeable
10	Childlike		Childlike	Tactful
11	Unsystematic	Unsystematic		
12	Conventional (negative)	Conventional (negative)		Truthful

The previous results suggested that the original conceptualisations of masculinity and femininity have shifted from Bem's original scale. Consequently, to provide a valid measure of gender identity for the present research, it was deemed most appropriate to define the characteristics of femininity and masculinity as that measured by factor one and factor two respectively. As per the original BSRI, the neutral items acted as "fillers" and were not used in further analyses. Having found suitable items that

measured masculinity and femininity, reliability analyses were rerun on the revised subscales. The results are presented in chapter six.

5.3.3.5 Correlation amongst the scales

Prior to the BSRI's revision in the present study, the Masculinity and Femininity scales were positively correlated (r = .41) at the 0.01% level of significance. This implied that higher scores on masculinity tended to be associated with higher scores on femininity, and is antithetical to the way in which Bem (1974) conceptualised masculinity and femininity as separate and independent. The revised version of the BSRI in the present research did, however, decrease this correlation to r = .34 significant at the 0.1% level of significance. Similar positive correlations have been reported in recent research using the BSRI (for example, Cameron, 2003; Zlotnick, 2002).

Cameron (2003) suggested this could be attributed to the apparent emphasis in schools on gender equality and the socialisation of androgynous traits for boys and girls. The present research disagreed with Cameron's supposition that the result was "expected" (as quoted in the review), as it has concerning implications for the ability of Bem's scale to discriminate between masculine and feminine characteristics. However, the present research does agree with Cameron's (2003) suggestion that the finding may point to the possibility that men today may be socialised to adopt an integrated set of traits rather than a polarised set, and hence this may limit versions of either masculinity or femininity. As such, individuals may be encouraged to become more androgynous or at least to see equal benefit in qualities of both genders. This latter argument can be substantiated when one engages with the way in which the subscales of the BSRI were reworked in the present research, as both subscales provided evidence of both masculinity and femininity as being desirable. It was also seen that more integrative versions of gender identity appear to be more common in the present sample (see Table 5.1). Despite the positive correlation between the subscales, it was still deemed appropriate to use the BSRI as the measure of gender identity as each subscale still measured antithetical constructs, and so was still able to differentiate between masculine and feminine traits.

5.3.3.6 The calculation of androgyny

While the calculation of the total masculinity and femininity score were fairly straightforward, Androgyny proved somewhat more challenging and was calculated using the following equation:

$$\text{Androgyny} = \left| \frac{Z_{Mas} - Z_{Fem}}{Z_{Mas} + Z_{Fem}} \right|$$

where

$$Z_{Mas} = \frac{\text{Masculinity} - \overline{X}_{\text{Masculinity}}}{S_{Masculinity}}$$

$$Z_{Fem} = \frac{\text{Femininity} - \overline{X}_{\text{Femininity}}}{S_{\text{Femininity}}}$$

In order to calculate the androgyny measure the raw femininity and masculinity scores were standardized to Z scores. This reflects a large difference between masculinity and femininity as a measure of androgyny. The consequence of this equation is that higher androgyny is denoted by lower androgyny scores.

5.3.4 The Psychosocial Inventory of Ego Strength (PIES)

The PIES, developed according to Erikson's theory, was designed to measure specific and overall ego strength as explained in the literature. There are eight questions for each of the eight ego strengths in the theory, yielding a total of 64 items. Four of the eight items for each subscale are negatively phrased and hence reverse scored. The response format is a Likert-type scale ranging from 1 ("does not describe me well") to 5 ("describes me very well"). Items are summed to obtain a score for each specific subscale, as well as a total for overall ego strength where higher scores indicate higher levels of ego strength. For the combined 64 items, Cronbach's alpha was reported at .94 (Markstrom et al., 1997). A shorter version of the scale also exists, consisting of 32-items. In this version there are four, rather than eight, questions per subscale, two of which are negatively phrased and hence reverse-scored. To avoid an unnecessarily lengthy and sometimes redundant questionnaire, the 32-item version was used as a measure of overall ego strength in the present research. It was not deemed feasible to calculate specific ego strength scores for each subscale as they now consisted of only four items each. While the reported internal consistency reliability for the shorter

scale is very high at .91, reliability for the PIES was re-analysed in the context of this study.

The initial stage of developing the PIES began with the generation of themes underlying the eight ego strengths in Erikson's theory and their antipathies (Markstrom et al., 1997). This led to the creation of numerous positively phrased items to represent the particular ego strength, as well as negatively phrased items to represent the corresponding antipathy. While the items that lacked face validity were discarded, the more acceptable items were scrutinised and rephrased to best capture the themes. Face validity was established using six graduate students who assessed the appropriateness of the items and identified problematic areas. Eriksonian scholars were used to provide expert opinion on the measure's ability to adequately tap into the content domain and so assessed its content validity. The PIES was then administered to a sample of predominantly female undergraduate students at Canadian University, and then later to a second sample of mixed male and female undergraduate students (Markstrom et al., 1997).

The Cronbach alpha coefficient for the combined 64 items at the first validation was .93, while the reliability of the eight individual subscales ranged from acceptable to very good, with only *purpose* having a low coefficient of .52. By the second set of validation the coefficient improved slightly to .94 for the 64 item version. When an attempt was made at this stage to collapse the scale to a 32 item version, the coefficient for overall was found to be .91. According to Markstrom et al. (1997), convergent validity for the scale was shown through significant positive correlations with measures of identity achievement, purpose in life, internal locus of control, and self-esteem, as well as negative correlations between overall ego strength and hopelessness, identity diffusion and moratorium, and personal distress.

5.3.4.1 Reliability analysis of the PIES

Calculating the reliability of the PIES was lengthier than the previous scales, as 16 items needed to be reverse scored as per Markstrom et al.'s (1997) guidelines. Having done so, the reliability stood at .77 (raw and standardised). PIES3, PIES7, and PIES13 were still, however, negatively correlated. In looking at the items, it was clear that PIES7 should be reversed scored, but that PIES3 and PIES13 should not be reverse

scored. While reverse scoring for PIES7 was then effected, PIES3 and PIES13 remained negatively correlated. The requirements of validity were deemed to outweigh the possible benefits of the increased reliability that would occur should the items be reverse scored. Importantly, the reliability remained adequate without reverse scoring the items. As the scale was used only to attain a measure of overall ego strength, reliability analyses were not conducted on the individual subscales, and the result of the overall reliability is presented in the next chapter.

5.4 Research Design

This is a non-experimental correlational research design as there is no control group, no random assignment and the researchers chose not to manipulate an independent variable (Niele & Liebert, 1986). This is appropriate as the aim of the research was to provide an account of the relationships between the variables of interest. The results may, however, serve as a pre-cursor for further experimental inquiry. It was also a cross-sectional design as data was collected at a single point in time.

5.5 Ethical Considerations

While no risks were envisaged for the participants in this research, the following ethical considerations were adhered to in protecting their welfare. In terms of informed consent, the information sheet outlined the nature and purpose of the study. It explained that the area of focus was men's feelings toward their bodies and how this translated into feelings about themselves. Students were informed that participation would require a questionnaire to be completed that would take no more than 20 minutes, and were encouraged to read the information sheet of the questionnaire. The latter stated that participation was completely voluntary and that they would be neither advantaged nor disadvantaged by participating. This was verbally reiterated by the researcher. It also stated that students were not obliged to answer any or all the questions if they chose not to, and that no identifying information was to be written anywhere on the questionnaire to ensure anonymity. It was also explained that as a result, no individual feedback could be given as the researcher could not trace responses back to particular participants. However, it stipulated that a summary of the overall results would be made available to students once the data had been analysed and reported. In order to keep responses confidential, students willing to participate were asked to stagger themselves around the room

before completing the questionnaire, and then to seal their questionnaire in the envelope provided before handing it back. Contact details of a counselling centre were provided for participants if they felt they required this service after completing the questionnaire. Only the researcher and research supervisor had access to the responses.

5.6 Statistical Procedures Used

In view of the research questions posed and type of design used in the present study, the most appropriate statistical procedures were correlation, t-tests and multiple regression. Correlation was used to assess the relationships between body image, gender identity and ego strength, while a t-test was used to determine how sexual orientation was related to gender identity and body image. Finally, multiple regression was used to assess whether there are particular combinations of gender identity and ego strength which predict higher scores on body image dissatisfaction. The logic and assumptions of these statistical methods will now be discussed.

5.6.1 Correlation

Correlation refers to the extent to which one variable is related to another variable, as measured by a correlation coefficient (Howell, 1999). While different kinds of correlation coefficients can be used to assess the presence of a relationship, the Pearson's product moment correlation (also known as Pearson's *r*) and Spearman's rank correlation coefficient (also known as Spearman's *rho*) are the most common and were used in the present research (Howell, 1999). While the value of the coefficient expresses the degree of linear relationship between the two variables in both correlation coefficients, Spearman's *rho* technically only requires a monotonic rather than a linear relationship. It is, however, linear in the ranks of items (Kendall, 1962).

A perfect positive relationship is denoted by + 1, while a perfect negative relationship is denoted by – 1. The closer a coefficient is to 1 or -1, the stronger the linear relationship. A coefficient of zero implies that no linear relationship exists (Howell, 1999). The sign of the coefficient (either positive or negative) indicates the direction of the relationship, but has no bearing on the strength of the relationship. A positive relationship suggests that higher scores on one variable tend to be associated with

higher scores on the other variable. In contrast, a negative relationship suggests that lower scores on the one variable tend to be associated with higher scores on the other variable (Howell, 1999). The significance of the correlation is also important, as it suggests that the correlation was "greater than likely to have arisen by chance" (Kline, 1994, p. 21). The more significant a correlation is, the more confident one can be that there truly is a relationship between the variables.

Pearson's *r* is a parametric test which requires certain assumption to be met, while Spearman's *rho* is the non-parametric equivalent typically used when the parametric assumptions fail (Howell, 2002). Pearson's *r* requires the assumptions of normality, equality of variance, and that the data is measured on an interval scale. Spearman's *rho* does not require these assumptions, and is appropriate for use on ordinal data (Pietersen & Damianov, 1998).

To test the assumption of normality, a histogram was generated for each measure. Where normality is met, the histogram should show a symmetric distribution designated by a bell-shaped curve. Further indicators of normality include the skewness and kurtosis values, which should be between 1 and -1 to denote adequate normality. A symmetric distribution is also denoted by equal values of the mean, median and mode, though a significant discrepancy between these values suggests that the data is skewed (Howell, 1999). The second assumption, equality of variance, designates that the variance of Y for each value of X be constant (Howell, 1999). To test this, a scatter diagram for each of the measures was generated, and an assessment of any outliers or influential plots was made. Pearson's *r* was carried out where the parametric assumptions were met, failing which Spearman's *rho* was calculated.

The histograms of ego strength (PIES scores), the BMI scores, masculinity, femininity and androgyny demonstrated no concerns with skewness or kurtosis and were approximately normally distributed. As they are interval scale variables and no problems with equality of variance were found, it was appropriate to use Pearson's *r*. However, both of the ACQ-R subscales demonstrated problems with normality from the histogram, which was then confirmed by the skewness values of 1.3 (ACQ-R Image) and 1.9 (ACQ-R Dysmorphia) respectively. While both variables are an interval scale of measure, additional problems with equality of variance were picked

up from the scatter plot with respect to ACQ-R Dysmorphia. Spearman *rho* was consequently deemed appropriate for these latter variables.

5.6.2 t - test

A two independent sample *t*-test was used to compare heterosexual and gay men with regard to gender identity and body image concerns. This test can be used to establish whether any difference in means found between two independent groups is sufficiently large to justify the conclusion that the two samples are drawn from different populations (Howell, 1999). It requires the assumption of homogeneity of variance, that "the sampling distribution of differences between means is normal", and that the measures are an interval scale (Howell, 2004, p. 318). In the case of homogeneity, a simple adjustment to the degrees of freedom can be applied to provide a reasonable test in the presence of heteroscedasticity (SAS Institute, 1993). While the gender identity variables met all the assumptions for the test, problems were found with normality with regard to the ACQ-R_Image, and normality and homogeneity of variance with regard to the ACQ-R_Dysmorphia. While caution was applied when interpreting any significant result found with regard to these subscales, the Central Limit Theorem suggests that the test may nevertheless be valid in view of the sample size (Howell, 2004).

5.6.3 Multiple regression

The second set of statistics conducted was multiple regression. This is a procedure that allows one to assess whether a linear combination of several predictor variables can predict the criterion variable (Minium, 1978). In other words, the regression equation will state the predicted value of Y (the criterion variable) on the basis of simultaneous knowledge of several predictor variables ($X_1, X_2, X_3..., X_p$) (Howell, 2002). In this research, the criterion variable was scores on the ACQ-R_Image (body image dissatisfaction), and the predictor variables were masculinity, femininity, androgyny and ego strength. In order to show whether changes in the criterion variable can be attributed to changes in the predictor variables, the predictor variables need to be correlated with the criterion variable (Howell, 2002). If the correlations are high, the actual value will cluster more closely about the predicted value, while low correlations imply that there may be considerable variation of actual values about the predicted values (Minium, 1978). This was assessed by looking at the results of the

correlations discussed in the previous section. Assuming this type of correlation exists, one can then calculate a line of regression.

While a regression equation can be constructed in several ways, this research used a variable selection processes known as stepwise backward elimination. In backward elimination, all predictor variables are initially included in the analysis. From the results of the significance tests, the predictor variable that contributes least to the model is identified and removed. The model is then rerun without the removed variable, and the predictor that again does not significantly contribute to the model will be removed. This process is continued until all remaining predictor variables are significant (Howell, 2002).

The output of a multiple regression typically begins with an ANOVA table that is used to assess whether any of the variables can predict body image dissatisfaction. This is accompanied by a value of R -squared, which indicates how much variation in the dependent variable can be predicted on the basis of the predictor variables (Minium, 1978). Assuming the values in this table are statistically significant, one can then begin to track which variables were removed and retained in the analysis.

The interpretation of the regression results requires consideration of several things. For example, the overall sample size for each combination is essential to consider when interpreting the regression. For Howell (1999), a rule of thumb is that there should be at least 10 observations for every predictor, though a more specific concern is that regression is most accurate in areas of the domain where there are a substantial number of observations. This is because the regression will produce predictions even when the domain is poorly defined, and caution must be taken not to predict beyond the range of the data. This is known as extrapolation.

> "*Extrapolation refers to predictions or estimates which we make for x (independent variable) scores which fall outside the range of the x scores in the data used to calculate our equation. Extrapolation should be avoided, or very cautiously interpreted, because we have no idea what the relationship between the variables is outside the range of the data*"
> (Pitman, 2003, p. 3).

A further concern with multiple regression is *multicollinearity*. This implies a high degree of correlation amongst the predictor variables, and suggests "that the regression equation is very unstable from one sample of data to another" (Howell, 2004, p. 240). Where this occurs, the substantive meaning of the individual regression coefficients becomes hard to interpret, as the contribution of any one predictor may be accounted for by another (Kerlinger & Pedhazur, 1973). A condition index was generated to test for multicollinearity in the present research, where a higher index suggests a stronger presence of multicollinearity. While multicollinearity makes the meaning of the individual parameter estimates difficult to interpret, it does not affect the predictive power of the regression. Nevertheless, given that the parameters of the regression are affected by multicollinearity, it was deemed appropriate to construct a table of means in order to facilitate the interpretation of the multiple regression in the present research.

In this table of means, the domain of *X* scores is divided into eight quadrants determined by median splits on each of the three predictor variables which remained in the model. The mean score for each quadrant was calculated in two ways. Firstly, the regression score represents the predicted values from the regression equation assuming the *X* value (the predictor variable) is either one standard deviation below the mean (denoting low scores in the table) or one standard deviation above the mean (denoting high scores in the table). Secondly, as a confirmatory score, the mean criterion score of all observations with predictor values below the median (low scores) and above the median (high scores) were presented.

5.6.3.1 Assumptions of multiple regression

The regression model requires normality of the residuals, homogeneity of variance of the residuals, linearity, and that the scales are an interval measure (Howell, 2002). Residuals are important indicators of the error of prediction, as they represent the "difference between the obtained and predicted values of *Y*" (Howell, 1999, p. 205). To test the normality of the residuals of ACQ-R_Image, a histogram of the residuals is generated. The distribution should appear in the form of a bell-shaped curve, with approximately equal measures of central tendency (Howell, 1999). Equality of variance is determined by a plot of residuals against predicted values for ACQ-R_Image. This plot will indicate whether the points are spread evenly around the line

or whether some variables are more spread than others. Problems with equality of variance can be seen in the latter scenario.