Store Image: Scale development
Part 2

RONEL DU PREEZ
ELIZABETH VISSER
HESTER JANSE VAN NOORDWYK
Department of Industrial Psychology
Stellenbosch University
South Africa

Correspondence to: Ronel du Preez
e-mail: rdp@sun.ac.za

ABSTRACT

The aims of this article (the second in a three-part series) are threefold, namely to (1) develop a scale for the measurement of the perceived importance of store image dimensions , (2) purify the developed scale to illustrate acceptable reliability and (3) develop and refine this scale for practical implementation in the apparel retail environment. A four-phase approach was documented for scale development. The provisional scale was purified and tested by means of two pilot studies and the data was subjected to Cronbach alpha and confirmatory factor analysis (CFA). A revised model of apparel store image was proposed. Model fit results indicated that fit can still be improved. Results culminated in a 55-item Apparel Store Image Scale that showed good reliability.

Keywords: store image, dimensions, apparel, conceptual model, multi-dimensional

Part 1 of this series of articles delineated the domain of store image and described the dimensions and subdimensions. Two models were proposed and a definition of store image was formulated to serve as point of departure for the following phase in scale development. The importance of this first phase in scale development cannot be overstated, since it is a prerequisite for determining the validity of a measurement scale and, more specifically, content validity (Netemeyer, Bearden & Sharma, 2003).

The development of measurement scales with desirable reliability and validity properties is a critical element in the evolution of a fundamental body of knowledge in a specific field of study. Churchill (1979) proposed a framework that is often employed as point of departure in measurement scale development (Blankson & Kalafatis, 2004; Grace, 2005; Li, Edwards & Lee, 2002; Parasuraman, Zeithaml & Berry, 1988; Shimp & Sharma, 1987). Based on Churchill's framework, as well as drawing from recommendations made by DeVellis (2003), Hair, Black, Babin, Anderson and Tatham (2006) and Netemeyer *et al.* (2003), four broad phases were identified in the scale development process (see Figure 1).

Part 1 of this series of articles discussed Phase 1. Phases 2 and 3 are reported in this paper while Phase 4 will be discussed in Part 3 of this series.

The aims of this paper are threefold, namely to
- develop a scale for the measurement of the perceived importance of the apparel store image dimensions;
- purify the developed scale to illustrate acceptable reliability; and
- develop and refine this scale for practical implementation in the apparel retail environment.

Generation and judging of measurement items (Phase 2)

The main objective of the second phase is the generation of measurement items (based on the work reported in the first article) that adequately represent the store image construct and domain, as well as the consequent judging of measurement items, as recommended by Churchill (1979), DeVellis (2003), Hair *et al.* (2006) and Netemeyer *et al.* (2003). The appropriate

operationalisation of a construct is imperative for valid empirical results and interpretation (Little, Lindenberger & Nesselroade, 1999; MacCallum & Austin, 2000). Therefore, the primary focus of this phase, in conjunction with the first phase, was to establish content and face validity of the measurement instrument (DeVellis, 2003; Netemeyer *et al.*, 2003). This was accomplished through initial item pool generation using existing literature and review by both expert and sample population judges. The domain sampling model was used as a basis for generating measurement items, since consequent reliability assessment reflects this model. Therefore, items were generated systematically to sample all content areas of store image as defined by the model.

Generation of measurement items

Scale development studies frequently report item generation from a review of the literature (Bearden, 2001; Dabholkar, Thorpe & Rentz, 1995; Grace, 2005; Lastovicka, Bettencourt, Hughner & Kuntze, 1999; Li *et al.*, 2002; Lichtenstein, Ridgeway & Netemeyer, 1993; Little *et al.*, 1999; Terblanché & Boshoff, 2004). Some reported extensive qualitative procedures to generate items used in empirical studies on store image (Amirani & Gates, 1993; Birtwistle & Siddiqui, 1995; Janse van Noordwyk, 2002; Thompson & Chen, 1998; Zimmer & Golden, 1988). For the purposes of this study, it was decided to rely heavily on extant literature for generating the initial item pool. The researcher, expert judges and sample population judges added a few items.

A composite list of attributes, previously employed as items in store image research, was compiled based on the model of store image. The inclusion of items was based on criteria reported in the reviewed store image studies, as well as guidelines from scale development literature. These criteria are summarised in Table 1.

Items generated from qualitative research were also reviewed (Birtwistle & Siddiqui, 1995; Thompson & Chen, 1998), as well as items with previous empirical support in store image literature reported by Lindquist (1974–1975) and Martineau (1958). Items that were not generated from the review of literature but deemed relevant by the researcher were also included.
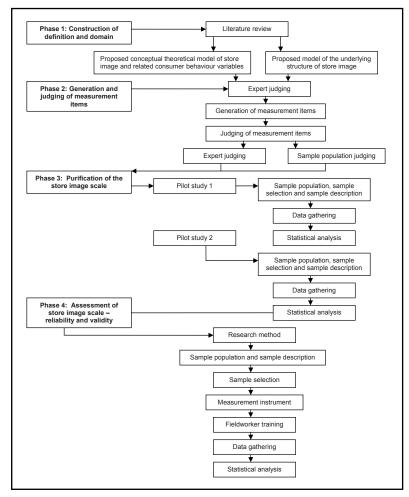
FIGURE 1
Scale development process
(adapted from Schlechter, 2005, p. 148)

The initial generation of items resulted in a composite list of 371 items. This was a cause for concern, but it was decided to retain the item pool based on the following three considerations: Firstly, in the early stage of the scale development process, it was preferable to be over-inclusive. Secondly, the internal consistency of a scale is a function of how strongly items correlate with each other. At this stage of scale development, however, the correlation between items was unknown. Lastly, the scale was submitted for judging of the measurement items, which would assist the process of refining the scale, as recommended by DeVellis (2003) and Netemeyer *et al*. (2003).

Each item from the composite list was rewritten into a declarative statement to which respondents have to indicate varying degrees of endorsement of the statement (Likert-type scale). Subsequently, each item was revised for clarity, ambiguity and double-barrel statements. This procedure was in accordance with recommendations in the literature (Brace, 2004; Bradburn, Sudman & Wansink, 2004; Churchill & Iacobucci, 2005; DeVellis, 2003; Frazer & Lawley, 2000; Kerlinger & Lee, 2000; Netemeyer *et al*., 2003; Oppenheim, 1992; Synodinos, 2003).

The 371 items were grouped within each dimension and subdimension, as guided by the model of apparel store image (see Figure 3, Part 1). The final item pool was reviewed to ensure that a sufficient number of items were included to adequately measure each dimension and subdimension, namely eight to ten items for each dimension as recommended by Netemeyer *et al*. (2003, p.147).

Judging of measurement items

Judging of the generated items served three distinct purposes, namely to ensure relevancy of the items to measure perceptions of the importance of store image, evaluate items for clarity and conciseness and identify possible areas of the store image domain that were not previously captured (DeVellis, 2003). Two experts in the fields of consumer psychology and apparel consumer behaviour conducted the first expert review. It comprised an evaluation of the initial measurement items (including format and layout of the scale) and the proposed model of store image.

The store image scale consisted of three sections. Section A covered the store image items categorised under the various store image dimensions. In Section B, respondents were requested to rate the individual dimensions on the same five-point response format. A demographic section, Section C

TABLE 1
Criteria for item inclusion from reviewed literature

CRITERIA FOR ITEM INCLUSION	
Coefficient alpha	≥ 0.7
Factor analysis	Eigen value > 1 Factor loadings ≥ 0.4 Coefficient alpha ≥ 0.7
Mean importance scores	Higher than average of given scale
Number of citations	Summed number higher than 90%

(e.g. gender, age, population group and home language), was included at the end of the measurement scale to avoid alienating respondents by asking for personal information at the outset of the measurement scale (Synodinos, 2003). A cover letter was compiled to convey important information regarding the research and completion of the scale to the respondents.

During this review, the experts gave consideration to the sample population of interest, namely the South African consumer. Within the South African context, the literacy and educational levels of consumers may vary significantly. English is not the home language of all respondents and a more refined scale could potentially result in inaccurate information given in response. This review culminated in a reduced item pool of 284 items and the retainment of the five-point Likert-type scale.

A panel of experts from different fields of study (apparel consumer behaviour, consumer psychology and statistics) undertook the second expert review. They were familiar with the research problem and objectives as well as the proposed store image model and scale. Based on their feedback, 57 items were eliminated and a further three items were generated, resulting in 230 items. Changes were made to the response format and rating scale. A six-point scale was used instead of the original five-point scale. Only the first and fifth point were anchored, namely 1 = unimportant and 5 = very important due to difficulties experienced with the inadequate description of the five anchor points. A sixth point was added to allow respondents a neutral response, namely 6 = unable to rate, and a visual presentation was added to the rating scale to aid responses, as recommended in the literature (Churchill & Iacobucci, 2005; DeVellis, 2003; Netemeyer *et al.*, 2003; Nunnally, 1978; Synodinos, 2003). The cover letter was adapted to reflect these changes.

Sample population judging serves the purpose of assessing the practical implementation of the scale with respondents similar to those employed in the administration of the scale in the subsequent phases of the scale development process. Therefore, two student group sessions were conducted. Based on the feedback of these group sessions, two more items were added to the scale (resulting in a scale consisting of 232 items) and smaller technical changes were made.

Purification of the measurement scale (phase 3)

Concerns regarding the length of the measurement scale had to be addressed in this phase. However, reducing the scale length

TABLE 2
Summary of goodness-of-fit-indices*

ABSOLUTE FIT MEASURES	
Minimum Fit Function of Chi-Square	A non-significant result indicates model fit
Normal Theory Weighted Least Chi-Square	A non-significant result indicates model fit
Root Mean Square Error of Approx. (RMSEA)	Values between 0.08 or below indicate acceptable fit. Values below 0.05 indicate good fit. Values below 0.01 indicate outstanding fit
Standardised Root Mean Residual (RMR)	Lower values indicate better fit with values below 0.05 indicating good fit
Goodness of Fit Index (GFI)	Higher values indicate better fit with values > 0.9 indicating good fit
Adjusted Goodness of Fit Index (AGFI)	Higher values indicate better fit with value > 0.9 indicating good fit
INCREMENTAL FIT MEASURES	
Non-Normed Fit Index (NNFI)	Higher values indicate better fit with values > 0.9 indicating good fit
Comparative Fit Index (CFI)	Values closer to 1 indicate better fit with values > 0.9 indicating good fit

*Adapted from Schlechter, 2005, p. 148

had to be done in conjunction with reliability, as reliability is a function of the number of items included in a scale (Churchill & Iacobucci, 2005; DeVellis, 2003; Netemeyer *et al.*, 2003). The purification phase included two separate pilot studies.

METHODOLOGY

The methodology for each pilot study will be discussed followed by a discussion of the results thereof.

Pilot study 1

The aim of the first pilot study was to obtain initial estimates of reliability as a basis for scale purification, as well as to aid in optimising scale length. A student sample population (n = 89) was deemed appropriate, since this study is concerned with the apparel consumer's perception of store image. Students were not considered entirely nonrepresentative, since they qualify as apparel consumers and form part of apparel market segments. This phase was concerned with providing evidence of internal consistency of the store image scale for which the sample was appropriate in terms of providing accurate results. Student samples are frequently employed in scale development research, which further serves to justify the student sample

TABLE 3
Reliability - pilot study 1

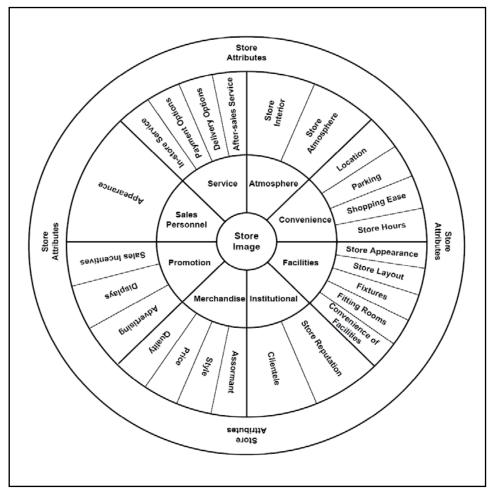| CONSTRUCT	DIMENSIONS	SUBDIMENSIONS	NO. OF ITEMS	COEFFICIENT ALPHA
Store image			232	0.90
	Atmosphere		15	0.57
		Store interior	11	0.75
		Store atmosphere	4	0.50
	Convenience		42	0.61
		Transportation	3	0.59
		Location	11	0.76
		Parking	6	0.75
		Shopping ease	18	0.82
		Store hours	4	0.87
	Facilities		41	0.78
		Store appearance	7	0.71
		Store layout	4	0.76
		Fixtures	14	0.84
		Fitting rooms	8	0.81
		Convenience of facilities	8	0.77
	Institutional		19	0.76
		Store reputation	9	0.73
		Clientele	10	0.75
	Merchandise		28	0.85
		Assortment	14	0.86
		Style	7	0.79
		Price	4	0.63
		Quality	3	0.60
	Promotion		33	0.81
		Advertising	17	0.89
		Displays	6	0.82
		Sales incentives	10	0.85
	Sales personnel		18	0.56
		Appearance	9	0.82
		Interaction	9	0.80
	Service		36	0.78
		In-store service	17	0.86
		Payment options	7	0.79
		Delivery options	5	0.90
		After-sales service	7	0.83

FIGURE 2
Revised model of apparel store image (after pilot study 1)

(Bearden, 2001; Grace, 2005; Li *et al*., 2002). The sample size was deemed adequate for this stage of the scale development process based on recommendations from literature (Blankson & Kalafatis, 2004; Dhurup, Venter & Oosthuyzen, 2005; Venter & Dhurup, 2005).

Statistical analysis

Statistica (version 8) was used for the analyses (StatSoft Inc., 2007). Coefficient alphas, item-total correlations and inter-item correlations were calculated for all items included within each subdimension. The cut-off value for coefficient alpha value was set at 0.7. The acceptable benchmark level for item-total correlations was set at above 0.3, with reports in the literature ranging from higher than 0.3 to higher than 0.5. The criterion for inter-item correlations was set at a range of 0.2–0.5 (Blankson & Kalafatis, 2004; DeVellis, 2003; Dhurup *et al*., 2005; Grace, 2005; Kerlinger & Lee, 2000; Netemeyer *et al*., 2003; Nunnally, 1978; Terblanché & Boshoff, 2004). The internal consistency of the subdimensions within each dimension was considered. Based on the results, the items in Section A were reduced to 214. No changes were made to sections B and C or the cover letter of the questionnaire.

Pilot study 2

The aim of the second pilot study was to provide additional evidence of scale reliability for scale purification, as well as to further reduce the scale length. A similar methodology was employed in the second pilot study. A convenience sample of students was recruited (n = 176). The 214-item measurement scale derived from the first pilot study was employed. A split-

sample approach was followed based on a 60:40 ratio, resulting in a training data set (n = 110) and a test data set (n = 66). This approach allowed for the purification of the scale based on the statistical analysis of the training data set, to be cross-checked by the statistical analysis from the test data set, as recommended by De Vellis (2003).

However, a scale that was representative of all subdimensions proposed in the model of apparel store image (refer to Figure 3 in Part 1) was still considered too long. In addition, to perform confirmatory factor analysis, each individual subdimension had to be represented by four items to allow for model identification (Hair *et al*., 2006). For the 27 identified subdimensions, this would result in, at least, a 108-item measurement scale, which would still be considered too long for practical usability. Therefore, the statistical analysis was only performed on

TABLE 4
Reliability - pilot study 2 (training data set - 214-item store image scale)

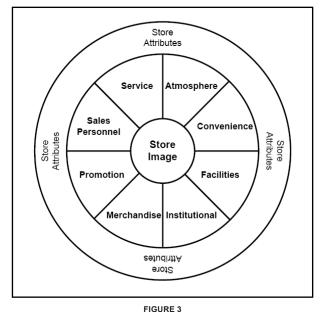| CONSTRUCT	DIMENSIONS	NO. OF ITEMS	COEFFICIENT ALPHA
Store image		214	0.90
	Atmosphere	11	0.84
	Convenience	38	0.93
	Facilities	37	0.93
	Institutional	17	0.85
	Merchandise	26	0.91
	Promotion	33	0.94
	Sales personnel	8	0.84
	Service	44	0.95

each of the eight broad dimensions associated with the store image construct. This was deemed acceptable to arrive at a measurement scale with optimum length whilst maintaining acceptable reliability.

Exploratory factor analysis (EFA): The training data set was employed in this statistical analysis procedure. Literature proposes that EFA and CFA be used in conjunction with one another (Fabrigar, Wegener, MacCallum & Strahan, 1999; Gorsuch, 1997). The model of apparel store image eliminated the need for EFA to establish the dimensionality of store image (refer to Figure 3 in Part 1). Therefore, the training data set was submitted to the principal axis factoring procedure and the analysis was constrained *a priori* to one factor for the separate investigation of each dimension. This is in accordance with previous studies employing this method for scale purification and optimising scale length (Bearden, 2001; Lastovicka *et al.*, 1999; Parasuraman *et al.*, 1988), as well as suggestions by Churchill (1979) to employ EFA as a means to confirm the number of conceptualised dimensions empirically after initial item evaluation through coefficient alphas and item-total correlations.

The cut-off value for factor loadings was set at a minimum of > 0.5 based on recommendations in the literature (Bearden, 2001; Blankson & Kalafatis, 2002; Grace, 2005; Hair *et al.*, 2006; Lastovicka *et al.*, 1999; Shrimp & Sharma, 1997; Tabachnick & Fidell, 2007). The training data set was further analysed through reliability measures and the criteria as per the previous pilot study were maintained. The results of all the statistical analyses were considered concurrently and concluded in a shortened measurement scale consisting of 55 items (refer to Appendix 1). A correlation analysis was done between the 214-item and the 55-item measurement scales.

Confirmatory factor analysis (CFA): Consequently, CFA was done on the test data set, employing the shortened measurement scale. For the purposes of this phase of the study, each dimension was submitted to CFA separately to allow for the investigation of individual items for further scale purification.

The measurement models for each dimension were tested through CFA using LISREL (version 8.8) (Jöreskog & Sörbom, 2006). The method of estimation was diagonally weighted least squares. This method was deemed appropriate for studies employing a Likert-type rating scale (Diamantopoulos & Sigauw, 2000; Steenkamp & Van Trijp, 1991). The CFA results

provided insight into model fit and evidence on items to be considered for deletion. Firstly, model fit was assessed through the examination of a combination of goodness-of-fit (GOF) measures, i.e. absolute and incremental fit indices. Table 2 provides a summary of these fit indices and also indicates the acceptable values used as guidelines for assessing GOF (adapted from Schlechter, 2005, p. 148).

TABLE 5
Factor loadings (55 items) - pilot study 2 (training data set)

DIMENSIONS	NO. OF ITEMS	ITEMS	FACTOR LOADINGS
Atmosphere	6	Item 2	0.639
		Item 3	0.647
		Item 4	0.718
		Item 5	0.685
		Item 6	0.726
		Item 11	0.565
Convenience	7	Item 12	0.613
		Item 19	0.621
		Item 22	0.613
		Item 32	0.644
		Item 38	0.664
		Item 41	0.617
		Item 46	0.607
Facilities	7	Item 53	0.689
		Item 56	0.690
		Item 57	0.701
		Item 60	0.710
		Item 75	0.634
		Item 76	0.625
		Item 80	0.620
Institutional	6	Item 95	0.649
		Item 96	0.691
		Item 98	0.673
		Item 100	0.642
		Item 101	0.698
		Item 102	0.628
Merchandise	8	Item 104	0.685
		Item 105	0.603
		Item 107	0.643
		Item 108	0.633
		Item 111	0.582
		Item 117	0.582
		Item 122	0.585
		Item 128	0.592
Promotion	8	Item 132	0.634
		Item 143	0.603
		Item 144	0.658
		Item 148	0.667
		Item 152	0.673
		Item 153	0.644
		Item 155	0.622
		Item 156	0.613
Sales personnel	5	Item 165	0.800
		Item 167	0.658
		Item 168	0.817
		Item 169	0.708
		Item 170	0.713
Service	8	Item 173	0.659
		Item 174	0.746
		Item 180	0.634
		Item 188	0.698
		Item 189	0.654
		Item 190	0.650
		Item 204	0.649
		Item 208	0.651

FIGURE 3
Simplified model of apparel store image – pilot study 2 (training data set)

TABLE 6
Reliability - pilot study 2 (training data set - 55-item store image scale)

CONSTRUCT	DIMENSIONS	NO. OF ITEMS	COEFFICIENT ALPHA
Store image		55	0.89
	Atmosphere	6	0.83
	Convenience	7	0.80
	Facilities	7	0.86
	Institutional	6	0.84
	Merchandise	8	0.83
	Promotion	8	0.84
	Sales personnel	5	0.88
	Service	8	0.86

It is critically important to examine parameter estimates in conjunction with model fit (Kelloway, 1998; MacCallum & Austin, 2000). Further scale purification was considered by investigating path estimates and standardised residuals to identify individual scale items for possible deletion. The cut-off value for completely standardised loadings was set at a minimum of > 0.5. The cut-off values for variance extracted (VE) and construct reliability (CR) were set at > 0.5 and > 0.7 respectively. Those items with standardised residuals of less than |2.5| were not considered for deletion. Where standardised residuals were between |2.5| and |4|, items were investigated but retained if there was no additional indication that these items should be deleted. Items with associated standardised residuals of higher than |4| were considered for deletion. These criteria were developed in accordance with recommendations by Diamantopoulos and Sigauw (2000) and Hair *et al.* (2006).

RESULTS AND DISCUSSION

Pilot study 1

Respondents (n = 89) were predominantly between 18 and 21 (82%). The majority belonged to the coloured population group (51%), followed by the black population group, while 56% indicated English as their home language, followed by isiXhosa and Afrikaans. Respondents bought clothes when needed or on a monthly basis, spending approximately R400 per month on clothing.

An investigation of the coefficient alphas of the dimensions revealed that Atmosphere (α = 0.57), Convenience (α = 0.61) and Sales personnel (α = 0.56) fell outside the set criteria of > 0.7. The subdimensions included in the Sales personnel dimension had high coefficient alphas and no items were identified for deletion. When considering the literature review, it is evident that the two dimensions Sales personnel and Service overlap with Employee service (Grace & O'Cass, 2005; Koo, 2003), Salespeople service (Kleinhans, 2003), Salesperson/ service (Manolis, Keep, Joyce & Lambert, 1994), Service – Sales associates attributes (Lee & Johnson, 1997) and Service – Store associates attributes (Lee & Johnson, 1997). Subsequently, it was decided to include the Interaction subdimension within the In-store service subdimension, since it could be justified as being conceptually related, as per literature recommendations (Blankson & Kalafatis, 2004; Parasuraman *et al.*, 1988).

The changes to the theoretical structure of the model of apparel store image (refer to Figure 3 Part 1) suggested by the statistical analysis were effected and are presented in Figure 2 (two subdimensions namely Transportation and Interaction were omitted from the Convenience and Sales personnel dimensions respectively). A 214-item store image scale was derived from the statistical analysis in the first pilot study. This scale was employed in the second pilot study.

TABLE 7
Reliability and item-total correlations - pilot study 2 (test data set)

CONSTRUCT	DIMENSIONS	ITEMS	NO. OF ITEMS	COEFFICIENT ALPHA	ALPHA IF DELETED	ITEM-TOTAL CORRELATIONS
Store image			55	0.83		
	Atmosphere		6	0.72		
		2			0.68	0.48
		3			0.67	0.49
		4			0.68	0.46
		5			0.67	0.51
		6			0.68	0.45
		11			0.71	0.37
	Convenience		7	0.62		
		12			0.57	0.37
		19			0.56	0.40
		22			0.57	0.38
		32			0.57	0.40
		38			0.58	0.35
		41			0.58	0.37
		46			0.66	0.09
	Facilities		7	0.61		
		53			0.59	0.26
		56			0.54	0.43
		57			0.49	0.62
		60			0.52	0.48
		75			0.57	0.32
		76			0.61	0.21
		80			0.66	0.09
	Institutional		6	0.69		
		95			0.66	0.42
		96			0.67	0.37
		98			0.64	0.47
		100			0.59	0.62
		101			0.66	0.41
		102			0.70	0.29
	Merchandise		8	0.64		
		104			0.64	0.18
		105			0.65	0.18
		107			0.57	0.45
		108			0.52	0.59
		111			0.60	0.37
		117			0.58	0.47
		122			0.62	0.28
		128			0.65	0.15
	Promotion		8	0.68		
		132			0.65	0.40
		143			0.71	0.14
		144			0.64	0.45
		148			0.66	0.39
		152			0.66	0.36
		153			0.65	0.41
		155			0.62	0.53
		156			0.64	0.44
	Sales personnel		5	0.79		
		165			0.80	0.41
		167			0.80	0.40
		168			0.70	0.72
		169			0.73	0.65
		170			0.71	0.70
	Service		8	0.59		
		173			0.55	0.32
		174			0.51	0.46
		180			0.59	0.18
		188			0.51	0.41
		189			0.56	0.26
		190			0.52	0.41
		204			0.56	0.27
		208			0.61	0.12

Pilot study 2

The sample profile (n = 176) was similar to that of pilot study 1. Most of the respondents (93%) were in the age group 18 to 21. The majority (81%) belonged to the white population group and indicated Afrikaans as their home language.

The data obtained were split into a training data set (n = 110) and a test data set (n = 66). Once again, the statistical analysis was only performed on each dimension associated with the store image construct. Therefore, the model was adapted to exclude all the subdimensions that focus on the broad dimensions of store image. This model is represented in Figure 3 and was employed in all further statistical analysis.

Training data set (n = 110)

Coefficient alphas for each dimension and the total scale are presented in Table 4. All item-total correlations met the adopted criteria of > 0.3. Inter-item correlations were within the set criteria of 0.2-0.5, except for the Sales personnel dimension at 0.62.

EFA was performed employing the principal axis factoring procedure and constraining the analysis to *a priori* one factor for each dimension. The factor loadings for items in the 214-item scale ranged from 0.41 to 0.72 for Atmosphere, 0.36 to 0.66 for Convenience, 0.28 to 0.71 for Facilities, 0.20 to 0.69 for Institutional, 0.36 to 0.68 for Merchandise, 0.43 to 0.67 for Promotion, 0.44 to 0.81 for Sales personnel and 0.39 to 0.74 for Service. Given that the cut-off value for factor loadings was set at a minimum of > 0.5, the results highlighted that individual items required closer scrutiny.

Item reduction was undertaken by considering item factor loadings in conjunction with item-total correlations. Aiming at scale purification and optimal scale length, items with the highest factor loadings and corresponding high item-total correlations were retained. This resulted in the deletion of 159 items across all dimensions and 55 items being retained. Table 5 presents the factor loadings of the individual items retained in the 55-item scale.

Coefficient alphas, inter-item correlations and item-total correlations were again calculated for the 55-item scale. Coefficient alpha for the total scale was recorded at 0.89 and ranged from 0.80 to 0.88 for the individual dimensions, thus exceeding the cut-off value of > 0.7, as presented in Table 6. All alpha values were lower for the shortened scale compared to the 214-item scale, except for Sales personnel. This is to be expected, since alpha increases with the number of items (Netemeyer *et al.*, 2003). The inter-item correlations were all

within the adopted criterion ranging from 0.2–0.5 and the item-total correlations were all above the cut-off value of > 0.3.

A correlation analysis was performed between the 214-item and the 55-item store image scales. The correlations between the various dimensions indicated satisfactory values, namely Atmosphere (r = 0.94), Convenience (r = 0.90), Facilities (r = 0.90), Institutional (r = 0.87), Merchandise (r = 0.90), Promotion (r = .90), Sales personnel (r = 0.94) and Service (r = 0.92) (see Section 3.4.2.3). The 214-item store image scale was representative of all the subdimensions initially proposed in the revised model of apparel store image (Figure 2). The 55-item scale did not represent all of these subdimensions but only the broad dimensions of store image. The high correlations between the longer and shorter store image scales provide support for the shortened version and confirm that the 55-item scale performs satisfactorily.

Test data set (n = 66)

Coefficient alphas for the 55-item scale were recorded at 0.83 for the total scale and ranged from 0.59 to 0.79 for the individual dimensions with only Atmosphere (α = 0.72) and Sales personnel (α = 0.79) satisfying the accepted cut-off value of > 0.7 (refer to Table 7).

The inter-item correlations of Convenience (0.192), Merchandise (0.192) and Service (0.170) fell outside the set criterion ranging from 0.2 to 0.5 with 14 individual items not achieving the set cut-off value of > 0.3, namely Convenience (item 46), Facilities (items 53, 76 and 80), Institutional (item 102), Merchandise (items 104, 105, 122 and 128), Promotion (item 143) and Service (items 180, 189, 204 and 208).

CFA was performed for each dimension of the test data set using diagonally weighted least squares as the method of estimation. Table 8 summarises the indices of model fit for each of the dimensions. Root mean square error of approximation (RMSEA) for Convenience (0.0) and Promotion (0.034) demonstrates good fit, whilst Institutional (0.054) meets the set criteria for acceptable fit. The RMSEA values for Atmosphere (0.096), Facilities (0.12), Merchandise (0.16), Sales personnel (0.12) and Service (0.16) fall outside of the set criteria for acceptable fit. The Standardised root mean residual (RMR) value for all dimensions exceeded 0.05, indicating poor fit. The Goodness-of-fit index (GFI) value for all the dimensions exceeded the cut-off value of 0.9, suggesting good model fit. The GFI value for all the dimensions exceeded the cut-off value of 0.9, suggesting good model fit. The Adjusted goodness-of-fit index (AGFI) criteria of > 0.9 for all the dimensions were met except for Merchandise (0.85) and Service (0.81). All the dimensions met

TABLE 8
Model fit indices of CFA on individual dimensions - pilot study 2 (test data set)

MODEL FIT INDICES	ATMOSPHERE	CONVENIENCE	FACILITIES	INSTITUTIONAL	MERCHANDISE	PROMOTION	SALES PERSONNEL	SERVICE
Absolute Fit Measure								
Degrees of Freedom	9	14	14	9	20	20	5	20
Normal Theory Weighted Least Squares Chi-Square	40.02 p < 0.01	23.87 p < 0.05	53.20 p < 0.01	23.35 p < 0.01	99.72 p < 0.01	41.04 p < 0.01	24.12 p < 0.01	96.17 p < 0.01
Root Mean Square Error of Approximation (RMSEA)	0.096	0.0	0.12	0.054	0.16	0.034	0.12	0.16
Standardised Root Mean Square Residual (RMR)	0.10	0.078	0.12	0.093	0.16	0.093	0.084	0.15
Goodness of Fit Index (GFI)	0.97	0.97	0.95	0.97	0.92	0.96	0.99	0.90
Adjusted Goodness of Fit Index (AGFI)	0.92	0.95	0.90	0.93	0.85	0.93	0.97	0.81
Incremental Fit Measures								
Non-Normed Fit Index (NNFI)	0.93	1.00	0.86	0.97	0.68	0.99	0.94	0.65
Comparative Fit Index (CFI)	0.96	1.00	0.91	0.98	0.77	0.99	0.97	0.75

the cut-off value of 0.9 on the Non-normed fit index (NNFI) except for Facilities (0.86), Merchandise (0.68) and Service (0.65). The Comparitive fit index (CFI) values for all dimensions exceed 0.9 and demonstrate good fit. The set criterion of > 0.9 for the CFI measure was met by all dimensions except for Merchandise (0.77) and Service (0.75).

Results from the absolute fit measures indicated that the model did not adequately reproduce the observed data. None of the dimensions met the set criteria for the normal theory weighted least chi-square statistic or the standardised RMR. Only the Convenience, Promotion and Institutional dimensions indicated acceptable fit based on the RMSEA measure. All dimensions met the set criteria for the GFI and AGFI measure, except for Merchandise and Service, which did not meet the set parameter for the AGFI index. All the dimensions met the set criteria for the incremental fit measures except for Facilities, Merchandise and Service. By implication, the specified models for all dimensions apart from Facilities, Merchandise and Service provided a better fit than the null model. Overall, however, the CFA results do not support adequate model fit.

Subsequently, the path estimates and standardised residuals of individual items were considered to identify those items contributing to the poor model fit. These items should be considered for deletion to further purify the store image scale. The cut-off values are indicated in Table 9 based on recommendations by Hair *et al.*, (2006). The parameter for item deletion based on the standardised residuals was set at higher than |4|. Items with standardised residuals between |2.5| and |4| were considered for deletion only if there was additional support for their deletion.

Atmosphere: Item 11 should be considered for deletion. The standardised residual between items 5 and 6 (3.44) exceeded the cut-off value of > |2.5|. However, no further support for the deletion of these items was recorded and since these items did not exceed the cut-off value of > |4|, they should be retained.

Convenience: The deletion of item 46 is supported by its item-total correlation not meeting the set criterion of > 0.3 (refer to Table 7). None of the standardised residuals exceeded the cut-off value of > |2.5|.

Facilities: Four of the eight items (53, 75, 76 and 80) should be considered for deletion. The item-total correlations for items 53, 76 and 80 did not meet the set criterion of > 0.3 and provide further support for their deletion. The standardised residual between items 75 and 76 was 3.25. This exceeded the cut-off value of > |2.5|. Although this standardised residual did not exceed the higher cut-off value of > |4|, the deletion of this item was supported by additional evidence from the completely standardised loadings and inter-item correlations.

Institutional: Completely standardised loadings for items 96 and 102 can be deleted. Additional support for the deletion of item 102 is provided by the item-total correlation not meeting the set criterion of > 0.3. The standardised residual between items 95 and 96 (2.58) exceeded the cut-off value of > |2.5|. This provides further support for the deletion of item 96.

Merchandise: The deletion of items 104, 105, 122 and 128 is supported by their item-total correlations not meeting the set criterion of > 0.3. The standardised residuals between items 108 and 111 (3.53) and items 117 and 128 (2.61) exceed the cut-off value of > |2.5|. The deletion of items 108 and 111 is not supported by any other results, and therefore these items should be retained. However, item 128 is not only associated with a high standardised residual but its completely standardised loading and item-total correlation further support its deletion. The standardised residual between items 107 and 108 (4.92) exceeded the adopted criterion of > |4|, suggesting that either one of these items should be deleted. Since item 108 also shares

a standardised residual of > |2.5| with item 111, this item should be considered for deletion.

TABLE 9
Summary: completely standardised loadings, VE and CR – pilot study 2 (test data set)

DIMENSION	CRITERIA FOR DELETION OF ITEMS AND EVALUATION OF DIMENSIONS		
	Item number	Completely standardised loading < 0.5	Variance extracted (VE) < 0.5	Construct reliability (CR) < 0.7
Atmosphere	2	0.63	0.39	0.79
	3	0.66		
	4	0.64		
	5	0.68		
	6	0.64		
	11	0.46		
Convenience	12	0.54	0.28	0.71
	19	0.67		
	22	0.50		
	32	0.56		
	38	0.53		
	41	0.60		
	46	0.14		
Facilities	53	0.42	0.34	0.75
	56	0.75		
	57	0.83		
	60	0.72		
	75	0.47		
	76	0.36		
	80	0.22		
Institutional	95	0.52	0.36	0.76
	96	0.46		
	98	0.67		
	100	0.81		
	101	0.60		
	102	0.43		
Merchandise	104	0.26	0.33	0.79
	105	0.28		
	107	0.71		
	108	0.94		
	111	0.62		
	117	0.52		
	122	0.35		
	128	0.23		
Promotion	132	0.56	0.33	0.78
	143	0.20		
	144	0.49		
	148	0.52		
	152	0.53		
	153	0.69		
	155	0.76		
	156	0.66		
Sales personnel	165	0.46	0.52	0.83
	167	0.44		
	168	0.93		
	169	0.85		
	170	0.78		
Service	173	0.61	0.26	0.70
	174	0.81		
	180	0.47		
	188	0.54		
	189	0.30		
	190	0.59		
	204	0.26		
	208	0.12		

Values not meeting the criteria are highlighted

Promotion: Items 143 and 144 should therefore be considered for deletion. Item 143 also had an item-total correlation that did not exceed the cut-off value of > 0.3, thus providing further support for its deletion. None of the standardised residuals exceeded the cut-off value of > |2.5|.

Sales personnel: Two of the five items for this dimension (165 and 167) did not meet the > 0.5 cut-off value and should be considered for deletion. None of the standardised residuals exceeded the criterion of > |2.5|.

Service: Four items namely 180, 189, 204 and 208 should be deleted based on the completely standardised loadings. This is supported by their item-total correlations not meeting the set criterion of > 0.3.

CONCLUDING REMARKS

The results from the item-total correlations, path estimates and standardised residuals provided support for the deletion of 20 items. In addition, the VE for all dimensions did not meet the set criterion. By implication, a higher amount of variance in the items was captured by measurement error compared to the underlying dimension. This result further supports the deletion of these items. However, all dimensions met the cut-off value for CR. This indicated that the items provide a reliable measurement of each dimension. The deletion of the suggested items should improve the model fit of the individual dimensions and further purify the store image scale.

However, the CFA results raised concerns that needed to be addressed before the next phase in the study. Firstly, the deletion of the items would lead to the Facilities, Merchandise and Sales personnel dimensions being under-identified, i.e. there would be less than four measurement items associated with these dimensions to allow for model identification. Secondly, the small sample size (n = 66) cast doubt on the CFA results, since the literature recommends sample sizes of 100 and more (Hair *et al.*, 2006). Therefore, the decision was made to retain the 55-item scale for the next phase in the scale development process.

In the following article, Phase 4 of this study will be presented. The 55-item Apparel Store Image Scale derived from the two pilot studies will be employed in the fourth phase. The reliability and validity of the Apparel Store Image Scale will be assessed through practical implementation in an apparel retail environment.

REFERENCES

Amirani, S. & Gates, R. (1993). An attribute-anchored conjoint approach to measuring store image. *International Journal of Retail and Distribution Management*, 21(5), 30–39.

Bearden, W.O. (2001). Consumer self-confidence: refinements in conceptualization and measurement. *Journal of Consumer Research*, 28(1), 121–135.

Birtwistle, G. & Siddiqui, N. (1995). Store image – characteristics for menswear fashion retailers. *The Home Economist*, 14(6), 20–22.

Blankson, C. & Kalafatis, S.P. (2004). The development and validation of a scale measuring consumer/customer-derived generic typology of positioning strategies. *Journal of Marketing Management*, 20, 5–43.

Bradburn, N., Sudman, S. & Wansink, B. (2004). *Asking questions: the definitive guide to questionnaire design – for market research, political polls, and social and health questionnaires*. San Francisco: Jossey-Bass.

Brace, I. (2004). *Questionnaire design: how to plan, structure and write survey material for effective market research*. London: Kogan Page.

Churchill, G.A. Jr. (1979). A paradigm for developing better measures of marketing construct. *Journal of Marketing Research*, 16, 64–73.

Churchill, G.A. Jr. & Iacobucci, D. (2005). *Marketing research: methodological foundations* (9th Ed.). Mason:Thomson Southwestern.

Dabholkar, P.A., Thorpe, D.I. & Rentz, J.O. (1995). A measure of service quality for retail stores: scale development and validation. *Journal of the Academy of Marketing Science*, 24(1), 3–16.

DeVellis, R.F. (2003). *Scale development: theory and applications* (2nd Ed.). Thousand Oaks: Sage Publications.

Dhurup, M., Venter, P.F. & Oosthuyzen, A. (2005). A factor analytical service quality measurement scale for supermarkets in South Africa. *South African Journal of Economic and Management Science*, 8(2), 140–153.

Diamantopoulos, A. & Sigauw, J.A. (2000). *Introducing LISREL: a guide for the uninitiated*. London: Sage Publications.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C. & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299.

Frazer, L. & Lawley, M. (2002). *Questionnaire design and administration*. Brisbane: John Wiley.

Gorsuch, R.L. (1997). Exploratory factor analysis: its role in item analysis. *Journal of Personality Assessment*, 67, 532–560.

Grace, D. (2005). Consumer disposition toward satisfaction (CDS): scale development and validation. *Journal of Marketing Theory and Practice*, 13(2), 20–31.

Grace, D. & O'Cass, A. (2005). An examination of the antecedents of repatronage intentions across different store formats. *Journal of Retailing and Consumer Services*, 12(4), 227–243.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. & Tatham, R.L. (2006). *Multivariate data analysis*. (6th Ed.). Upper Saddle River: Pearson/Prentice Hall.

Janse van Noordwyk, H.S. (2002). Perceived importance of retail store image attributes to the female large-size apparel consumer in a multicultural society. Unpublished master's thesis. Stellenbosch: Stellenbosch University.

Jöreskog, K. & Sörbom, D. (2006). *LISREL 8.8*. Chicago, IL: Scientific Software International Inc.

Kelloway, E.K. (1998). *Using LISREL for structural equation modelling: a researcher's guide*. Thousand Oaks: Sage Publications.

Kerlinger, F.N. & Lee, H.B. (2000). *Foundations of behavioural research* (4th Ed.). United States, Wadsworth.

Kleinhans, E.H. (2003). Black female student consumers' perception of clothing store image attributes. Unpublished masters thesis. Stellenbosch: Stellenbosch University.

Koo, D. (2003). Inter-relationships among store images, store satisfaction, and store loyalty among Korea Discount Retail patrons. *Asia Pacific Journal of Marketing and Logistics*, 15(4), 42–71.

Lastovicka, J.L., Bettencourt, L.A., Hughner, R.S. & Kuntze, R.J. (1999). Lifestyle of the tight and frugal: theory and measurement. *Journal of Consumer Research*, 26, 85–98.

Lee, M., & Johnson, K.K.P. (1997). Customer expectations for service at apparel retail outlets. *Journal of Family and Consumer Sciences*, Winter, 26–30.

Li, H., Edwards, S.M. & Lee, J. (2002). Measuring the intrusiveness of advertisements: scale development and validation. *Journal of Advertising*, 31(2), 37–47.

Lichtenstein, D.R., Ridgway, N.M. & Netemeyer, G. (1993). Price perceptions and consumer shopping behaviour: a field study. *Journal of Marketing Research*, 30(2), 234–245.

Lindquist, J.D. (1974–1975). Meaning of image. *Journal of Retailing*, 50(4), 29–38.

Little, T.D., Lindenberger, U. & Nesselroade, J.R. (1999). On selecting indicators for multivariate measurement and modelling with latent variables: when "good" indicators are bad and "bad" indicators are good. *Psychological Methods*, 4(2), 192–211.

MacCallum, R.C. & Austin, J.T. (2000). Applications of structural equation modelling in psychological research. *Annual Review of Psychology*, 51, 201–226.

Manolis, C., Keep, W.W., Joyce, M.L. & Lambert, D.R. (1994). Testing the underlying structure of store image scale. *Educational and Psychological Measurement*, 54(3), 628–645.

Martineau, P. (1958). The personality of the retail store. *Harvard Business Review*, 36, 47–55.

Netemeyer, R.G., Bearden, W.O. & Sharma, S. (2003). *Scaling procedures: issues and applications*. Thousand Oaks: Sage Publications.

Nunnally, J.C. (1978). *Psychometric theory*. London: McGraw-Hill Book Company.

Oppenheim, A.N. (1992). *Questionnaire design, interviewing and attitude measurement*. London: Continuum.

Parasuraman, A., Zeithaml, V.A. & Berry, L.L. (1988). SERVQUAL: a multi-item scale for measure consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12–37.

Schlechter, A.F. (2005). The influence of transformational leadership, emotional intelligence, trust, meaning and intention to quit on organisational citizenship behaviour. Unpublished doctoral thesis. Stellenbosch: Stellenbosch University.

Shimp, T.A. & Sharma, S. (1987). Consumer ethnocentrism: construction and validation of the CETSCALE. *Journal of Marketing Research*, 24(3), 280–289.

StatSoft Inc. (2007). STATISTICA 8.0. [computer software]. www.statsoft.com.

Steenkamp, J.E.M. & Van Trijp, H.C.M. (1991). The use of LISREL in validating marketing constructs. *International Journal of Research in Marketing*, 8, 283–299.

Synodinos, N.E. (2003). The "art" of questionnaire construction: some important considerations for manufacturing studies. *Integrated Manufacturing Systems*, 14(3), 221–237.

Tabachnick, B.G. & Fidell, L.S. (2007). *Using multivariate statistics* (5th Ed.). United States of America: Pearson Allyn and Bacon.

Terblanché, N.S. & Boshoff, C. (2004). The in-store shopping experience: a comparative study of supermarket and clothing store customers. *South African Journal of Business Management*, 35(4), 1–9.

Thompson, K.E. & Chen, Y.L. (1998). Retail store image: a means-end approach. *Journal of Marketing Practice*, 4(6), 161–173.

Venter, P.F. & Dhurup, M. (2005). Consumer perceptions of supermarket service quality: scale development and validation. *South African Journal of Economic and Management Sciences*, 4, 424–436.

Zimmer, M.R. & Golden, L.L. (1988). Impressions of retail stores: a content analysis of consumer images. *Journal of Retailing*, 64(3), 265–293.

APPENDIX 1
Examples of items: apparel store image scale

ATMOSPHERE
1. fashionability of store interior
3. attractiveness of décor in store

CONVENIENCE
9. accessibility of store (e.g. location within mall)
12. ease of finding merchandise items

FACILITIES
14. accessibility of store entrance/exit
20. ease of shopping with family in mall where store is situated

INSTITUTIONAL
22. store's appeal to my friends
25. similarity between store image and self image

MERCHANDISE
28. availability of imported merchandise
33. availability of styles suitable to my age

PROMOTION
38. spaciousness of in-store displays
40. sales with marked-down prices

SALES PERSONNEL
43. fashionability of sales personnel
46. similarity in age between sales personnel and myself

SERVICE
49. courteousness of sales personnel
54. availability of mail-order service