

HOFSTEDE'S VSM-94 REVISITED: IS IT RELIABLE AND VALID?

TERESA KRUGER
GERT ROODT

*Department of Human Resource Management
Rand Afrikaans University*

ABSTRACT

The main objective of this study was to investigate the metric properties of the Hofstede Value Survey Module-94 (VSM-94). The questionnaire was evaluated against criteria for test construction. The VSM-94 consisting of 20 items was administered to 231 female managers in a large telecommunications organisation. The inter-correlations of the item scores were empirically investigated. Anti-image inter-correlations were executed on the 20 item-scores. Items with low measures of sampling adequacy (MSA) were omitted and a factor analysis and an anti-image inter-correlation conducted on the remaining eight items. Uniformly item-response distributions and low item inter-correlations resulted in poor reliability coefficients on the final factors, which suggests that the VSM-94 is probably not suitable for use in the South African context.

OPSOMMING

Die hoofdoel van die studie was om die metriese eienskappe van die Hofstede Value Survey Module-94 (VSM-94) te ondersoek. Die vraelys is teen kriteria vir toetskonstruksie geëvalueer. Die VSM-94 bestaande uit 20 items, is toegepas op 231 vroulike bestuurders in 'n groot telekommunikasie-organisasie. Die interkorrelasies van itemtellings is empiries ondersoek. Teenbeeld interkorrelasies is op die 20 item-tellings uitgevoer. Items met lae maatstawwe van steekproeftoereikendheid (MSA) is uitgeskakel en 'n faktorontleding en teenbeeld interkorrelasie is uitgevoer op die oorblywende agt items. Eenvormige item-responsverdelings en lae item-interkorrelasies het swak betroubaarheidskoëffisiënte op die finale faktore tot gevolg wat suggereer dat die VSM-94 waarskynlik nie geskik is vir gebruik in die Suid-Afrikaanse konteks nie.

An infinitive number of studies have been conducted using the Value Survey Module developed by Geert Hofstede measuring cultural values according to five dimensions. Hofstede's groundbreaking research is also widely reported and recognised (Gibson, Ivancevich & Donnelly, 1994; Hellriegel, Slocum & Woodman, 1998; Ivancevich & Matteson, 1996; Nelson & Quick, 1997; Robbins, 1996) in measuring differences between national cultures as well as value differences between ethnic groups. The original objective of the study was to investigate the influence of cultural values on female leadership behaviour. For this purpose the Value Survey Module 1994 (VSM-94) was used to measure cultural values and the Multiple Factor Leadership Questionnaire (MLQ-5) was used to assess leadership behaviour. However, early results have raised suspicion regarding the construction of the VSM-94 itself. This has prompted the researchers to re-direct their focus towards the metric properties of the VSM-94.

Few people would question Hofstede's work and the value he has added to the field of cultural studies, cross-cultural research and the development of cross-cultural theory that were needed to fill gaps in knowledge. It can not be assumed that instruments with cross-cultural equivalence in one country would also be effective in another culture (*cf.* Berry & Triandis, 1980; Roodt, Kinnear, Erwee, Smith, Lynch & Millet, 2001). The fact remains that much of the empirical research that is being reported in South Africa (e.g. Thomas & Bendixen, 2000; Newman & Nollen, 1996) show no evidence of validations done on the VSM-94. The manual accompanying the survey contains no information on the reliability or validity of the instrument either.

The objective of the current study is therefore to evaluate the metric properties of the VSM-94 against accepted criteria of test construction. The wording of the items, as well as the item response format, is scrutinised in this investigation.

Culture

Hofstede (1993) found that unique cultural characteristics such as language, beliefs, values, religion, and social organisation are

generally presumed to necessitate distinct actions in different cultural settings. Culture is therefore a complex web of norms, values, assumptions, attitudes, and beliefs that are characteristic of a particular group (Kuchinke, 1999), implying that one way of acting is preferable to another. In such, cultural differences are an important consideration in understanding human behaviour.

Hofstede (1983) identified five dimensions of cultural variability that provide broad explanations for differences between cultures. The fifth dimension was added only later in his work.

Power distance.

'Power distance' is defined as the extent to which less powerful members of a group accept and expect that power be distributed unequally between leaders and followers. This dimension deals directly with people's belief about the unequal distribution of power and status as the proper way for organising social systems (Newman & Nollen, 1996). Items 3, 6, 14 and 17 measure this dimension.

Uncertainty avoidance.

'Uncertainty avoidance' indicates the extent to which a culture programs its members to feel either uncomfortable or comfortable in unstructured situations (Elenkov, 1998). It refers to the degree to which members of a culture are threatened by uncertainty. These cultures will ultimately find ambiguity and changes stressful. Cultures high on 'uncertainty avoidance' have clear norms and people behave in accordance with these norms (Offermann & Hellmann, 1997). Items 13, 16, 18 and 19 measure this sub-scale.

Newman and Nollen (1996) identified this dimension as the most problematic of Hofstede's dimensions arguing that 'uncertainty avoidance' is an artefact of the time in which it was developed, namely the late 1960s and early 1970s. The cold war in Europe during this period may have resulted in higher 'uncertainty avoidance' scores for the European sample. It most probably also had an influence on the overall results of Hofstede's study.

Individualism-collectivism.

This dimension distinguishes between cultures that value individual effort as opposed to collective team effort.

'Individualism-collectivism' indicates the extent to which identity is developed by the self or collectively. Bchner and Hesketh (1994) found that individualistic cultures reported less likelihood of working in a team compared to people from collectivist cultures. The items measuring this dimension are items 1,2,4 and 8.

Masculinity.

'Masculinity' refers to cultures in which assertiveness, challenge and ambition are highly valued as opposed to feminine cultures where the emphasis is placed on good working conditions and relationships. This dimension is measured by items 5, 7, 15 and 20 on the VSM-94. Competitive orientation vary between these two extremes, where masculine cultures have an assertive and highly competitive orientation and feminine cultures are non-assertive and caring towards others (Elenkov, 1998). In masculine cultures members of society would act and acquire, rather than think and observe.

Long-term orientation.

This fifth dimension refers to a person's outlook on the future. It emphasises the degree to which a group is orientated towards 'long-term' results rather than short-term gratification (Kuchinke, 1999). Cultures with high scores on items 9 to 12 are characterised by patience, perseverance, and a respect for tradition (Hofstede, 1991).

Hofstede's dimensions could conceptually be linked to variables identified in other studies (Theron, 1993). Hofstede and Bond (1984) reported 'uncertainty avoidance' to be inadequate in discriminating amongst cultures while Newman and Nollen (1996) found the same results in an earlier study while investigating the fit between management practices and national culture in South Africa. Kuchinke's (1999) findings show that culture measured through Hofstede's framework might not be such a good predictor of other variables. In this regard Kuchinke acknowledged that the VSM-94 is not a valid instrument in all circumstances. Two of the dimensions namely 'power distance' and 'uncertainty avoidance' failed reliability tests while 'masculinity' and 'long-term orientation' yielded results that differ significantly from the original research done by Hofstede. Thomas and Bendixen (2000) did not report any reliability in their study using the VSM-94 on a South African sample. They however found that the instrument is inadequate in distinguishing between individualism and communalism and reported a similarity between the cultural groups they measured on most of Hofstede's dimensions. The following are possible explanations for the findings: that South Africa's global culture is stronger than any of the indigenous sub-cultures; that their sample of 586 was too small; or that the instrument used does not adequately distinguish between these groups. The latter explanation seems to be the most valid (especially in the light of current findings).

The most relevant international study that evaluated the metric properties of the VSM-94 came from Spector, Cooper and Sparks (2001). These authors used a sample of 6,737 participants across 23 countries. On both English and translated versions of the instrument, internal consistency was poor and the factors extracted differed significantly from Hofstede's work.

Cross cultural research

Elenkov (1998) identified one of the biggest obstacles in cross-cultural research as the transferability of these studies to other cultures. Individuals come from different cultural groups, which also affects their mindset and framework (Machet, 1996) and they would therefore interpret stimuli in different ways. This variance in interpretation could have a significant influence on the results. In this regard Berry and Triandis (1980) argued that it should be possible to compare two groups on a single dimension, meaning they should have some common features or equivalence. The groups should however also show differences on the same dimension in order to make comparisons between them. These two authors have identified three kinds of equivalence, all of which need to be present in order to have construct validity.

Functional equivalence exists "when two or more behaviours are related to functionally similar problems" (Berry & Triandis, 1980, p. 11). These problems (or factors) need to transcend cultural boundaries in order for the researcher to make inferences about the construct being measured. If these behaviour indicators vary in the same way across all groups, it provides a basis for comparison. This platform for comparison allows researchers to apply instruments across various groups. It would be of no use to measure a construct between groups if it only applies to one particular group. The authors argue that functional equivalence is a pre-existing phenomenon and cannot be created or manipulated.

Conceptual equivalence lies in the common meaning of stimuli, concepts or behaviours and is also a pre-condition for comparison. *Translation equivalence* is of importance where an existing research instrument is translated by using a bilingual translator. The translator would translate the instrument into the new language and then reverse the translation into the original language (Greer & Greer, 1998). Differences can occur because individuals speaking the same language can easily misunderstand each other. Any anomalies will show support for conceptual non-equivalence. To use an example, the word *thrift* on the VSM-94 needs to have a common meaning and be commonly understood by all individuals being measured in order to establish whether there are reliable value differences between the cultural groups. *Semantic equivalence* uses a bipolar adjective scale to indicate the meaning of a concept across languages. Maintaining semantic equivalence is one of the greatest problems in cross-cultural research (Bedell, Van Eeden & Van Staden, 1999). Another problem regarding language is that although an important language may be used in business, it might not be the overwhelming language used by the population. This is true for South Africa with 11 official languages. Although English is spoken in the business world, other languages could take preference in the social network, contributing to possible discrepancies in understanding the same language.

Metric equivalence considers the psychometric properties of sets of data, which should reflect the same structure. Essentially this means that measuring instruments must be structured in similar ways within one group in order to make valuable inter-group comparisons (Berry & Triandis, 1980). In analysing score comparability it is important that the same construct is measured across the different groups (Bedell et al., 1999). Inter-group differences on a test must therefore reflect real difference on the measured construct, excluding factors relevant to the situation or other factors pertaining to the test.

Survey construction

Sound questionnaire construction is essential in finding valid and reliable data that can be used in research. Not all instruments can be applied equally successfully across cultural groups. Various essential criteria exist that any instrument should adhere to, of which only the most relevant will be discussed (cf. Schepers, 1992).

Define the construct.

The construct should be clearly defined and the developer must know exactly what the test should measure and how the instrument is to differ from others (Gregory, 1996). In the case of the VSM-94, values need to be conceptualised. Values are basic convictions that a specific mode of conduct is socially preferred to an opposite mode (Hofstede, 1991). It is very important here that the researcher explicitly explains the purpose of the instrument before starting the construction of the items.

Identify the domain.

The domain of an instrument is linked to the construct (Swart, Roodt & Schepers, 1999) since it provides a framework of the construct being measured. In the case of the VSM-94 the domain is cultural values. Hofstede (1980a) defined culture as "the collective mental programming of people in an environment".

Identify sub-domains.

Identifying sub-domains signifies that the test developer needs to find all the elements underpinning the construct (Schepers, 1992). Through extensive analysis Hofstede arrived at the five sub-domains of cultural values, referred to as dimensions of culture, namely 'power distance', 'individualism', 'masculinity', 'uncertainty avoidance', and 'long-term orientation'.

Identify behaviour indicators.

Swart et al. (1999) see this phase as the operationalisation of the sub-domains into observable and quantifiable behaviour. Through numerous cultural studies, Hofstede has managed to assign indicators to each of the five sub-domains. The 'power distance' dimension would include behaviour such as having good relationships with others (item 3), being consulted by superiors (item 6), being afraid to express disagreement (item 14) and working with more than one superior (item 17). The 'individualism' index underlines indicators such as more time for personal and family life (item 1), good working conditions (item 2), security of employment (item 4) and adventure in the job (item 8). Co-operation (item 5), job advancement opportunities (item 7), trust (item 15) and accountability for failure (item 20) are all associated with the 'masculinity' sub-domain. The 'uncertainty avoidance' dimension highlights behaviour such as the harmful attitude towards competition (item 18), adhering to work rules (item 19), indeterminate management (item 16) and anxiety at work (item 13). Behaviour that is reflected in the 'long-term orientation' sub-domain includes personal stability (item 9), thrift (item 10), perseverance (item 11) and respecting tradition (item 12).

Item construction.

A test as a whole is only as good as the items included in the test. The choice and number of items to include as well as the item format is crucial to the validity and reliability of the instrument. According to Schepers (1992) value scales, like the VSM, usually take on two formats, either questions or statements. Items 1 to 12 on the VSM-94 state the question in the format of "how important is it to..." followed by the item content. Questions 13 and 14 are asked as individual questions with response scales different from those in item 1 to 12.

Statements, like items 15 through to 20, can be positively or negatively stated and are usually responded to on a Likert scale which present the respondent with options ranging from agree/disagree or approve/disapprove (Gregory, 1996). Likert scales pose some problems and Schepers (1992) argues that the equal-interval quality of this scale declines when more than two of the points on the scale are anchored. Another problem is that statements with a strong positive or negative connotation will be endorsed without evaluating the content of the statement. These statements would not only increase response bias, but Swart et al. (1999) mentioned that item response distributions might tend towards a bimodal curve.

The objectives of the current study is therefore to evaluate the VSM-94 by focusing on:

- criteria for test construction
- item and response format
- differential item skewness
- item inter-correlations

METHOD

Sample

To eliminate the effects of organisational culture in this study only one company was targeted and the questionnaire was sent to the entire population of female managers in a large telecommunications institution. The sample included women from various cultural groups including Zulu, Indian, Afrikaans, English, Sotho, Xhosa and Coloured South Africans. Of the 461 questionnaires that were sent out, 231 were returned, yielding a response rate of 50%.

TABLE 1
BIOGRAPHICAL CHARACTERISTICS OF THE SAMPLE

Variable	Frequency	Percentage
Mother tongue		
Zulu	22	9.6
Jewish	1	.4
Xhosa	16	7.0
Afrikaans	97	42.5
Indian	25	11.0
Sotho	20	8.8
English	31	13.6
Other	16	7.0
Missing Values	3	
Total	231	100.0
Age		
24-35 Years	129	55.8
36-45 Years	70	30.3
+ 45 years	31	13.4
Missing Values	1	.4
Total	231	100.0
Marital status		
Single – never married	56	24.2
Married	135	58.4
Divorced	23	10.0
Missing Values	17	7.4
Total	231	100.0
Education		
No tertiary education	52	22.5
Tertiary education	163	70.6
Missing Values	16	6.9
Total	231	100.0

The biographical characteristics of the sample are described in Table 1. From this data it can be seen that the majority of the sample were White, Afrikaans and a large proportion of the sample were between 24 and 35 years of age.

Instruments

Hofstede developed the original Value Survey Module in a major United States-based organisation using 116 000 participants in 40 countries (Theron, 1993). The VSM-82 with its 47 questions was extensively analysed through factor analysis and yielded the four dimensions of 'individualism', 'power distance', 'uncertainty avoidance', and 'masculinity'. Theron's (1993) validation of the earlier VSM by means of factor analysis produced completely different results in South Africa. Only two factors with eigenvalues greater than one according to Kaiser's (1961) criterion were obtained by using principal axis factoring with varimax rotation. However, when the data was subjected to rotated factoring four factors were obtained. Factors 2,3 and 4 had inadequate item loadings. Factor 1 therefore accounted for the majority of the items. Theron (1993) tested the reliability of the early VSM by means of the split-half method with Spearman-Brown correction, which yielded a reliability coefficient of 0,88 for the unequal lengths. An alpha coefficient of 0,90 was also obtained.

The VSM-94 has been modified and contains only 20 items and six demographic questions. This shortened questionnaire does have implications for reliability because a reduction in items shortens the measure, compromising reliability (Theron, 1993). All items on the VSM-94 use a five-point Likert scale of which all the points are anchored.

The shortened version of the Multifactor Leadership Questionnaire (MLQ-5) with 45 questions was also used since

the initial objective of the study was to investigate the influence of cultural values on leadership behaviour. The response scale varies on a five-point scale from "frequently, if not always" to "not at all". The MLQ-5 proved to be a good comparative instrument used on the same sample.

Procedure

To increase the number of responses on the questionnaire, it was distributed via electronic mail to all the participants. The sample was dispersed across the country and it would have been difficult administering the questionnaire directly. To further increase the response rate, a cover letter was prepared, stating the purpose of the research and the potential advantages to the organisation. A follow-up reminder was sent out twice prior to the deadline date. Both the VSM-94 and the MLQ-5 were incorporated into one neat questionnaire pack in order to minimise the perception that there were two questionnaires to complete which could have decreased the response rate. The questionnaire pack was designed with the objective of making the completion of the questionnaire as simple as possible therefore taking up minimal time from the managers. A back-up questionnaire pack was sent out with the original in the event of any respondent not being able to open the large file. As English has become the language of business in South Africa both questionnaires were administered in English to all participants.

RESULTS

Each item's frequency has been calculated and from Table 2 it is clear that item-scores of items 1 to 12, and 17 on the VSM-94 are positively skew leaning towards one end of the scale. Most of the remaining items (15-20) tend towards a bimodal distribution with extreme frequencies on both ends of the scale.

TABLE 2
ITEM RESPONSE DISTRIBUTIONS

	Of utmost importance	Very important	Moderate importance	Of little importance	Of no importance
q1	94	108	26	1	
q2	96	106	24	3	
q3	117	103	9		
q4	85	96	46	1	
q5	75	121	33		
q6	70	123	36		
q7	122	87	19	2	
q8	99	102	28		
q9	114	102	10	2	
q10	30	117	64	4	2
q11	82	127	15	1	4
q12	28	87	83	22	8
	Never	Seldom	Sometimes	Usually	Always
q13	8	58	130	26	8
	Very seldom	Seldom	Sometimes	Frequently	Very frequently
q14	12	40	74	92	12
	Strongly agree	Agree	Undecided	Disagree	Strongly disagree
q15	10	92	27	97	4
q16	21	104	13	73	19
q17	86	81	14	47	1
q18	11	49	27	132	11
q19	32	71	25	84	17
Q20	17	92	26	82	13

The inter-correlation of the items on each sub-domain have shown poor results as can be seen from Tables 3 to 7 below.

TABLE 3
CORRELATION ON THE INDIVIDUALISM DIMENSION

		q1	q2	q4	q8
Q1	Pearson Correlation	1.000	.119	.076	-.071
	N	229	229	229	229
Q2	Pearson Correlation	.119	1.000	.321(**)	.078
	N	229	229	229	229
Q4	Pearson Correlation	.076	.321(**)	1.000	-.038
	N	229	229	229	229
Q8	Pearson Correlation	-.071	.078	-.038	1.000
	N	229	229	229	229

** Correlation is significant at the 0.01 level (2-tailed).

TABLE 4
CORRELATION ON THE POWER DISTANCE DIMENSION

		q3	q6	q14	q17
q3	Pearson Correlation	1.000	.367(**)	-.062	.094
	N	229	229	229	228
q6	Pearson Correlation	.367(**)	1.000	.065	-.022
	N	229	229	229	228
q14	Pearson Correlation	-.062	.065	1.000	.026
	N	229	229	230	229
q17	Pearson Correlation	.094	-.022	.026	1.000
	N	228	228	229	229

** Correlation is significant at the 0.01 level (2-tailed).

TABLE 5
CORRELATION ON THE MASCULINITY DIMENSION

		q5	q7	q15	q20
q5	Pearson Correlation	1.000	.212(**)	-.022	-.118
	N	229	229	229	228
q7	Pearson Correlation	.212(**)	1.000	.062	-.101
	N	229	230	230	230
q15	Pearson Correlation	-.022	.062	1.000	-.010
	N	229	230	230	230
q20	Pearson Correlation	-.118	-.101	-.010	1.000
	N	229	230	230	230

** Correlation is significant at the 0.01 level (2-tailed).

TABLE 6
CORRELATION ON THE LONG-TERM ORIENTATION DIMENSION

		q9	q10	q11	q12
q9	Pearson Correlation	1.000	.212(**)	.108	.268(**)
	N	228	217	227	228
q10	Pearson Correlation	.212(**)	1.000	.079	.092
	N	217	217	217	217
q11	Pearson Correlation	.108	.079	1.000	.341(**)
	N	227	217	229	22
q12	Pearson Correlation	.268(**)	.092	.341(**)	1.000
	N	228	217	227	228

** Correlation is significant at the 0.01 level (2-tailed).

TABLE 7
CORRELATION ON THE UNCERTAINTY AVOIDANCE DIMENSION

		q13	q16	q18	q19
q13	Pearson Correlation	1.000	.037	.100	.114
	N	230	230	230	229
q16	Pearson Correlation	.037	1.000	-.143(*)	-.134(*)
	N	230	230	230	229
q18	Pearson Correlation	.100	-.143(*)	1.000	-.023
	N	230	230	230	229
q19	Pearson Correlation	.114	-.134(*)	-.023	1.000
	N	229	229	229	229

* Correlation is significant at the 0.05 level (2-tailed).

An investigation into the inter-item correlation showed that in many cases there is little or no correlation among items from the same sub-domain. Items that fail to consistently relate to one another, raises suspicion regarding the overall construct validity (Spector, Cooper & Sparks, 2001). On the 'individualism' dimension only two items correlate significantly with each other namely item 2 and 4. On the 'power distance' dimension a significant correlation was found between items 3 and 6 only. A significant correlation was also found between items 5 and 7 on the 'masculinity' index.

'Long-term orientation' yielded the best inter-correlation where 3 of the 4 items showed statistical significant correlations namely between item 9 and items 10, 11 and 12. An inter-correlation could however not be found between items 10 and 12 which both correlates significantly with item 9. The 'uncertainty avoidance' index showed a negative inter-correlation for items 16 and 18 and 16 and 19. Although all the correlations between items reported here are statistically significant, the correlations are very low for all items. This indicates that the items found in each of the sub-domains have very little in common, suggesting that they do not measure the same construct, as they should.

In order to establish if the item inter-correlation will comply with the criterion of sampling adequacy set for factor analysis an anti-image correlation was conducted for all the items

on the VSM-94 shown in Table 8. An anti-image correlation is the negative value of the partial correlation between variables. The partial correlation is the correlation that exists between variables when all other variable effects are taken into account (Hair, Anderson, Tatham & Black, 1998). Linked to the anti-imaging correlation is the measure of sample adequacy (MSA). The principal axis in Table 8 shows the MSA for each item.

The scores on the MSA can range from 0 to 1. A variable with a score of 1 is perfectly predicted without error from other variables (Hair et al., 1998). The authors propose the following guidelines in interpreting MSA scores: (0,80+ meritorious/outstanding), (0,70+ middling), (0,60+ mediocre), (0,50+ miserable) and (0,50- unacceptable).

Table 9 demonstrates that only 9 of the 20 items have MSA values of 0,60 or greater signifying that more than half of the items on the VSM-94 are inadequate, ranging from miserable to unacceptable. Items 2 to 6, 8 to 10 and 19 were extracted because of their acceptable MSA values and another anti-image correlation was done for these nine items. Item 8 produced a poor MSA score of 0,595 and was therefore omitted from further analysis. Table 9 shows the MSA values for the eight remaining items. It was decided to subject the eight remaining items to a principal factor analysis, a varimax rotation and an iterative item analysis.

TABLE 9
ANTI-IMAGE CORRELATION FOR ITEMS WITH HIGH MSA VALUES

	q2	q3	q4	q5	q6	q9	q10	q19
q2	.749(a)	-.045	-.206	.029	-.118	-.116	-.109	-.132
q3	-.045	.702(a)	-.113	-.162	-.294	-.032	-.051	.021
q4	-.206	-.113	.724(a)	-.124	.040	-.252	-.024	-.173
q5	.029	-.162	-.124	.743(a)	.138	-.072	-.073	-.164
q6	-.118	-.294	.040	-.138	.631(a)	-.027	-.069	.131
q9	-.116	-.031	-.252	-.073	-.027	.745(a)	-.179	-.108
q10	-.109	-.051	-.024	-.073	-.069	-.179	.660(a)	.171
q19	-.132	.021	-.173	-.164	.131	-.108	.171	.606(a)

a Measures of Sampling Adequacy(MSA)

TABLE 8
ANTI-IMAGE CORRELATION FOR VSM-94 (20 ITEMS)

	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	q11	q12	q13	q14	q15	q16	q17	q18	q19	q20
q1	.593a	-.096	-.069	.010	-.100	.070	.030	.074	-.080	-.147	.130	.030	-.044	.069	.036	-.007	.018	-.096	.075	-.030
q2	-.096	.668a	.004	-.245	-.022	-.092	.061	-.053	-.125	-.107	-.127	.057	-.094	.060	-.094	.024	.141	-.002	-.135	-.037
q3	-.069	.004	.650a	-.142	-.187	-.268	.012	.016	-.030	-.028	-.105	.081	-.065	.135	-.074	.032	-.042	.077	.019	-.172
q4	.010	-.245	-.142	.727a	-.056	.016	-.105	.109	-.198	-.017	.069	-.148	-.012	-.009	.120	-.016	-.098	-.048	-.118	.068
q5	-.100	-.022	-.187	-.056	.685a	-.144	.004	-.140	-.006	-.025	.031	-.109	.028	-.086	.062	-.011	-.231	-.094	-.151	.156
q6	.070	-.092	-.268	.016	-.144	.632a	-.221	-.045	.082	-.031	-.054	.141	-.000	-.054	-.041	.021	.138	-.013	.108	-.047
q7	.030	.061	.012	-.105	.004	-.221	.597a	-.273	.094	-.122	-.012	-.123	-.114	-.068	.009	-.002	-.184	-.028	.004	.158
q8	.074	-.053	.016	.109	-.140	-.045	-.273	.610a	.024	-.026	-.171	.054	.044	.040	-.056	-.103	.010	.003	.112	-.048
q9	-.080	-.125	-.30	-.198	-.006	-.082	.094	.024	.720a	-.121	-.051	-.215	.045	-.022	-.056	.066	-.223	.062	-.065	.044
q10	-.147	-.107	-.028	-.017	-.025	-.031	-.122	-.026	-.121	.684a	.021	-.041	.074	-.114	-.098	.056	-.098	.030	.150	-.133
q11	.130	-.127	-.105	.069	.031	-.054	-.012	-.171	-.051	.021	.593a	-.275	.069	-.018	.033	.070	.116	-.087	-.049	-.194
q12	.030	.058	.081	-.148	-.109	.141	-.123	-.054	-.215	-.041	-.275	.575a	-.065	.053	-.008	.018	.174	.120	-.091	.025
q13	-.044	-.094	-.065	-.012	.028	-.000	-.114	.044	.046	.074	.069	-.065	.460a	-.157	-.120	-.105	.000	-.150	-.129	-.072
q14	.069	.060	.135	-.009	-.086	-.054	-.068	.040	-.022	-.114	-.018	.053	-.157	.430a	-.155	.231	-.041	.319	.013	-.038
q15	.036	-.094	-.074	.120	.062	-.041	.009	-.056	-.056	-.096	.033	-.008	-.120	-.155	.547a	-.056	.021	-.090	.117	.048
q16	-.068	.024	.032	-.016	-.011	.021	-.002	-.103	.066	.055	.070	.018	-.105	.231	-.056	.468a	-.041	.207	.120	-.087
q17	.018	.141	-.042	-.098	-.231	.138	-.184	.010	-.223	-.098	.116	.174	.000	-.041	.021	.041	.561a	-.138	-.064	-.061
q18	-.096	-.002	.077	-.048	-.094	-.013	-.028	.003	.062	.030	-.087	.120	-.150	.319	-.090	.207	-.138	.404a	.070	.042
q19	.075	-.135	.019	-.118	-.151	.108	.004	.112	-.065	.150	-.049	-.091	-.129	.013	.117	.120	-.064	.070	.656a	.000
q20	-.030	-.037	-.172	.068	.156	-.047	.158	-.048	.044	-.133	-.194	.025	-.072	-.038	.048	-.087	.061	.04	.000	.490a

The eight items were inter-correlated and eigenvalues were calculated. According to Kaiser's (1961) criterion three factors were postulated. According to Table 10 the third factor was non-determined. The internal consistencies (Cronbach Alphas) of a remaining two-factor solution were 0,5491 and 0,5497 for factor 1 and 2 respectively, which is not satisfactory either.

TABLE 10
FACTOR MATRIX FOR 8 ITEMS WITH HIGH MSA SCORES

	Factor		
	1	2	3
q4	.586	-.233	
q9	.544	-.123	-.243
q3	.471	.324	.220
q2	.458		-.101
q5	.447		.168
q19	.334	-.527	.242
q6	.367	.461	.154
q10	.309	.237	-.313

Extraction Method: Principal Axis Factoring.
a Attempted to extract 3 factors. More than 25 iterations required.
(Convergence=1.361E-03). Extraction was terminated.

The MLQ-5 administered to the same sample yielded results consistent with other findings. The current study postulated four factors with acceptable reliabilities on factor 1 (0,9094) and factor 2 (0,7228) while alpha coefficients on factor 3 (0,4939) and factor 4 (0,5800) were low. Results have suggested that the latter two factors can be collapsed into one factor.

DISCUSSION

The purpose of this study was to investigate the metric properties of the VSM-94. The results suggest several factors that require consideration before using the VSM-94 in a South African context. The results of the MLQ-5 show findings consistent with Visser (1992) who extracted 4 factors and Ackermann, Schepers, Lessing and Dannhauser (2000) who postulated 3 factors with high reliabilities (0,944, 0,736 and 0,803). The consistent results obtained from the MLQ-5 indicate that there were no sample errors in this study. This consistency could however not be found for the VSM-94 on the same sample. These results highlight the difficulties in cross-cultural research, but also questions the VSM-94 used in measuring cultural values.

To address the first objective of the study, the VSM-94 has been evaluated against accepted criteria of test construction. According to the results of the study, Hofstede has succeeded in applying the first four criteria.

- The construct (values) has been clearly identified and theoretically investigated.
- According to the requirements of the second criteria the domain underlying the instrument was identified as cultural values.
- The sub-domains of 'power distance', 'individualism', 'masculinity', 'uncertainty avoidance' and 'long-term orientation' have been identified and adequately explained.
- Behaviour indicators for each of the sub-domains have been clearly stated and subsequently the instrument has been based on these indicators.

The construction of the items that need to measure the behaviour indicators and the response format on the VSM-94 poses some problems. The problem relates to a large extent to the formulation of the questions. The survey is made up of a combination of questions and statements. Items 1 to 12 are stated in a question format "how important is it to..." followed by the item content. The problem here is that respondents will

find most of the items important, resulting in all items being endorsed as of utmost importance. Swart et al. (1999) also mention that what a person sees as important is largely influenced by the current situation. In the case of the VSM-94 the context of the items is not clearly specified.

Items 13 and 14 are individually stated questions with individual response scales. There is also less skewness on these two items. Items 15 to 20 are statements prompting a response ranging from strongly agree to strongly disagree. Statements that are extremely positively or negatively directed will elicit response from the participants without fairly evaluating the content of the item. From Table 2 it is clear that the respondents answered in a relatively homogenous manner on the VSM-94. The item and response format of the VSM-94 therefore contributes to the problem of item response distributions. In order to reduce this effect Swart et al. (1999) propose that each item be transformed into an individually asked question like item 13 and 14.

The response format used in this instrument also poses problems although not to the same extent as the item format. All points on the five-point Likert scale used in the VSM-94 are anchored and Schepers (1992) noted that this eliminates the equal interval property of the scale. Interval scales supply a metric for measuring the differences between ranks (Gregory, 1996). Therefore equal interval scales have better discriminatory value than ordinal scales in comparative research because the intervals are constant across all groups measured. Swart et al. (1999) also propose that the five-point Likert scale be changed to a seven-point scale with only the two extreme points on the scale anchored. This will retain the equal interval property of the scale and thus yield improved statistical results.

The third objective of the study was to investigate the inter-correlation between the items. Results show little correlation amongst the items on each dimension. Only 'long-term orientation' and 'uncertainty avoidance' produced inter-correlations for more than 2 of the 4 items. Deleting these items will shorten the sub-scale to such an extent that it would hardly describe the values intended by Hofstede. These findings are consistent with the study by Spector et al. (2001) who also found redundant items within each sub-domain. Their study incorporated data from 23 countries including South Africa (135 sampled).

This study has further shown that most of the items on the VSM-94 can not be used in South Africa. The MSA scores of 12 of the 20 items are unacceptable as per Hair et al. (1998). The items with MSA scores of less than 0,60 range across all five the cultural dimensions. The result of this is that only one or two items are useful on each dimension, directly impacting on the reliability of each dimension and the instrument as a whole. The eight items remaining were factor analysed and have yielded unacceptably low reliabilities.

In investigating the fourth objective of transferability of the survey to other cultures the cross-cultural equivalence of the instrument was evaluated. The aim of the VSM-94 is to assess whether there are cultural value differences between two or more groups. In order to do this the instrument must be uniform for use across the various cultures.

Problems arise with conceptual equivalence where a common meaning of a stimulus or concept is a precondition for comparison. Instruments which are reliable in one country could contain phrases that are not interpreted consistently (Roodt et al., 2001). Although most people in the world understand English, it can not be assumed that everyone comprehends the same phrase in the same manner. This was shown with the word "thrift" used in the VSM-94. This question had 14 missing values probably because respondents did not understand the meaning of the word. The researchers had many queries regarding the meaning of this phrase in item 10. Even

when told the meaning of the word, respondents will interpret it within the context of their own framework. The latter signifies that the VSM-94 lacks semantic equivalence as well because the meaning is not consistent across cultures, even if the same language is used in the instrument. According to Greer and Greer (1998) it is preferred that the instrument emerges from the culture in which it is used, rather than carried over from another culture.

The biggest problem is with metric equivalence because the data in this study does not reflect the same structure as other data sets of previous research. Results have shown conflicting findings across studies done in South Africa and internationally. Although the instrument should be able to reflect real differences between groups on the same construct, it appears that situational factors and factors pertaining to the test play a major role. Swart et al. (1999) argue that the situation and the time when the survey is completed do play integral parts in contaminating the results. From the analysis, it can be seen that the items from the sub-domains were dispersed across all the factors indicating that the items on each dimension do not measure the same construct. Factor 1 consists of items from the 'power distance' (item 3), 'individualism' (Items 2 and 4) 'masculinity' (item 5) and 'long-term orientation' (item 9) sub-scales. 'Uncertainty avoidance' (item 9) and 'power distance' (item 6) made up factor 2 while factor 3 contained only one item (10) from the 'long-term orientation' dimension. It can therefore be said that the VSM-94 lacks metric equivalence because previous research as well as the current study do not show consistent factor extraction as per five dimensions postulated by Hofstede.

Even when including only highly correlating items in the factor analysis, the reliability of the VSM-94 can still not be justified for use in South Africa. Spector et al. (2001) also found unacceptable Cronbach alpha coefficients on the sub-domains even at the sample level, which included a South African sample. Unfortunately, these findings are not limited to South Africa as discovered in their study where a replication of Hofstede's dimensions failed to support the five sub-scales. Spector et al (2001) also reported very poor item inter-correlations on each of the sub-scales. Only 'long-term orientation' has shown consistent inter-correlations between the items. Since reliability is a major requirement for validity, it can be concluded that the VSM-94 would lack construct validity as well.

Kuchinke (1999) correctly observed that Hofstede did his original work in the late 1960's and early 1970's. The scores obtained from that study used to produce the original VSM have probably changed due to social, economical and political factors. This highlights the dynamic nature of values. Hofstede also used a single multinational corporation to collect the data. Although this would partially eliminate the influence of organisational culture, it does not adequately discriminate between national cultures.

The current study shows some limitations. The questionnaires returned were non-random and interpretations of the results of this study are therefore limited by both the size and the non-random nature of the sample. Other factors such as socio-economical status, education level and quality, language proficiency and acculturation should also be reckoned with. Although the sample was made up of females on managerial level, most of which have a tertiary education, the quality of the education and career opportunities could have played a major role.

The study has pointed out gaps in cross-cultural research and questionnaire construction and enough evidence is provided to oppose the use of the VSM-94 in scientific research. The lack of internal consistency also raises questions regarding the interpretation of existing results. Further research is required to validate the VSM-94 for use in South Africa by improving

the existing questionnaire for future use. Careful consideration should be given to the construction of the items, including the item and response format. Caution should also be applied when analysing the research data. All instruments need to be validated by scrutinising the item inter-correlations. Only if these inter-correlations are significant can a factor analysis be done on the items from which significant conclusions can be drawn.

REFERENCES

- Ackermann, C.P., Schepers, J.M., Lessing, B.C. & Dannhauser, Z. (2000). Die faktorstruktuur van Bass se Veelfaktor-Leierskapvraelys in die Suid Afrikaanse konteks. *Tydskrif vir Bedryfsielkunde*, 26 (2), 58-65.
- Bedell, B., Van Eeden, R & Van Staden, F. (1999). Culture as moderator variable in psychological test performance: Issues and trends in South Africa. *Journal of Industrial Psychology*, 25 (3), 1-7.
- Berry, J.W. & Triandis, H.C. (eds.) (1980). *Handbook of cross-cultural psychology*, 2. Boston: Allyn & Bacon.
- Bochner, S. & Hesketh, B. (1994). Power distance, individualism/collectivism, and job-related attitudes in a culturally diverse work group. *Journal of Cross-Cultural Psychology*, 25, 233-257.
- Elenkov, D.S. (1998). Can American management concepts work in Russia? A cross-cultural comparative study. *California Management Review*, 40 (4), 133-153.
- Gibson, J.L., Ivancevich, J.M. & Donnelly, J.H. (1994). *Organizations: Behavior, Structure, and Processes*. Burr Ridge Ill.: Irwin.
- Greer, J.M. & Greer, T. (1998). Cultural issues in Instrumentation and Survey research: Problems and prospects, *The 3rd International Conference on Management (ICM '98)*, July 25-28, Shanghai, China.
- Gregory, R.J. (1996). *Psychological testing: History, principles and applications* (2nd Ed). Boston: Allyn & Bacon.
- Hair, J.F., Anderson R.E., Tatham, R.L. & Black, W.C. (1998). *Multivariate data analysis* (5th Ed). Upper Saddle River, N.J.: Prentice Hall.
- Hellriegel, D., Slocum, J.W. & Woodman, R.W. (1998). *Organisational Behavior*. USA: International Thomson Publishing.
- Hofstede, G. (1983). National cultures in four dimensions. *International Studies of Management and Organisations*, 13 (2), 52-73.
- Hofstede, G. (1991). *Cultures and organizations: Software of the mind: Intercultural co-operation and its importance for survival*. New York: McGraw-Hill.
- Hofstede, G. (1993). Cultural constraints in management theories. *Academy of Management Executive*, 7, 81-94.
- Hofstede, G. (1994). *Value Survey Module 1994 manual*. IRIC, University of Tilburg, Maastricht, The Netherlands.
- Hofstede, G. & Bond, M.B. (1984). Hofstede's cultural dimensions: An independent validation using Rokeach's value survey. *Journal of Cross-Cultural Studies*, 15 (4), 417-433.
- Ivancevich, J.M. & Matteson, M.T. (1996). *Organizational Behavior and Management*. Chicago: Irwin.
- Kaiser, H.F. (1961). A note on Guttman's lower bound for the number of common factors. *British Journal of Statistical Psychology*, 14(1), 1
- Kuchinke, K.P. (1999). Leadership and culture: Work-related values and leadership styles among one company's U.S. and German telecommunications employees. *Human Resource Development Quarterly*, 10 (2), 135-154.
- Machet, M.P. (1996). Problematic issues in cross-cultural research. *Mousaion*, 14 (3), 60-67.
- Newman, K.L. & Nollen, S.D. (1996). Culture and congruence: The fit between management practices and national culture. *Journal of International Business Studies*, 27 (4), 753-780.
- Nelson, D.L. & Quick, J.C. (1997). *Organizational Behavior: Foundations, Realities and Challenges*. Minneapolis/St. Paul: West Publishing Company.

- Offerman, L.R. & Hellmann, P.S. (1997). Culture's consequences for leadership behaviour. *Journal of Cross-Cultural Psychology*, May 1997, 28 (3), 342-348.
- Robbins, S.P. (1996). *Organizational Behavior: Concepts, Controversies, and Applications*. Englewood Cliffs, N.J.: Prentice-Hall.
- Roodt, G., Kinnear, C., Erwee, R., Smith, D., Lynch, B. & Millett, B. (2001). Measurement of organisational inertia: Portability of a South African scale in an Australian context. *Journal of Industrial Psychology*, 27 (2), 57-61.
- Schepers, J.M. (1992). *Toetskonstruksie: Teorie en praktyk*. Johannesburg: Randse Afrikaanse Universiteit.
- Spector, P.E., Cooper, C.L. & Sparks, K. (2001). An international study of the psychometric properties of the Hofstede Survey Module 1994: A comparison of individual and country/province results. *Applied Psychology-an International Review*, 50 (2), 269-281.
- Swart, C., Roodt, G. & Schepers, J.M. (1999). Item-formaat, differensiële itemskeefheid en die faktorstruktuur van 'n selfvoltooiingsvraelys. *Tydskrif vir Bedryfsielkunde*, 25(1), 33-43.
- Theron, S.W. (1993). *A cross-cultural analysis of work-related values in a mining industry*. Department of Psychology: University of Pretoria.
- Thomas, A. & Bendixen, M. (2000). The management implications of ethnicity in South Africa. *Journal of International Business Studies*, 31 (3), 507-520.
- Visser, W.A. (1992). *Replication factor analysis of the Bass Multifactor Leadership Questionnaire*. Department of Psychology: University of Cape Town.