




Evidence-based case selection: An innovative knowledge management method to cluster public technical and vocational education and training colleges in South Africa



Authors:

Margaretha M. Visser¹ 
 Judith A. Van Biljon² 
 Marlien Herselman^{2,3} 

Affiliations:

¹Education and Skills Development Research Programme, Human Sciences Research Council, South Africa

²College of Science, Engineering and Technology, School of Computing, University of South Africa, South Africa

³Council for Scientific and Industrial Research, University of South Africa, South Africa

Corresponding author:

Mariette Visser,
 mmvisser@hsrc.ac.za

Dates:

Received: 03 Apr. 2016
 Accepted: 13 Dec. 2016
 Published: 22 Mar. 2017

How to cite this article:

Visser, M.M., Van Biljon, J.A. & Herselman, M., 2017, 'Evidence-based case selection: An innovative knowledge management method to cluster public technical and vocational education and training colleges in South Africa', *South African Journal of Information Management* 19(1), a751. <https://doi.org/10.4102/sajim.v19i1.751>

Read online:



Scan this QR code with your smart phone or mobile device to read online.

Background: Case studies are core constructs used in information management research. A persistent challenge for business, information management and social science researchers is how to select a representative sample of cases among a population with diverse characteristics when convenient or purposive sampling is not considered rigorous enough. The context of the study is post-school education, and it involves an investigation of quantitative methods of clustering the population of public technical and vocational education and training (TVET) colleges in South Africa into groups with a similar level of maturity in terms of their information systems.

Objectives: The aim of the study was to propose an evidence-based quantitative method for the selection of cases for case study research and to demonstrate the use and usefulness thereof by clustering public TVET colleges.

Method: The clustering method was based on the use of a representative characteristic of the context, as a proxy. In this context of management information systems (MISs), website maturity was used as a proxy and website maturity model theory was used in the development of an evaluation questionnaire. The questionnaire was used for capturing data on website characteristics, which was used to determine website maturity. The websites of the 50 public TVET colleges were evaluated by nine evaluators. Multiple statistical techniques were applied to establish inter-rater reliability and to produce clusters of colleges.

Results: The analyses revealed three clusters of public TVET colleges based on their website maturity levels. The first cluster includes three colleges with no websites or websites at a low maturity level. The second cluster consists of 30 colleges with websites at an average maturity level. The third cluster contains 17 colleges with websites at a high maturity level.

Conclusion: The main contribution to the knowledge domain is an innovative quantitative method employing a characteristic (in this case website maturity) as a proxy for MIS maturity. The method can be applied in quantitative, qualitative and mixed methods research, to group a population and thereby simplify the process of sample selection of cases for further in-depth investigation.

Introduction

'A scientific discipline without a large number of thoroughly executed case studies is a discipline without systematic production of exemplars, and a discipline without exemplars is an ineffective one.' (Flyvbjerg 2006:1)

Case research is a mature methodology in information and knowledge management research, and there is agreement that case selection is critical in ensuring high-quality case research (Benbasat, Goldstein & Mead 1987). Furthermore, the case selection should be based on characteristics of the context in which the phenomenon being investigated is found, because cases cannot be translated from one context to another, for example, from a developed country context to a developing country context (Mudzana & Maharaj 2015). What has received less attention is the development of methods for rigorous selection of case studies when there are a large number of cases from which to select (Seawright & Gerring 2008).

Copyright: © 2017. The Authors. Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

The study contributes to improving information and knowledge management practice, especially if one wants to ensure objectivity in case selection. The purpose of the study is to propose an evidence-based quantitative method for the selection of cases for case study research and to demonstrate this by clustering technical and vocational education and training (TVET) colleges in South Africa into groups with information systems at similar levels of maturity. The research context is public TVET colleges in South Africa. A broad overview of public TVET colleges in South Africa is provided in the next section.

Samples are data sets, taken from a wider data universe using a particular procedure, in order to generalise about the wider universe with a particular level of confidence (Ben-Zvi, Bakker & Makar 2015). To make the selection of a representative sample of public TVET colleges more rigorous, the use of an innovative sampling method was investigated and applied.

It is often not possible or feasible to survey all cases in a context because of cost implications and time constraints. Therefore, sampling is a key factor in making reliable statistical inferences about the universe. It is incumbent on the researcher to clearly define the target population and sample selection approach, but sometimes sampling is an underestimated part of a research study (Field & Hole 2003; StatPac 2014). The sample population is defined in keeping with the objectives of the specific study, and although guidelines are available, the researcher has to rely on logic and judgement to make the appropriate sample selection for the study, the latter being a subjective process. The method described here informs an evidence-based approach to case study selection.

Mouton (2001) identifies the use of biased sampling, because of the use of non-probability sampling techniques, as one of the main errors encountered in selecting data sources. Despite the maturity of this issue, the development of statistical reasoning in terms of samples and sampling in the education community deserves further attention (Ben-Zvi et al. 2015). Popular techniques for sample selection in qualitative research where case study strategies are implemented are convenient or purposive sampling, which only allows the researcher to generalise the research findings to the specific case under investigation. Serious challenges are likely to develop if a very small sample is randomly selected from a large population of cases without any prior stratification being done (Seawright & Gerring 2008; Williamson 2003).

The research question for the study was therefore formulated as follows: *How can an evidence-based quantitative method be used for case selection in information systems research?*

The process that was used to address the problem is based on web maturity models (WMM) theory (Rhoads 2008). An evaluation questionnaire was developed to investigate the

level of maturity of websites of the public TVET colleges. The colleges were then clustered by using algorithmic methods and statistical analysis software.

The article is presented in four parts. The first part provides the background to the investigation undertaken into the available literature. The methodological approach is then discussed, followed by the presentation of the findings. The article concludes by discussing the findings, with specific reference to their implications for improving rigor in case selection.

Literature review

The aim of the literature review is to provide background on public TVET colleges and to offer conceptual clarification of the key concepts and theoretical frameworks underpinning the empirical study.

Background on public TVET colleges

The TVET sector is defined by the United Nations Educational, Scientific and Cultural Organisation (UNESCO) as:

a comprehensive term referring to those aspects of the educational process involving, in addition to general education: the study of technologies and related sciences; as well as the acquisition of practical skills, attitudes, understanding, knowledge relating to occupations in various sectors of economic and social life. (Del Mar 2011:4)

Public TVET colleges in South Africa were formerly known as public further education and training colleges and were renamed in 2014 (Department of Higher Education and Training [DHET] 2013). Public TVET colleges were established and operated under the authority of the Continuing Education and Training Act 16 of 2006, and they are subject to the authority of the DHET (2013).

The college sector is seen as central to the provision of post-school education and training in South Africa, and strengthening and expanding this sector is the DHET's highest priority. A number of areas require strengthening at the colleges, including improving and developing information systems (DHET 2013; Dzvapatsva, Mitrovic & Dietrich 2014; Visser, van Biljon & Herselman 2013). There are 50 registered and accredited public TVET colleges in South Africa; these operate on more than 264 campuses in both rural and urban areas of the country (TVET Colleges South Africa 2015).

Sample selection and types of sampling

Sampling entails the selection of a subset of cases from the population under investigation. It is not always possible or feasible to conduct a census (which is an official count or survey that collects data from the entire population), because of time and cost constraints. A subset of the population under examination, if representative of the population and if large enough, can provide statistically significant results.

Theory on sampling deals with two domains: probability and non-probability sampling. Probability samples contain some type of randomisation and consist of simple, stratified, systematic, cluster, complex multi-stage or sequential sampling types (Oates 2006; Summers 1991). Non-probability samples lack randomisation and consist of the following sample types: convenience or accidental, purposive, quota, accessible, judgement, volunteer or self-selection, snowball and expert. The core distinctions between the two domains are that probability sampling study findings can be generalised to the target population and that it is mainly used in quantitative methods. Non-probability sampling study findings can only be generalised to the institution from which the sample was drawn, and it is primarily implemented when using qualitative research methods (Feild et al. 2006; Oates 2006; Summers 1991). Seddon and Scheepers (2012) emphasise the importance of sample representativeness and the need for researcher judgement when any claim is made about the likely truth of sample-based knowledge claims in other settings.

Seawright and Gerring (2008) emphasise the importance of the selection process when case selection has to be implemented. In most case selection techniques discussed in the literature, in-depth familiarity of each case is needed before a non-probability sample can be drawn.

Case selection in case study research has the same objectives as random sampling, which are: (1) to acquire a representative sample and (2) a useful variation on the dimensions of theoretical interest (Seawright & Gerring 2008). The selection of cases is therefore guided by the position of the case along these two dimensions within the population of interest. Seawright and Gerring (2008) identified the following case selection techniques: typical, diverse, extreme, deviant, influential, most similar and most different cases, and defined these as indicated in Table 1.

Miles, Huberman and Saldana (2014) provide criteria that can be used to evaluate the sample strategy that was used in a research study. The authors suggest that the sample should be relevant to the conceptual frame and research question(s) and that the researcher should make sure that the phenomena under investigation will appear in the sample. The sample plan should be viable in terms of cost, time, access to people

TABLE 1: Cross-case technique of case selection.

Technique	Definition
Typical	Cases (one or more) are typical examples of some cross-case relationship.
Diverse	Cases (two or more) exemplify diverse values.
Extreme	Cases (one or more) exemplify extreme or unusual values relative to some univariate distribution.
Deviant	Cases (one or more) deviate from some cross-case relationship.
Influential	Cases (one or more) with influential configurations of the independent variables.
Most similar	Cases (two or more) are similar in terms of specified variables.
Most different	Cases (two or more) are different in terms of specified variables.

Source: Seawright, J. & Gerring, J., 2008, 'Case selection techniques in case study research: A menu of qualitative and quantitative options', *Political Research Quarterly* 61(2), 297. <http://dx.doi.org/10.1177/1065912907313077>

and the researcher's own work style. The sampling plan should be ethical in terms of issues such as potential risks and benefits, informed consent and the relationship with informants. The researcher should also evaluate: if the plan will enhance generalisability of findings, either through conceptual power or representativeness; and if the findings will produce believable descriptions and explanations, which would be true in real life.

This brief literature review on sampling provides evidence that sampling is essential to the rigor of any research that relies on statistical inference, and also in research where the selection of the case study needs to be justified as non-convenient as often applies to research in information management and educational contexts. Sample selection is critical in ensuring research rigor yet non-trivial and that provides the rationale for this study into innovate, evidence-based sampling methods.

Web maturity models theory

The literature revealed that information and communication technology (ICT) is an essential enabler for economic and social development in an organisation and that it enhances the competitiveness of organisations (UNCTAD & TNSO 2008). ICT, which includes websites and management information systems (MISs), improves communication, operational efficiencies, sales turnover and information quality in organisations (Burgess, Sellitto & Karanasios 2009). Hence, the maturity and sophistication level of an organisation's website is indicative of the level of sophistication of its ICT, including its MISs.

Maturity models reveal the degree of technological sophistication and organisational transformation (Ziembra & Papaj 2013). Organisations also recognise the strategic importance of knowledge management and sharing, and the practise of using platforms such as the World Wide Web for this purpose (Mannie, Van Niekerk & Adendorff 2013).

The term 'maturity' relates to the degree of formality and optimisation of processes. The concept of maturity is fundamental to the evaluation of systems, and maturity models are used in different fields, such as business, education and information systems, to evaluate and monitor progress. The capability maturity model (CMM) identifies different levels of maturity in organisations; for most of the CMMs, the maturity scale structure includes five levels (Esterhuizen, Schutte & du Toit 2012; Paulk et al. 1993; Rhoads 2008). CMMs exist for many applications used in organisations, such as software development; information technology development; management; and project, data, business and knowledge management (Esterhuizen et al. 2012).

The WMM builds on the CMM and can also be used to assess the maturity of an organisation's website (Rhoads 2008).

The European Union eGovernment WMM represents the degree of technological sophistication and organisational

transformation in government agencies (Ziembra & Papaj 2013). Fath-Allah et al. (2014) compared 25 eGovernment maturity models and identified *presence*, *interaction*, *transaction* and *integration* as the criteria that differentiated the first four maturity levels in most of those websites. The levels of service and complexity are similar to those described for the European Union eGovernment model, but the latter proposes five levels because it distinguishes between one-way interaction and two-way interaction (Ziembra & Papaj 2013). The general website maturity model of the University of British Columbia (UBC) (UBC 2015) identifies *initial*, *repeatable*, *defined*, *managed* and *continual process* improvement as the five stages, and it becomes clear that the stages depend on the context and the purpose of the website.

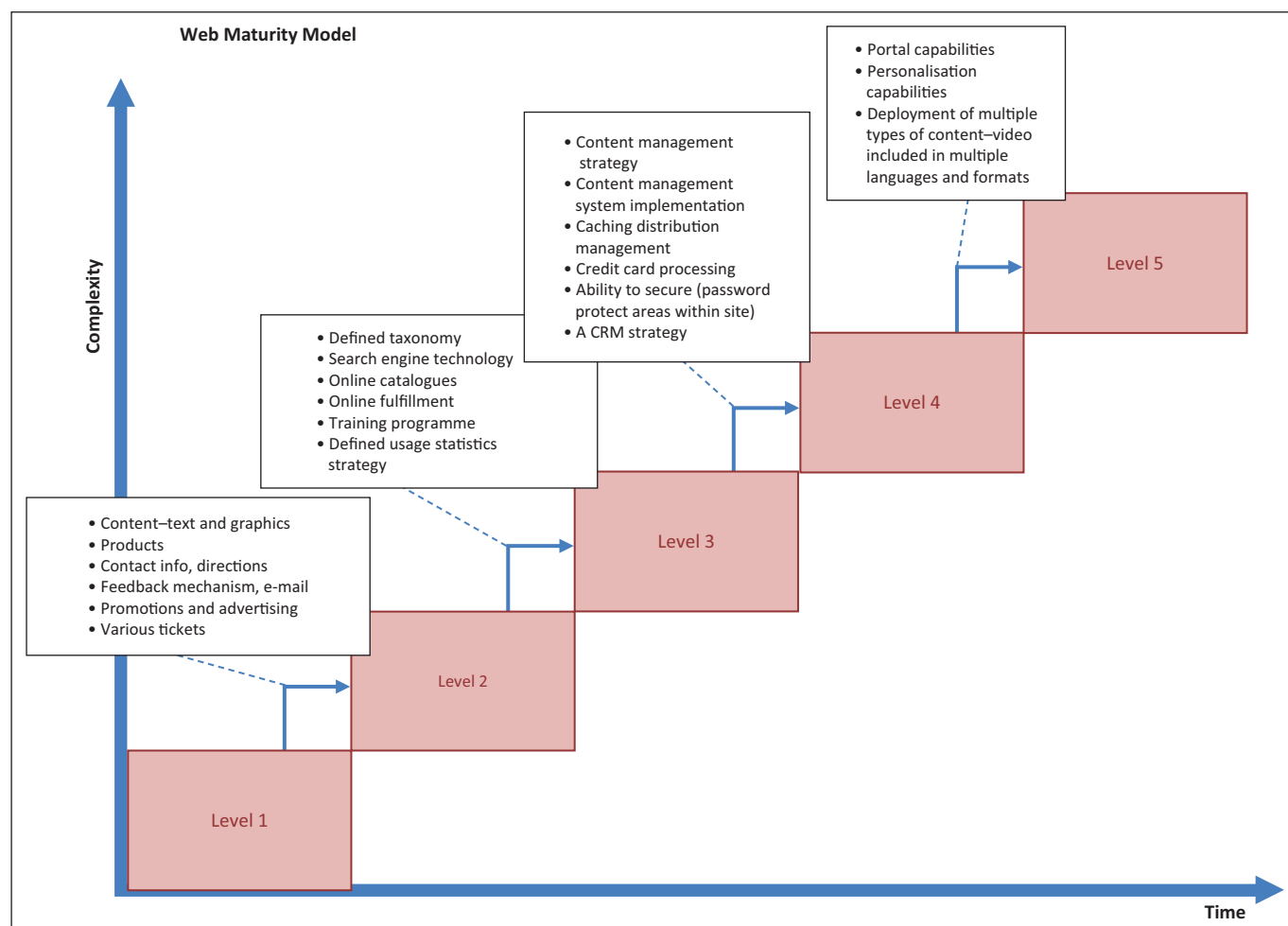
The design of an evaluation questionnaire for the evaluation of public TVET colleges' websites was modelled on Paulk et al. (1993) and Rhoads (2008), who focused on process control. The levels are as follows:

- level 1 relates to ad hoc business practices with no control at all
- level 2 relates to stable processes with a repeatable level of statistical control

- level 3 relates to defined processes to ensure consistent implementation
- level 4 relates to managed result metrics
- level 5 relates to active optimisation of the processes.

The WMM suggests that a website develops in maturity from level 1 to level 5, as depicted in Figure 1. Websites at level 1 provide basic introductory information about the institution (*presence*). The website evolves to level 2, if it includes text or information about the organisation, graphics, contact details and a feedback mechanism (*interaction*); websites develop to level 3 (*transactional*) if there is a search engine and more detailed information on what is offered by the institution (e.g. courses, training programmes and catalogues); websites develop to level 4 (*integration*) if they have systems such as content and distribution management, and evidence customer relationship management strategies and credit card processing functionality. A website maturity level 5 offers portal capability and personalised capability and contains multimedia content such as videos and multiple language choices.

In addition to evidence of specific components (functionality) on the website, as suggested by the WMM, Remenyi (2002) provides 10 main evaluation criteria for website maturity; each of these includes a number of sub-criteria. The main



Source: Rhoads, V., 2008, *Web maturity models, business, technology*, viewed 19 March 2015, from <http://www.slideshare.net/vtrhoads/web-maturity-models>

FIGURE 1: Web maturity model.

evaluation criteria include the following (Remenyi 2002): first impressions, navigation, content, attractors, findability, making contact, browser compatibility, knowledge of users, user satisfaction and other useful information. Most of these criteria relate to *user experience*.

The concepts *website maturity*, *user-centred design* and *usability* are intricately related (Earthy, Jones & Bevan 2001). Tullis and Albert (2013) view *usability* as the ability of the user to use the product to successfully perform a task; this is measured in terms of effectiveness, efficiency and user satisfaction in completing a task. *User experience* takes a broader view of the entire interaction with the product and includes thoughts, feelings and perceptions (Tullis & Albert 2013). This aligns with the view of Rubinoff (2004), who describes *user experience* as consisting of four factors, namely usability, content, branding and functionality. A discussion of *user experience* and *usability* is beyond the scope of this study, but it suffices to say that the study aligns with the view of *user experience* subsuming *usability*, with the focus here being on *usability*, *content* and *functionality*, for purposes of identifying the maturity attributes. Furthermore, *content* is measured in terms of its authority, purpose, usefulness, coverage, currency, objectivity and accuracy (Dalhousie University 2015).

Inter-rater reliability

The reliability of a measure is defined as the ability to produce the same results under the same conditions (Field & Hole 2003). Whenever humans are involved in conducting an evaluation, concerns are raised about the reliability and consistency of the results. Inconsistency in ratings is possible because the evaluators or raters could have been distracted, they become tired of doing repetitive tasks or they could have misinterpreted the evaluation criteria. However, there are ways in which consistency among raters can be determined.

Four general classes of reliability estimates are identified (Trochim 2006). The inter-rater or inter-observer reliability procedure is used to measure the degree to which the same phenomenon receives a consistent score by different raters. In statistics, inter-rater reliability, inter-rater agreement and concordance are described as the extent of agreement among raters. It shows how much homogeneity, or consensus, there is in the ratings. The second class of reliability estimates includes test-retest reliability, which is a procedure used to assess the consistency of a measure from one time to another. A single rater can, for instance, rate a phenomenon twice at different times. The third class includes parallel-forms reliability. This procedure is used to measure consistency in the results of two tests constructed in the same way using the same content domain. The fourth class consists of internal consistency reliability procedures, which assess the consistency of results across items within a test.

The most commonly known procedures that can be used for inter-rater reliability tests are the joint probability of agreement test (Uebersax 1987); Cohen's (1960) Kappa statistics, which works for two raters; and Fleiss' Kappa (Fleiss 1971), which is an improvement on Cohen's Kappa (1960)

and which works on any fixed number of raters. Correlation coefficients such as Spearman's ρ (rho) and Pearson's r can also be used to consider pairwise correlation among raters when using a scale that is ordered. Another method by which reliability testing can be performed is intraclass correlation coefficients (ICC); this is done by calculating the proportion of variance of an observation because of between subject variability in the true scores (Field 2006; Landis & Koch 1977; Ludbrook 2010).

Field (2009) notes two common uses of ICC: firstly, comparing paired data on the same measure; and secondly, assessing the consistency between the ratings provided by raters for a set of objects.

Calculation of the ICC depends on whether a measure of consistency (in which the order of scores from a source is considered but not the actual value around which the scores are anchored) or absolute agreement (in which both the order of scores and the relative values are considered), and whether the scores represent averages of many measures or just a single measure, is required (Field 2009).

Cluster analysis

Cluster analysis is a group of multivariate statistical techniques used to group data from a population into groups with similar characteristics. Three types of clustering techniques exist: hierarchical, non-hierarchical and a combination of hierarchical and non-hierarchical clustering techniques (Caccam & Refran 2012).

In hierarchical cluster analysis, the algorithm initially creates a cluster for each record or case in the database and then groups the cases together on the basis of similarities. It is a stepwise procedure that results in the construction of a hierarchy or tree-like structure of clusters (Sadiq 2012; Sarstedt & Mooi 2014). In non-hierarchical clustering analysis, a pre-determined number of categories are created, based on a selected criterion; the cases are then sorted into the categories or clusters based on similarities, using an iterative algorithm that optimises the chosen criterion (Sadiq 2012; Sarstedt & Mooi 2014).

The third technique, in which both hierarchical and non-hierarchical cluster methods are used, is called TwoStep cluster analysis (Elliott & Woodward 2007; Sadiq 2012). This method uses a hierarchical approach first and then a non-hierarchical approach. The hierarchical procedure produces the clusters; the non-hierarchical method then uses the produced clusters and clusters each case again to provide a more accurate cluster membership.

Research design

Data source and sample

An evaluation questionnaire was developed based on WMM theory, which included *user experience* attributes (cf. the section on WMM theory and Appendix 1). The aim of the

questionnaire was to gather data on the website characteristics of the total population of 50 public TVET colleges in South Africa. A Microsoft Access form was created to capture the collected data. Nine evaluators scored the different aspects of the college websites and captured their scores on the Microsoft Access form. The Statistical Package for the Social Sciences (IBM SPSS) programme was utilised to calculate the inferential statistics. Inter-rater reliability was established and the collected data on website characteristics were subjected to cluster analysis techniques, in order to present clusters of public TVET colleges with similar website characteristics. TVET colleges' website maturity was used as a proxy for TVET colleges' MIS(s) maturity.

The questionnaire consisted of 17 questions. Fifteen of the seventeen questions queried the websites in terms of the availability of specific components, such as establishment date, contact and direction details, having a feedback mechanism, having search engine technology and having social media links. These components are related to the different maturity levels of websites. The other two questions, together with their sub-questions (question 4 and sub-questions 4a to 4g; question 5 and sub-questions 5a to 5e), required the evaluators to provide a score from 1 (poor) to 5 (excellent) for the appearance, usability and content information presented on the websites.

A sample of nine evaluators from three institutions was selected to conduct the survey – seven female and two male evaluators. One of the evaluators was part of the research team, and the other eight were master's interns (5) and university students (3), who were chosen conveniently. One of the evaluators was a researcher, another was an expert in web design, five were studying in the domain of computer-user interaction and the other two had studied mechanical engineering. The evaluators were trained on how to rate each aspect of the website to ensure that they used the same evaluation criteria. The total population of public TVET college websites was evaluated.

Because the method used to cluster the public TVET colleges relied heavily on the scores given by the evaluators, it was important to check the scores for consistency and reliability. ICC were calculated to determine inter-rater reliability (cf. the section about inter-rater reliability) to establish absolute agreement in the average scores of the nine evaluators on questions 4 and 5.

After inter-rater reliability had been established, TwoStep cluster analysis (cf. the section on cluster analysis) was used to group the colleges, based on website maturity level. The level of maturity of a college's website was used as a proxy for the maturity level of the college's MIS(s).

Data management and analysis for this study were conducted by using IBM SPSS, Microsoft Excel and Microsoft Access. The results and findings are presented in the following section.

Results and findings

About the websites

An analysis of the year of establishment of the websites revealed that most of the websites were established recently. Seventeen (34%) of the websites were less than 2 years old; 10 (20%) were 2 to 3 years old; 4 (8%) were 4 to 5 years old and 4 (8%) were older than 5 years. Two (4%) TVET colleges did not own a website, and the dates of establishment were not displayed on 13 (26%) of the websites (Figure 2).

Table 2 presents the number of websites that contained the listed components. Evidently, most of the 48 college websites contained components associated with the second level of web maturity, such as contact details (98%), directions (96%) and a feedback mechanism (74%). The fact that 77% of the websites had social media links could indicate a high usage of this functionality by stakeholders (students, staff, suppliers, etc.). Just more than half (53%) of the websites made provision for students to download a registration form. The reason for this phenomenon could be related to the illegal practice of downloading and selling registration forms to potential students who are without Internet access (as explained on one of the college websites).

In the light of the current focus on e-learning, the utilisation of social media platforms for improving academic performance of TVET students in the country and the fact

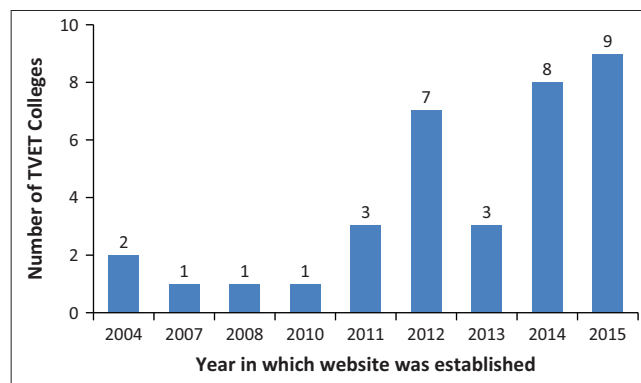


FIGURE 2: Year of establishment of TVET college websites.

TABLE 2: Number of websites by availability of specific components.

Components on website	Websites with capability		Web maturity level
	Number	%	
Contact details	46	98	2
Directions to college	45	96	2
Feedback mechanism	35	74	2
Search engine technology	27	57	3
Social media links	36	77	3
Registration form can be downloaded	25	53	3
Register online	6	12	4
Payment online	1	2	4
Portal capabilities for students	10	21	5
Deploy multiple types of content	7	15	5
Linked to a career portal or partners	17	36	5
Login facilities for staff or suppliers	16	33	5

that one of the biggest challenges identified by public TVET college lecturers was the lack of adequate contact hours for teaching (Dzvpatsva et al. 2014), it is surprising that only one in five websites (10 websites or 21%) had a portal capability for students. More than a third (17 or 36%) of the websites had electronic links to career portals.

The evaluators had to provide a score ranging from 1 (*poor*) to 5 (*excellent*) to questions and sub-questions 4 and 5 in the survey instrument. These questions examined the 'look-and-feel and usability' and 'content and information' of the websites, respectively (cf. Appendix 1). A weighted average index (WAI) was calculated for the scores provided by the evaluators at questions 4 and 5, as presented in Table 3. The WAIs of all evaluators were above the average of 2.5, which indicated that, on average, most of the websites were user-friendly and contained useful information. The minimum and maximum WAI values were (3.2 and 4.1) and (2.8 and 4.0), respectively, for the 'look-and-feel and usability' and value of 'content and information'. This indicates that more attention should be given to the 'content and information' on the websites.

Inter-rater reliability findings

An inter-rater reliability analysis was performed, using the ICC to determine absolute agreement among the average scores of the evaluators at questions 4 and 5 separately. The averages for questions 4a to 4g (variables were named $evalq4^i$

where $i \in \{1,2,3, \dots 9\}$) and questions 5a to 5e (variables were named $evalq5^i$ where $i \in \{1,2,3, \dots 9\}$) were calculated for each TVET college per evaluator. These average scores were used in the calculation of the ICC to establish inter-rater reliability.

The ICC were calculated by using a two-way random model, which controlled for evaluator effects and measured absolute agreement in which both the order of scores of evaluators and the relative values were considered. An ICC value of 0.7 and above is generally interpreted as acceptable, 0.8 and above is considered optimal and 0.9 and above is considered excellent (Field 2006, 2009).

Table 4 depicts the results of the ICC for the 'look-and-feel and usability' evaluation of the websites (from $evalq4^i$); Table 5 depicts the results for the 'content and information' evaluation (from $evalq5^i$). The tables include two sets of results: one for single measures and the other for average measures. We are interested in the average measures, which show how consistent the ratings were among all evaluators.

The inter-rater reliability on the scores for 'look-and-feel and usability' (Table 4) was found to be ICC (2,9) = 0.806 ($p < 0.000$), 95% CI (0.697, 0.885). In other words, the ICC value was 0.806, which indicates optimal agreement in the evaluators' scores. The confidence interval (CI) indicates that 95% of the samples of the data can be expected to have an ICC value between 0.697 and 0.885. The data thus seem to be reliable, with little variability between the evaluators.

TABLE 3: Frequencies of website scores provided by evaluators.

Category	Evaluator value	Score					Total	WAI
		Poor----->		-----Excellent				
		1	2	3	4	5		
Look-and-feel and usability	Evaluator 1	0	4	6	20	17	47	4.1
	Evaluator 2	2	8	21	11	5	47	3.2
	Evaluator 3	0	6	27	10	4	47	3.3
	Evaluator 4	0	9	23	9	6	47	3.3
	Evaluator 5	3	4	15	21	0	43	3.3
	Evaluator 6	1	3	28	11	2	45	3.2
	Evaluator 7	1	1	11	22	10	45	3.9
	Evaluator 8	1	1	12	19	12	45	3.9
	Evaluator 9	5	7	12	13	7	44	3.2
Content and information	Evaluator 1	0	1	13	16	17	47	4.0
	Evaluator 2	0	5	20	13	9	47	3.6
	Evaluator 3	0	11	19	15	2	47	3.2
	Evaluator 4	2	5	19	19	2	47	3.3
	Evaluator 5	6	4	24	10	0	44	2.9
	Evaluator 6	2	2	27	16	0	47	3.2
	Evaluator 7	1	2	8	19	15	45	4.0
	Evaluator 8	2	3	14	19	7	45	3.6
	Evaluator 9	9	10	8	15	2	44	2.8

WAI, weighted average scores.

TABLE 4: Intraclass correlation coefficient of rater scores for websites in terms of look-and-feel and usability.

Measure	Intraclass correlation \ddagger	95% confidence interval			F-test with true value 0		
		Lower bound	Upper bound	Value	df 1	df 2	Sig
Single	0.316a \dagger	0.204	0.460	6.410	40	320	0.000
Average	0.806	0.697	0.885	6.410	40	320	0.000

df, degrees of freedom; Sig, significance.

\dagger , The estimator is the same, whether the interaction effect is present or not; \ddagger , Type A intraclass correlation coefficients using an absolute agreement definition.

Note: Two-way random effects model where both people effects and measures effects are random.

TABLE 5: Intraclass correlation coefficient of rater scores for websites in terms of content and information.

Measure	Intraclass correlation \ddagger	95% confidence interval		Value	F-test with true value 0		
		Lower bound	Upper bound		df 1	df 2	Sig
Single	0.223 \dagger	0.126	0.359	4.908	40	320	0.000
Average	0.721	0.564	0.835	4.908	40	320	0.000

df, degrees of freedom; sig, significance.

\dagger , The estimator is the same, whether the interaction effect is present or not; \ddagger , Type A intraclass correlation coefficients using an absolute agreement definition.

Note: Two-way random effects model where both people effects and measures effects are random.

In addition, 80.6% of the variance in the means of the evaluators' scores was real and not because of chance.

The inter-rater reliability for evaluator scores on the 'content and information' (Table 5) was found to be ICC (2,9) = 0.721 ($p < 0.000$), 95% CI (0.564, 0.835). In other words, the ICC value was 0.721, which indicates an acceptable agreement in the evaluators' scores. The CI indicates that 95% of the samples of the data can be expected to have an ICC value between 0.564 and 0.835. The data thus seem to be reliable, with little variability between the evaluators. In addition, 72.1% of the variance in the means of the evaluators' scores was real and not because of chance.

After inter-rater reliability had been established and the data were found to be reliable, cluster analysis was performed on the data.

Cluster analysis findings

A TwoStep cluster analysis, which includes both hierarchical and non-hierarchical cluster methods, was performed on the data to find a suitable model to cluster the colleges. The TwoStep cluster analysis made use of the log-likelihood distance measurement and was based on Schwarz's Bayesian criterion.

Three input variables were used in the clustering analysis, that is, the overall average of all evaluators' scores for question four (variable named *Avgq4*) and for question five (variable named *Avgq5*), and a newly created variable called *Compfinal*. The variable *Compfinal* represented the number of components included on the websites of the colleges; this was generated by calculating the sum of the values for questions 6 and 8 to 17 in the questionnaire for each college. The values for the variable *Compfinal* could, therefore, be any number from 0 to 11 (dichotomous variables: 0 = no or 1 = yes). Descriptive statistics for the variable *Compfinal* were calculated: values ranged from 0 to 7, the mean was $M = 4.08$ and the standard deviation was $SD = 1.75$.

Figure 3 presents the results of the most suitable model of clusters. The model presents a silhouette measure of cohesion and separation of above 0.5, which indicates cluster quality and a good model fit. Usually, an acceptable ratio of cluster sizes is between 2 and 3 (Sadiq 2012; Sarstedt & Mooi 2014), but because two TVET colleges did not own a website, a separate cluster to contain those colleges and the colleges with websites at a low maturity level made sense, and the cluster ratio size of 10.00 was accepted.

The model presented three clusters: cluster 1 (6% or 3 colleges) represents colleges with no websites or websites at a very low maturity level; cluster 2 (60% or 30 colleges) represents colleges with websites at an average maturity level; cluster 3 (34% or 17 colleges) represents colleges with good quality websites. The technical report in which the outcome of the cluster analysis is presented can be viewed at <http://tinyurl.com/zde2fa5>.

The results can be summarised as follows:

- *Websites in cluster 1* had, on average, one of the 11 components; were evaluated at 0.55, on average, for 'look-and-feel and usability', were rated, on average, at 0.46 for 'content and information' quality.
- *Websites in cluster 2* had, on average, three of the 11 components; and scored, on average, 3.29 and 3.06 for 'look-and-feel and usability' and 'information and content', respectively.
- *Websites in cluster 3* had, on average, six of the 11 components; and scored, on average, 4.02 and 3.67 for 'look-and-feel and usability' and 'information and content', respectively.

The results are graphically presented in Figure 3, in which it is evident that the input variable *Avgq5* had the highest predictor importance (1.00) and the other two variables – 'mean score of evaluators on question 4' (*Avgq4*) and *Compfinal* – contributed 0.99 and 0.96 to the prediction of the clusters, respectively.

Ethical consideration

Ethical clearance for the research study was granted by the College of Science, Engineering and Technology's (CSET) Research and Ethics Committee, which is part of the University of South Africa (UNISA).

Discussion

This study has implications for methods of selecting cases for qualitative case study research. The study showed how the set of 50 public TVET colleges was grouped into groups with similar website characteristics. An evaluation questionnaire was developed, based on WMM theory. The evaluation questionnaire included user experience attributes and was used by nine evaluators to evaluate the websites of the 50 public TVET colleges in South Africa. Inter-rater reliability of the nine evaluators was established, with the calculation of ICC. Hierarchical and non-hierarchical cluster analysis (TwoStep) was used to cluster the population of public TVET colleges.

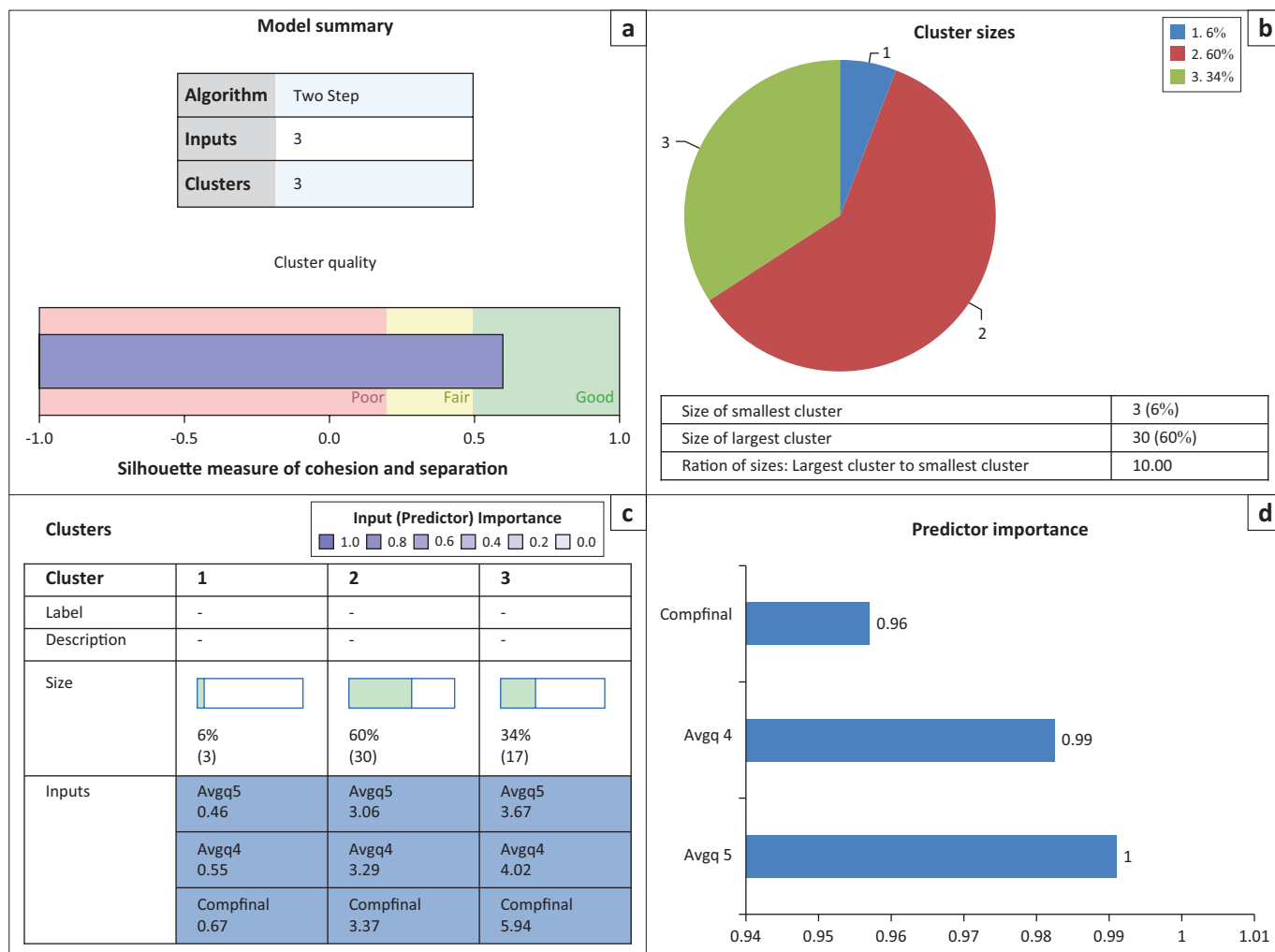


FIGURE 3: (a) Model summary, (b and c) cluster sizes and (d) predictor importance, as calculated using TwoStep cluster analysis.

Three groups emerged from the analysis (cf. Figure 3): colleges without a website or websites at a low maturity level (cluster 1); colleges with websites that achieved an average score for first impression of structure, appearance, navigation, ease of finding specific functionality and the comprehensiveness and usefulness of content and information about programmes offered and how to apply and register at the college (cluster 2); and colleges that met the usability requirements of user experience and satisfaction, information requirements and comprehensive functionality in terms of number of components available (cluster 3).

It is important to keep in mind that the outcome of statistical or technical procedures should (as far as possible) be verified with qualitative empirical evidence (Punj & Stewart 1983). Although we have to reject clusters that do not meet minimum statistical requirements, this does not mean that clusters that are statistically acceptable are the only meaningful clustering outcome (Klasterin 1983). Future studies could investigate other methods to cluster public TVET colleges into groups with similar website or MIS characteristics.

Furthermore, for future research that focuses on websites or MISs at public TVET colleges, researchers could select cases from the three clusters to ensure representation of all

categories. One should also take into consideration that college websites and MISs develop and grow in maturity over time, and therefore, it might be necessary to repeat this analysis in future to enable up-to-dating of the categorisation of public TVET colleges in groups with similar website or MIS characteristics.

Conclusion

The aim of the study was to propose an evidence-based quantitative method for the selection of cases for case study research. The method was demonstrated by clustering TVET colleges in South Africa based on website maturity models, as a proxy for information systems maturity.

The method of case selection presented in this article is a more rigorous alternative to convenient or purposive sample selection techniques and was intended to be uncomplicated and replicable by researchers and practitioners with different levels of proficiency in statistical analysis.

Technological developments have provided many tools for new and innovative ways of analysing data. However, to ensure rigor in the selection of cases in qualitative research studies, it is important that researchers also develop

innovative methods for sample selection to overcome subjectivity and bias. The evidence-based method, which used quantitative data, is a contribution in terms of basic practice for sample selection. Quantitative data were gathered through an evaluation questionnaire to generate a proxy for the level of IS development of individual public TVET colleges in South Africa. The presented method can be applied by any researcher to ensure that objectivity is ensured when selecting cases and to ensure rigour in the research methodology of a study.

Besides the practical contribution of clustering public TVET colleges, the value of this study resides on the evidence-based method, and the replicative value that it offers researchers and practitioners in any field of study, where cases for in-depth case studies, has to be selected. A summary of the method is provided in Appendix 2. The novelty lies in using a specific characteristic as proxy for clustering the cases, so that at least one case from each cluster can be selected. Clustering of the population in this study was based on the maturity level of the population's websites, but in other populations, other easily accessible and available relevant information could be collected, captured and used in cluster analysis. Future research is needed to triangulate the findings from the quantitative clustering with a qualitative method, for example, clustering by expert review.

Acknowledgements

The authors would like to express their sincere gratitude to the nine evaluators who took the time to evaluate the 50 public TVET college websites. Our appreciation also goes to Dr. Linda Zuze, for comments made on the draft version of this article. The study was partly funded by UNISA and the Human Sciences Research Council (HSRC).

Competing interests

The authors declare that they have no financial or personal relationship that may have influenced them inappropriately in writing this article.

Author's contributions

The study was conducted as part of the doctoral degree requirements of Mariette Visser. Judy van Biljon acted as the supervisor, and Marlien Herselman acted as the joint supervisor for the study.

References

- Benbasat, I., Goldstein, D.K. & Mead, M., 1987, 'The case research strategy in studies of information systems', *MIS Quarterly* 11(3), 369–386. <https://doi.org/10.2307/248684>
- Ben-Zvi, D., Bakker, A. & Makar, K., 2015, 'Learning to reason from samples', *Educational Studies in Mathematics* 88, 291–303, viewed 11 April 2015, from <http://link.springer.com/article/10.1007/s10649-015-9593-3/fulltext.html>
- Burgess, S., Sellitto, C. & Karanasios, S., 2009, *Effective web presence solutions for small businesses: Strategies for successful implementation*, IGI Global, Hershey, PA.
- Caccam, M. & Refran, J., 2012, *Cluster analysis*, Slideshare, 1–100, viewed 19 March 2015, from <http://www.slideshare.net/jewelrefran/cluster-analysis-15529464>

- Cohen, J., 1960, 'A coefficient for agreement for nominal scales', *Education and Psychological Measurement* 20, 37–46. <https://doi.org/10.1177/001316446002000104>
- Dalhousie University, 2015, *Criteria for websites. Evaluating web resources*, Dalhousie University, viewed 19 March 2015, from http://libraries.dal.ca/using_the_library/evaluating_web_resources/6_criteria_forwebsites.html
- Del Mar, V., 2011, *Introducing UNESCO's Technical Vocational Education and Training (TVET) definition and strategy*, United Nations Educational, Scientific and Cultural Organization (UNESCO), viewed 13 September 2015, from <http://www.uis.unesco.org/StatisticalCapacityBuilding/>
- Department of Higher Education and Training (DHET), 2013, *White paper for post-school education and training: Building an expanded effective and integrated post-school system*, Department of Higher Education and Training (DHET), Pretoria.
- Dzvapatva, G.P., Mitrovic, Z. & Dietrich, A.D., 2014, 'Use of social media platforms for improving academic performance at Further Education and Training colleges', *South African Journal of Information Management* 16(1), Art. #604, 1–7. <https://doi.org/10.4102/sajim.v16i1.604>
- Earthy, J., Jones, B.S. & Bevan, N., 2001, 'The improvement of human-centred processes – Facing the challenge and reaping the benefit of ISO 13407', *International Journal of Human-Computer Studies* 55(4), 553–585, viewed 11 April 2015, from <http://www.idealibrary.com>
- Elliott, A.C. & Woodward, W.A., 2007, *Statistical analysis quick reference guidebook: With SPSS examples*, Sage Publication Ltd., London.
- Esterhuizen, D., Schutte, C. & du Toit, A., 2012, 'A knowledge management framework to grow innovation capability maturity', *South African Journal of Information Management* 14(1), Art. #495, 1–10. <https://doi.org/10.4102/sajim.v14i1.495>
- Fath-Allah, A., Cheikhi, L., Al-Qutaish, R.E. & Idri, A., 2014, 'E-government maturity models: A comparative study', *International Journal of Software Engineering & Applications* 5(3), 71–91.
- Feild, L., Pruchno, R.A., Bewley, J., Lemay, E.P. & Levinsky, N.G., 2006, 'Using probability vs. nonprobability sampling to identify hard-to-access participants for health-related research: Costs and contrasts', *Journal of Aging and Health* 18(4), 565–583. <https://doi.org/10.1177/0898264306291420>
- Field, A.P., 2006, 'Intraclass correlation', *Encyclopedia of Statistics in Behavioural Science*, Volume 2, pp. 948–954, John Wiley and Sons Ltd., Chichester, UK.
- Field, A.P., 2009, *Discovering statistics using SPSS*, 3rd edn., Sage Publication Ltd., London.
- Field, A.P. & Hole, G., 2003, *How to design and report experiments*, Sage Publications Ltd. London.
- Fleiss, J.L., 1971, 'Measuring nominal scale agreement among many raters', *Psychological Bulletin* 76(5), 378–382.
- Flyvbjerg, B., 2006, 'Five misunderstandings about case-study research', *Qualitative Inquiry* 12(2), 219–245, viewed 25 March 2016, from <http://qix.sagepub.com/content/12/2/219.abstract>
- Klastorin, T.D., 1983, 'Cluster analysis results', *Journal of Marketing Research* 20(1), 92–98. <https://doi.org/10.2307/3151416>
- Landis, J.R. & Koch, G.G., 1977, 'The measurement of observer agreement for categorical data', *Biometrics* 33, 159–174. <https://doi.org/10.2307/2529310>
- Ludbrook, J., 2010, 'Confidence in Altman–Bland plots: A critical review of the method of differences', *Clinical and Experimental Pharmacology and Physiology* 37(2), 143–149. <https://doi.org/10.1111/j.1440-1681.2009.05288.x>
- Mannie, A., Van Niekerk, H.J. & Adendorff, C.M., 2013, 'Significant factors for enabling knowledge sharing between government agencies within South Africa', *South African Journal of Information Management* 15(2), Art. #569, 1–8. <https://doi.org/10.4102/sajim.v15i2.569>
- Miles, M.B., Huberman, A.M. & Saldana, J., 2014, *Qualitative data analysis: A methods sourcebook*, Sage Publications, Inc., Thousand Oaks, CA.
- Mouton, J., 2001, *How to succeed in your master's and doctoral studies: A South African guide and resource book*, Van Schaik Publishers, Pretoria.
- Mudzana, T. & Maharaj, M., 2015, 'Measuring the success of business-intelligence systems in South Africa: An empirical investigation applying the DeLone and McLean Model', *South African Journal of Information Management*, 17(1), Art. #646, 1–7, viewed 21 March 2016, from <http://www.sajim.co.za/index.php/SAJIM/article/view/646>
- Oates, B.J., 2006, *Researching information systems and computing*, Sage Publication Ltd, London.
- Paulk, M.C., Weber, C.V., Curtis, B. & Chrissis, M.B., 1993, *Capability maturity model for software (Version 1.1)*, Technical Report, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, CMU/SEI-93-TR-024 ESC-TR-93-177.
- Punj, G. & Stewart, D.W., 1983, 'Cluster analysis in marketing research: Review and suggestions for application', *Journal of Marketing Research* 20(2), 134–148. <https://doi.org/10.2307/3151680>
- Remenyi, D., 2002, *Criteria to evaluate a website*, Management Centre International Limited, England, viewed 19 March 2015, from <http://www.mcil.co.uk/review/7-10-criteria.htm>
- Rhoads, V., 2008, *Web maturity models, business, technology*, viewed 19 March 2015, from <http://www.slideshare.net/vtrhoads/web-maturity-models>
- Rubinoff, R., 2004, 'Quantify user experience', *Design & UX*, viewed 12 June 2015, from <http://www.sitepoint.com/quantify-user-experience/>
- Sadiq, M.M., 2012, *Cluster analysis in SPSS (hierarchical, non-hierarchical & two-step)*, viewed 19 March 2015, from <https://www.youtube.com/watch?v=eW9KnSRDYTK>

- Sarstedt, M. & Mooi, E., 2014, *A concise guide to market research: The process, data, and methods using IBM SPSS statistics*, Springer-Verlag, Berlin, viewed 19 March 2015, from <http://www.guide-market-research.com/>
- Seawright, J. & Gerring, J., 2008, 'Case selection techniques in case study research: A menu of qualitative and quantitative options', *Political Research Quarterly* 61(2), 294–308. <https://doi.org/10.1177/1065912907313077>
- Seddon, P. & Scheepers, R., 2012, 'Towards the improved treatment of generalization of knowledge claims in IS research: Drawing general conclusions from samples', *European Journal of Information Systems* 21, 6–21. <https://doi.org/10.1057/ejis.2011.9>
- StatPac., 2014, *Survey sampling methods*, StatPac, viewed 11 April 2015, from <https://www.statpac.com/surveys/sampling.htm>
- Summers, S., 1991, 'Selecting the sample for a research study', *Journal of Post Anesthesia Nursing* 6(5), 355–363, viewed 17 September 2015, from <http://www.ncbi.nlm.nih.gov/pubmed/1833537>
- Trochim, W.M., 2006, *The research methods knowledge base*, 2nd edn., Research Methods: Knowledge Base, viewed 2 April 2015, from <http://www.socialresearchmethods.net/kb>
- Tullis, T. & Albert, W., 2013, *Measuring the user experience – Collecting, analyzing, and presenting usability metrics*, 2nd edn., Morgan Kaufman, 320 p.
- TVET Colleges South Africa, 2015, *Public TVET Colleges: Overview*, The official website for the DHET TVET Colleges, viewed 25 March 2015, from <http://www.fetcolleges.co.za/>
- UBC, 2015, *Web maturity model – Information technology*, University of British Columbia, viewed 5 December 2015, from <http://www.it.ubc.ca>
- Uebersax, J.S., 1987, 'Diversity of decision making models and the measurement of interrater agreement', *Psychological Bulletin* 101, 140–146. <https://doi.org/10.1037/0033-2909.101.1.140>
- UNCTAD & TNSO, 2008, *United Nations Conference on Trade and Development. Measuring the impact of ICT use in business: The case of manufacturing in Thailand. Prepared jointly by the UNCTAD secretariat and the Thailand National Statistical Office (TNSO)*, United Nations, New York, viewed 12 November 2016, from http://unctad.org/en/docs/sdteecb20073_en.pdf
- Visser, M.M., van Biljon, J.A. & Herselman, M.E., 2013, 'Evaluation of management information systems: A study at a Further Education and Training college', *South African Journal of Information Management* 15(1), Art. #531, 1–8. <https://doi.org/10.4102/sajim.15i1.531>
- Williamson, G.R., 2003, 'Misrepresenting random sampling? A systematic review of research papers in the Journal of Advanced Nursing', *Journal of Advanced Nursing* 44(3), 278–288. <https://doi.org/10.1046/j.1365-2648.2003.02803.x>
- Ziemba, E. & Papaj, T., 2013, 'A pragmatic approach to the e-government maturity in Poland – Implementation and usage of SEKAP', Academic Conferences International Limited, Varese, Italy, June 2013, pp. 560–570.

Appendix starts on the next page →

Appendix 1

Questionnaire: Evaluation tool

The following three pictures depict the pages in the Microsoft Access tool that was used in the evaluation of the TVET college websites.

EVALUATION OF TVET COLLEGES' WEBSITES

Page1 Page2 Page3

UniqueID:
 Province:
 College:
 URL:

q1: Does the college have a website?
 q2: When (year) was the website established?
 q3: Has the website been updated in 2015?

Please provide a score for each of the following questions from 1 to 5, where 1=Poor and 5=Excellent.

4a) Rate your first impression about the **STRUCTURE** (identifiable components which give form to the website) of the website?
 4b) Rate your first impression about the **APPEARANCE / GRAPHICS** (look and feel) of the website?
 4c) Rate to what extent does the website **CATER** for its main clients which are the **STUDENTS**?
 4d) How easy is it to **NAVIGATE** around on the website?
 4e) **HOW EASY** is it to find information on **HOW TO REGISTER** at the college?
 4f) **HOW EASY** is it to find information on the **PROGRAMMES** offered at the college?
 4g) Rate to what extent the **LINKS** on the website work? (1=found more than one link not working; 2=found one link not working; 3=website doesn't have many links but found no links not working; 5=complex website and all links selected worked well.)

q4: How would you rate the overall appearance of the text and graphics of the website in terms of appearance, clarity, ease of use and accessibility? Please use a scale from 1 to 5 where 1=Poor and 5=Excellent.

Go to page 2

EVALUATION OF TVET COLLEGES' WEBSITES

Page1 Page2 Page3

Please check if the following components are present on the website:

q6: TVET College contact details (Telephone no., Email address, etc.)?
 q7: Directions to TVET College and / or campuses (Physical address, Map,)?
 q8: Feedback mechanism to send an email to TVET College?
 q9: Is there a search function available?
 q10: Social media links (Facebook, Twitter, etc.)?
 q11: Application form for registration downloadable?
 q12: Is it possible for students to register online?
 q13: Is it possible to process payment to online?
 q14: Does the website have portal capabilities where a student can have his/her own space with specific study material?
 q15: Deploy multiple types of content (videos, multiple languages, etc.)?
 q16: Linked to career portal or partners (SETAs, service providers, etc.)?
 q17: Have login facilities for e.g. staff, suppliers, etc.?

Next record

EVALUATION OF TVET COLLEGES' WEBSITES

Page1 Page2 Page3

Please provide a score for each of the following questions from 1 to 5, where 1=Poor and 5=Excellent.

5a) To what extent does the website provide information about **STUDENT LIFE** (in terms of entertainment, sport or cultural activities) on the campus?
 5b) To what extent does the website provide information about the procedure to apply for **REGISTRATION**?
 5c) To what extent does the website provide information about the **PROGRAMMES and COURSES** offered at the college?
 5d) To what extent does the website provide information on available **TENDERS for SUPPLIERS** or service providers?
 5e) To what extent does the website provide information on **VACANCIES** (for prospective staff) at the college?

q5: How would you rate the general information and information on programmes, courses and procedures on how to apply and register of the website in terms of completeness? Please use a scale from 1 to 5 where 1=Poor and 5=Excellent.

Go to page 3

Appendix 2

Steps in the method proposed in this study

- Step 1: Decide which characteristics the clusters should be included, based on the research study for which the cases are required.
- Step 2: Do a literature study on the characteristics under investigation (in this study, the IS characteristics of public TVET colleges, based on the web maturity level of the websites).
- Step 3: Design and develop (based on extant literature) a survey or evaluation questionnaire to capture appropriate quantitative data on the characteristics.
- Step 4: Select and brief reviewers or data collectors.
- Step 5: Collect the data and compile into an electronic data set.
- Step 6: Import the data set into a statistical software package and perform descriptive and inferential statistical procedures to ensure a rigorous outcome.
- Step 7: To calculate the ICC, the following steps were followed separately for variables *evalq4ⁱ* and *evalq5ⁱ*: in IBM SPSS, select 'Analyse', 'Scale', 'Reliability Analysis', then select the appropriate variables (in this study, first *evalq4ⁱ* and then *evalq5ⁱ* in the second analysis), then click on 'Statistics'. Make sure that the box next to 'Intraclass Correlation Coefficient' is checked. After the ICC box has been checked, select the 'Two-way Random' (or context applicable) model and the 'Absolute Agreement' type. Then run the analyses.
- Step 8: Three variables were created for inclusion in a cluster analysis (*Avgq4*, *Avgq5*, *Compfinal*). The following steps were followed in IBM SPSS: select 'Analyse', 'Classify', 'TwoStep Cluster'. Then select the variables and place them in the applicable variable type box (categorical/ continuous). Change the 'Determine automatically maximum' clusters to the required number (in this study, 4 was chosen). Under 'Output', check 'Create cluster membership variable', to enable the creation of a variable for the classification of the TVET colleges.
- Step 9: Select cases from each cluster, as appropriate.