2012

A Critical Evaluation and Comparison of Two Formative Measures of System Quality Using Criterion Variables

Azim Ali
Victoria University of Wellington, azim.ali@telecom.co.nz

Mary Tate
Victoria University of Wellington, mary.tate@vuw.ac.nz

Ahmad Rabaa'i
Queensland University of Technology, a.rabaai@qut.edu.au

A Critical Evaluation and Comparison of Two Formative Measures of System Quality Using Criterion Variables

Azim Ali
School of Information Management
Victoria University of Wellington, New Zealand
azimali@gmail.com

Ahmad Raba'ii,
School of Information Technology
Queensland University of Technology (QUT),
Brisbane, Australia
a.rabaai@qut.edu.au

Mary Tate
School of Information Management
Victoria University of Wellington, New Zealand
Mary.tate@vuw.ac.nz

Meng Zhang
School of Information Technology
Queensland University of Technology (QUT),
Brisbane, Australia
Meng.zhang.ruc@gmail.com

Abstract

Evaluating the validity of formative variables has presented ongoing challenges for researchers. In this paper we use global criterion measures to compare and critically evaluate two alternative formative measures of System Quality. One model is based on the ISO-9126 software quality standard, and the other is based on a leading information systems research model. We find that despite both models having a strong provenance, many of the items appear to be non-significant in our study. We examine the implications of this by evaluating the quality of the criterion variables we used, and the performance of PLS when evaluating formative models with a large number of items. We find that our respondents had difficulty distinguishing between global criterion variables measuring different aspects of overall System Quality. Also, because formative indicators "compete with one another" in PLS, it may be difficult to develop a set of measures which are all significant for a complex formative construct with a broad scope and a large number of items. Overall, we suggest that there is cautious evidence that both sets of measures are valid and largely equivalent, although questions still remain about the measures, the use of criterion variables, and the use of PLS for this type of model evaluation.

Keywords
System quality, formative constructs, criterion variables, ISO 9126, IS-Impact

INTRODUCTION

Developing valid measures is one of the major challenges we face as scholars, and arguably one of the major advantages offered by 'scientific research' over the insights of reflective practitioners informed by experience. Nevertheless, different operationalisations of 'the same' construct abound, running the risk of meaning variance in the construct, where researchers "*may utter the same words, but the words have different meanings, so any logical comparison of their utterances is precluded*" (Curd et al. 1998, p 222). Conversely, it is not uncommon for two constructs with different names to be operationalised using synonymous indicators. This means that establishing the equivalence of two or more independently developed constructs (regardless of whether they have the same name) is fraught with difficulty. Establishing the equivalence of two formative constructs is even more challenging, as every item is expected to make a unique contribution to the overall meaning of the construct. Dropping a measure from a formative-indicator model may omit a unique part of the conceptual domain (Diamontopoulos et al. 2006). Similarly, the weight of formative indicators is partially determined by the number and nature of the dependent variables, so it has been argued that the meaning of a formative construct is partly determined by its nomological net (MacKenzie et al. 2011). This can result in meaning variance, where the meaning of the construct is unstable from model to model (MacKenzie et al. 2011). It could be argued that every formative measure is almost by definition a unique construction of its indicators and nomological net, and therefore cannot be generalised. However, this would preclude any direct comparison of formative measures. In this paper, we offer one of the first studies to critically evaluate and compare two formatively specified System Quality constructs from independent research streams – the IS-Impact model (Gable et al. 2008) and the ISO 9126 software quality standard. This is also one of the first studies to evaluate the discriminant validity of a formative information System Quality construct using criterion variables.

The aims of this paper are: 1) to operationalise and validate the ISO 9126 software quality standard as a formative index; 2) to critically evaluate and compare the ISO 9126 standard with the IS-Impact measure of System Quality; 3) to offer insights for users of both sets of measures; and 4) to derive insights for researchers wishing to conduct a meta-analysis of formative measures using criterion variables. Below we offer first a literature review of measurement issues regarding the two models we compare, followed by our research design for critically evaluating and comparing the models, followed by our results, a discussion and a conclusion.

LITERATURE REVIEW

The specification of constructs as formative or reflective has been well-rehearsed recently in research literature (Bagozzi 2011; Bollen 2011; Diamontopoulos 2011; MacKenzie et al. 2011). We offer here a brief review of some of the most important principles, as background to our research design. We then introduce the two models we use in the study.

Specification and Validation of Formative Constructs

Understanding of the use of formative variables in IS research has advanced considerably in the last 2-3 years. The basic characteristics of formative constructs have been well-rehearsed, i.e. 1) formative indicators characterise a set of distinct causes which are not interchangeable; each indicator captures a specific aspect of the construct's domain; 2) there are no specific expectations about patterns or magnitude of inter-correlations between the indicators; 3) formative indicators have no individual measurement error terms (i.e. they are assumed to be error-free in a conventional sense, the error is at the construct level or not at all); 4) while reflective measurement models with more than two indicators are identified and can be estimated, a formative measurement model, in isolation, is 'under-identified' and cannot be estimated (Diamontopoulos et al. 2001).

Two paths out of a formative model are required to identify it. This is often achieved by means of a MIMIC model. However, the interpretation and ontology of MIMIC models has been problematic, as the MIMIC model needs to do double duty as a composite index, which is formed out of its indicators, and a reflective latent variable with global reflective indicators that capture the meaning of the 'overall' latent construct, as shown in Figure 1, (Borsboom 2005; MacKenzie et al. 2011).

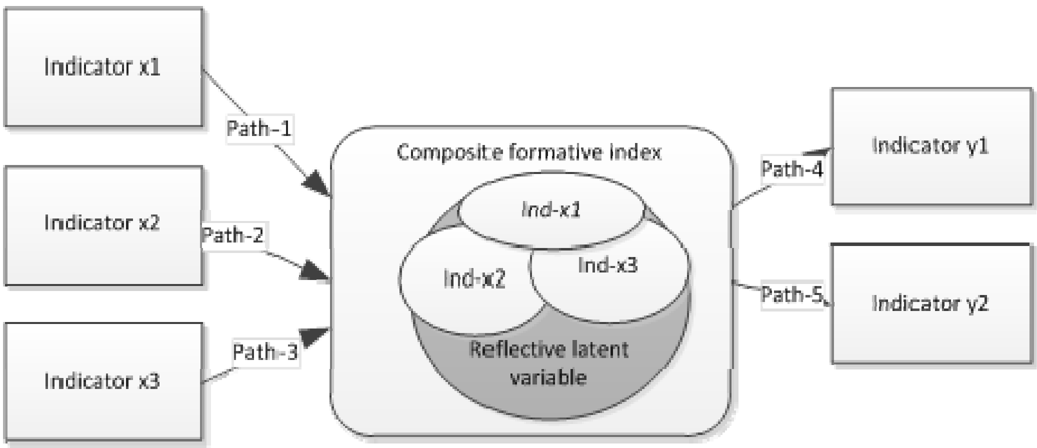

Figure 1: MIMIC Model with Unclear Ontology

The model represents three formative facets of a focal construct which is also measured with two criterion variables, or global reflective indicators (MacKenzie et al. 2011). This is clearer if it is represented as the statistically equivalent model of a composite formative variable predicting a single latent variable measured reflectively (Figure 2)[1].

These two variables now represent a composite System Quality index, and a user's perception of overall System Quality measured reflectively. *In this explanation, model validation "puts a premium on the content validity of the global reflective indicators, because they are essentially being used as criterion measures to establish the criterion-related validity of the formative indicators. Consequently, this test of validity will only provide useful information to the extent that the global reflective indicators faithfully capture the conceptual domain of the focal construct"* (MacKenzie et al. 2011, p 315). Validation criteria for this type of model include the strength of the regression co-efficient from causal indicator to latent variable, and the unique variance of the formative variable arising from the indicator (Bollen 2011). Discriminant validity exists when the items for the formative construct are more closely correlated with the posited criterion variables than for criterion variables of other constructs (Bollen et al. 1991; Diamontopoulos 2011; Diamontopoulos et al. 2001).

[1] This is also the usual way that MIMIC models are specified in PLS

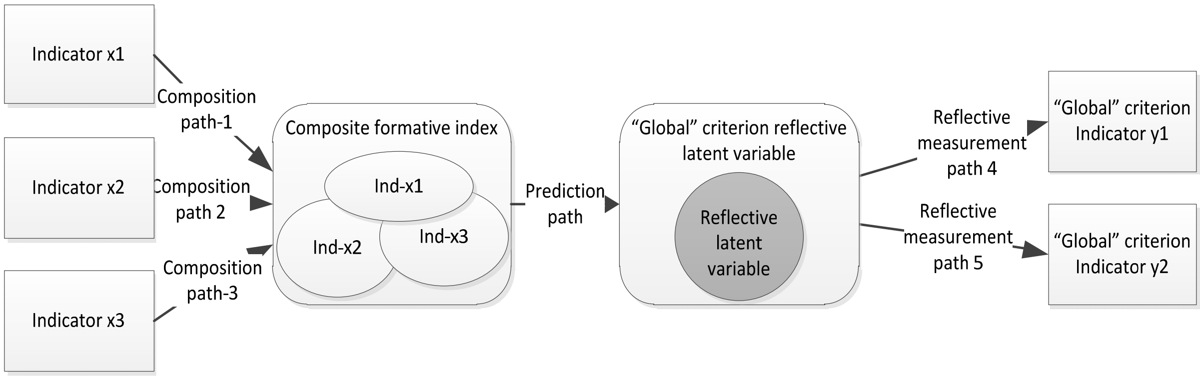

Figure 2: MIMIC Model Split into an Index and a Global Reflective Variable

In addition to providing for model identification, it has been suggested that the use of a MIMIC model with global reflective criterion measures reduces the risk of meaning variance or interpretational confounding. It is claimed that interpretational confounding "*is present to the extent that coefficients linking formative or reflective indicators with a focal construct significantly change depending on the other exogenous variables in the model (i.e. those caused by the focal construct)*" (Mackenzie et al. 2011, p 316), which is the case with formative variables. The formative variable is context specific and "*cannot be estimated without the incorporation of reflective dependent variables; the resulting context specificity may affect not only the co-efficients of the formative indicators, but the explained variance as well*" (Diamontopoulos 2011, p 344). Overall, the use of criterion variables has many advantages for identifying formative models, resolving the ontological issues implicit in MIMIC models, and reducing or eliminating the problem of interpretational confounding that is associated with formative constructs.

Model 1: IS-Impact

The provenance of the IS-Impact model is from the information systems management research community. The Delone and McLean model of information systems success (DeLone et al. 1992; DeLone et al. 2003) has been cited more than 7,000 times. The IS-Impact model and instrument, which is substantially based on the IS-Success model, is gaining traction in both academic and practitioner circles as a well-validated and parsimonious measure. The IS-Impact model draws on and extends the IS-Success model, and has been developed primarily in an ERP system context. The IS-Impact model is conceptualised as a formative, multi-dimensional index for use in an organisational context, with four success dimensions: 1) Individual Impact, a measure of the extent to which (the IS) has influenced the individual capabilities and effectiveness of key users s; 2) Organisational Impact, a measure of the extent to which (the IS) has promoted improvement in organisational results and capabilities; 3) Information Quality, which is a measure of the quality of (the IS) outputs, that is, the quality of the information the system produces in reports and on-screen; and 4) System Quality, which is a measure of the performance of the IS from a technical and design perspective. The IS-Impact model has been described as the most comprehensive and well-validated measure model for IS success (Petter et al. 2008).

Model2: The ISO 9126 standard

In contrast, the provenance of the ISO 9126 model is based in the practitioner and international standards communities. The International Standards Organisation (ISO) is the world's largest voluntary standards organisation. ISO standards exist for a wide range of products and services that are traded internationally, and aim to give state of the art specifications for products, services and good practice, helping to make industry more efficient and effective. Rather than being validated using rigorous academic methods, ISO standards could perhaps be better described as being validated by a crowdsourced Delphi method: "*Our standards are developed by the people that need them, through a consensus process. Experts from all over the world develop the standards that are required by their sector*"[2]. The ISO 9126 standard aims to cover 'Software Quality', and is effectively a taxonomy of quality dimensions and sub-dimensions. Software products are defined in a broad sense: encompassing executables, source code, architecture descriptions, and so on. "*ISO/IEC 9126 does not provide requirements for software, but it defines a quality model which is applicable to every kind of software*"[3]. The ISO 9126 model is intended to support both 'internal' (system specialists') views, and 'external' (system users') views. While the quality dimensions and sub-dimensions are generalisable across organisations, the ISO

---

[2] http://www.iso.org/iso/home.html

[3] *http://www.cse.dcu.ie/essiscope/sm2/9126ref.html*.

standards are intended to be implemented as 'attributes' that can be verified or measured in the software product. Specific attributes are not defined in the standard, as they can vary between different software products.

The generalised ISO 9126 model is depicted in Figure 3. The quality characteristics are defined as: 1) *Functionality* – set of attributes that bear on the existence of a set of functions and their specified properties. The functions are those that satisfy stated or implied needs; 2) *Reliability* – set of attributes that bear on the capability of the system to maintain its level of performance under stated conditions for a stated period of time; 3) ***Usability*** – set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users; 4) *Efficiency* – set of attributes that bear on the relationship between the level of performance of the system and the amount of resources used, under stated conditions; 5) *Maintainability* – set of attributes that bear on the effort needed to make specified modifications; 6) *Portability* – set of attributes that bear on the ability of the system to be transferred from one environment to another. While the ISO model is based in software engineering practice and the international standards community, it has also been adopted by researchers, primarily within the software engineering community, with some cross-over into general IS management research. We were unable to identify any previous studies that had evaluated the model empirically as dimensions of a generalisable formative software quality construct.

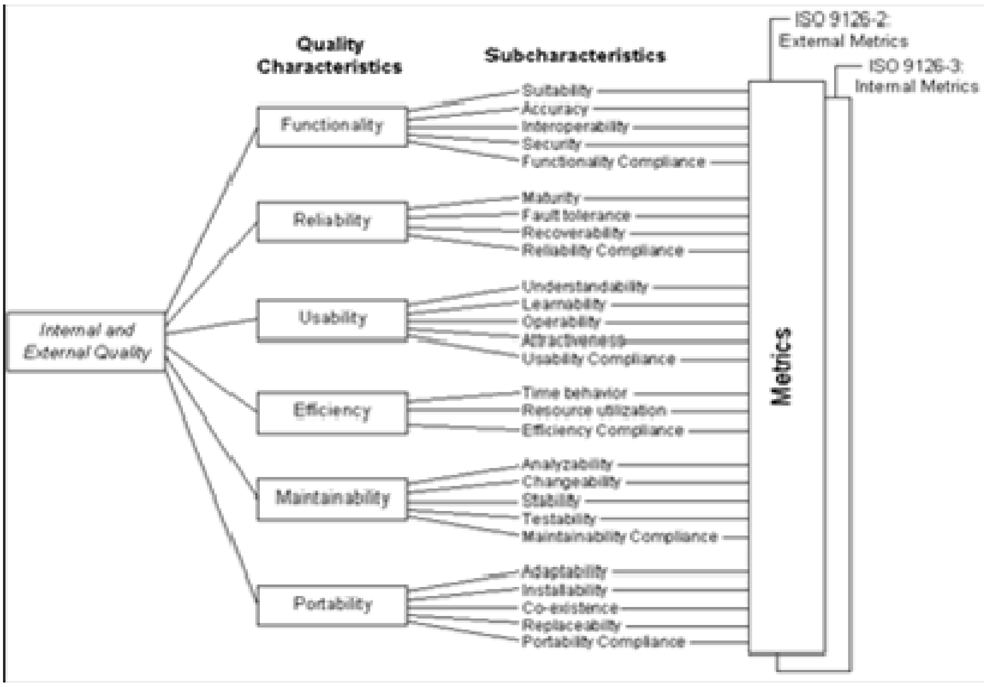

Figure 3: the ISO 9126 Model

Model Comparison

The two models were selected as good candidates for evaluating an 'overall' System Quality construct, as they were developed independently, both have a strong provenance, both appear to be formative and to have a comparable and overlapping scope (Figure 4). They can therefore be argued to be alternative specifications aiming at capturing the domain of a global System Quality construct. The equivalence between the two models is shown in Figure 4. Overall, the ISO 9126 standard covers a similar scope to the System Quality dimension of the IS-Impact model. The sub-dimensions and attributes for the ISO 9126 model are listed, with the IS-Impact items that cover a similar scope shown in brackets.

If the two models can be shown to be equivalent, and if they both have strong predictive power on a global System Quality construct, then 1) it provides triangulation and enhances the validity claims of both measures; 2) it provides confidence in the existence of a generalisable global System Quality construct that can be measured formatively (as opposed to assuming that measures of System Quality are unique to each context and organisation); 3) it suggests that either set of measures can be used effectively, depending on the organisation's preferences.

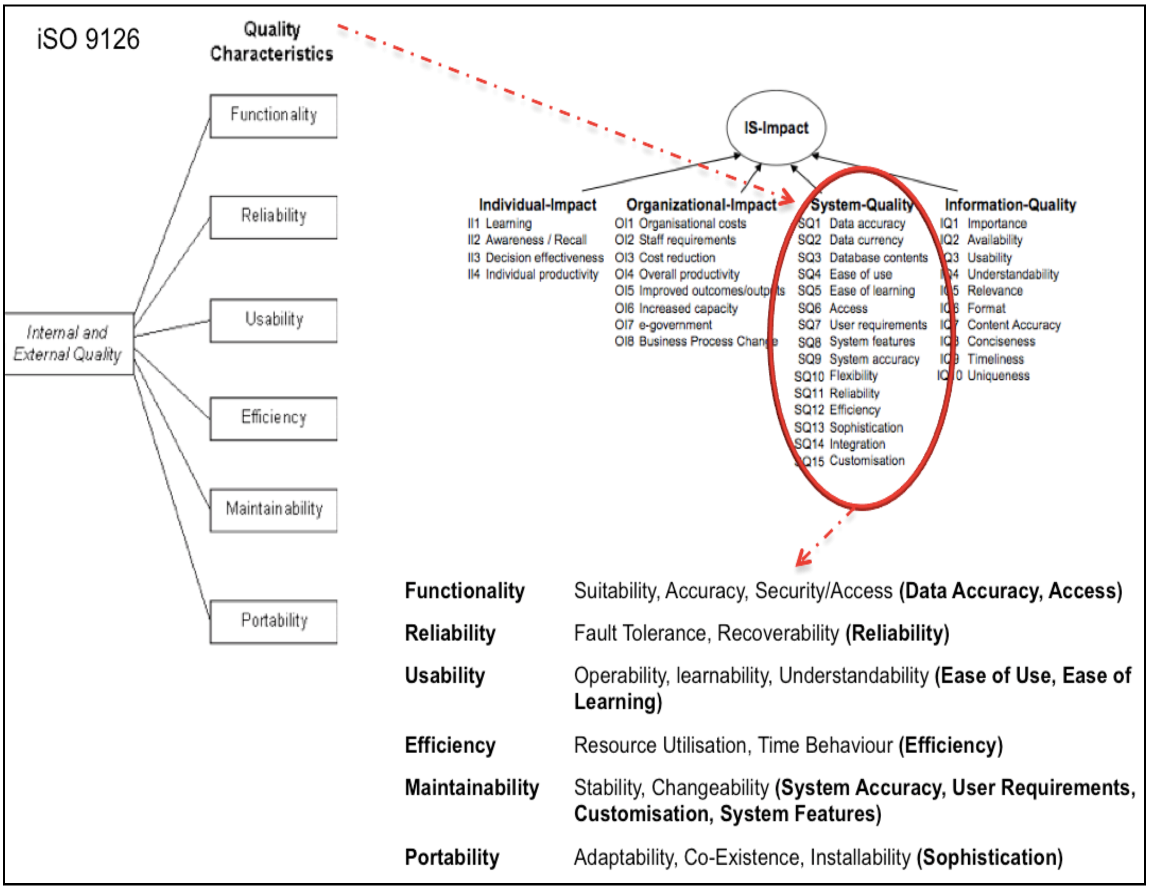

Figure 4: Comparison of the IS-Impact and ISO-9126 models

RESEARCH DESIGN

Survey and Sample

The constructs were measured using a 5-point Likert scale. Once the survey items were created to measure the constructs, a pilot test was completed. A random sample of 400 SAP users in a single large organisation was approached via email, from diverse modules such as Finance, Logistics, HR, Payroll and SRM. A total of 136 out of the 400 surveys were completed and returned, a respective response rate of 34%. Using participants within a single organisation ensured that the comparison would not be clouded by the use of different terminology for System Quality constructs that could occur within different organisational contexts.

Operationalisation of the Models

In our operationalisation of the IS-Impact model, we adopted the existing items and formative specification for System Quality. The IS-Impact model, however, had previously been validated without the use of criterion variables for the formative sub-dimensions. For the ISO 9126 quality standard, we were unable to find any well-established, generalisable operationalisations. However, we identified one study that developed a set of generalisable items (Padayachee et al. 2010). We used this study as the basis for our ISO 9126 System Quality items.

The external validity assessment for a formative construct refers to the extent to which the formative indicators actually capture the domain of the construct (Chin 1998). External validity can be assessed by regressing the formative construct on a reflective indicator of the same construct (Götz et al. 2010; Henseler et al. 2009). We developed global criterion variables for System Quality ("Overall, I believe the technical qualities of SAP will continue to support the organisation in the future", and "Overall, I believe SAP has excellent System Quality"). To allow us to test the discriminant validity of the formative construct, we also developed global criterion measures for overall organisational impact ("Overall, I believe SAP has been beneficial for the organisation", and "Overall, I believe SAP has made the organisation more effective"). We also added an overall summative question about the perceived success and impact of the SAP system ("Do you consider SAP to be a success in your organisation?").

Comparison

To assess and compare the validity of the constructs, Partial Least Squares (PLS) was conducted using SmartPLS 2.0 M3. First we evaluated the IS-Impact System Quality construct and the ISO 9126 software quality constructs. We specified each construct independently as a formative index predicting a global latent variable for System Quality with two reflective criterion variables in PLS.

Assessing formative measurement models raises the concern of whether each indicator contributes to the formative construct. Various statistical tests can be performed to determine whether an indicator should be included in the formative construct or not (e.g. Diamantopoulos et al. 2008; Diamantopoulos et al. 2001; Götz et al. 2010; Henseler et al. 2009; Petter et al. 2007; Urbach et al. 2010), such as assessing the degree of multicollinearity and assessing indicators' loadings on the dependent variable. High multicollinearity could mean that the formative indicator's information is redundant (Henseler et al. 2009). In order to check for multicollinearity, variance inflation factor (VIF) was calculated using SPSS (e.g. Götz et al. 2010; Henseler et al. 2009; Urbach et al. 2010). The ISO 9126 model was evaluated as a multi-dimensional formative construct, with formative sub-dimensions for functionality, reliability, usability, efficiency, maintainability and portability, and also as a first order formative index without the sub-dimensions.

We compared the two measures by combining the two sets of indicators into a single model to see if the predictive power on the global variable increased. This would indicate that the two sets of measures were additive, rather than alternative specifications of the same domain. We then evaluated the collinearity of the combined set of items. If the two sets of measures were equivalent measures of System Quality, then the items from the combined sets should demonstrate some collinearity. To determine the reliability of our criterion measures, we examined the cronbach's alpha for our reflective global variables for organisational impact and System Quality, and also the factor loading on each individual measure. To examine the external validity of the formative measure, we regressed the items on a global criterion indicator. The discriminant validity of the formative measure was assessed by evaluating the correlation of each of the individual System Quality items with the global criterion measures for both System Quality and Organisational Impact. Formative measures of System Quality should have a higher correlation with criterion measures of System Quality than with those for organisational impact (Bollen et al. 1991). Finally, based on some unexpected aspects of our results, we carried out some further analysis on the structure of our criterion measures.

RESULTS

Criterion Measures

The cronbach's alpha of the criterion measures for organisational impact and System Quality were .92 and .73 respectively. The loadings of the factors on each item were well above the recommended threshold of .7, indicating high reliability (Straub et al. 2004). We were therefore initially confident in our criterion measures.

Assessment of Each Measure of System Quality as Formative

We evaluated the models formatively using two criterion measures with PLS. Since the results were weaker than we expected, we explored the possibility that one of the criterion measures, ("Overall, I believe the technical qualities of SAP will continue to support the organisation in the future") was ambiguous. We re-evaluated the models with a single criterion measure. We also evaluated the ISO 9126 model as both a multi-dimensional formative construct including formative sub-dimensions of functionality, reliability and so on; and as a first order formative construct without sub-dimensions. Assessing formative measurement models raises the concern of whether each indicator contributes to the formative construct (Henseler et al. 2009). Various statistical tests can be performed to determine whether an indicator should be included in the formative construct or not (e.g. Diamantopoulos et al. 2008; Diamantopoulos et al. 2001; Götz et al. 2010; Henseler et al. 2009; Petter et al. 2007; Urbach et al. 2010), including: assessing the degree of multicollinearity and assessing indicators' weights as well as loadings.

While no minimum threshold values for formative indicator weights have been established (e.g. Rai et al. 2006), a high indicator weight suggests that the indicator is making a substantive contribution to the formative construct (Diamantopoulos 2006). Additionally, a significance level of at least 0.05 suggests that an indicator is relevant and valid for the construction of the formative construct (e.g. Urbach et al. 2010). Our models produced mixed results. The strongest predictive power was achieved with a single order formative construct with a single criterion variable. Even in this case, only half of each of the item sets showed statistically significant path co-efficients when the models were evaluated individually, and only one third of the items in the combined item sets. However, the ISO 9126 items collectively explained about 42% of the variance in the global variable (r-square 0.422), the IS-Impact System Quality items explained 49% of the variance (r-square 0.490) and the combined items explained approximately 52% (r-square 0.516).

If we had developed the measures from scratch, we might have been tempted at this point to assume that they were not good quality formative measures, as many of the items did not appear to be making a significant contribution to the overall measure. However, given the strong provenance of the measures prior to our study, we considered the possibility that this was not the only explanation for the relatively weak results. We also considered the possibility that there were issues with the global criterion variables, or with the nature of the tool used.

Comparing the Individual and Combined Item Sets

The presence of collinearity is one possible explanation for a relatively weak contribution from individual items. We examined each item set separately for collinearity, followed by the combined item set using the VIF statistic. A VIF>10 is often used as a threshold (e.g. Gefen et al. 2011; Götz et al. 2010; Henseler et al. 2009; Petter et al. 2007). Others recommend a more restricted heuristic such as VIF>4.0 (Garson 2008) to indicate that multicollinearity between items may be a problem. The VIF values are shown in Table 1 for each model and the combined item set. The majority of the single model scores were below the more restricted threshold of 4.0, suggesting that when considered individually, neither the ISO 9126 model nor the IS-Impact System Quality construct have significant issues with collinearity. When the item sets were combined, more than half of the combined items had a VIF above 4.0. This suggested that there was significant redundancy between the two item sets. This supports our assertion that the ISO and IS-Impact measures have a similar coverage and scope, and can be considered as valid and alternative formative measures of System Quality.

Table 1: VIF table

Item sets	ISO	Combined		IS-Impact	Combined
	VIF	VIF		VIF	VIF
Q1.9 SAP can perform the tasks required	2.628	3.658	Q1.28 All data within SAP is fully integrated and consistent	1.984	3.007
Q1.10 SAP produces results as expected	2.636	**4.900**	Q1.29 The SAP system is available 100% of the time	2.304	3.208
Q1.11 SAP can interact with other applications	1.819	3.376	Q1.30 Data from SAP often needs correction	1.978	3.038
Q1.12 SAP is compliant with standards	2.033	3.268	Q1.31 Data from the SAP is current enough	1.472	2.335
Q1.13 SAP prevents unauthorised access	1.991	3.139	Q1.32 SAP is missing key data	2.447	3.890
Q1.14 SAP is capable of handling errors	2.857	3.957	Q1.33 SAP is easy to use	*4.366*	**6.738**
Q1.15 SAP can resume working and restore lost data after failure	*4.672*	**7.342**	Q1.34 SAP is easy to learn	*4.922*	**5.958**
Q1.16 It is easy to comprehend how to use SAP	*8.656*	**15.061**	Q1.35 It is often difficult to get access to information that is in SAP	1.707	2.488
Q1.17 The user can use the system easily	*6.497*	**10.847**	Q1.36 SAP meets the units requirements	2.111	3.684
Q1.18 Using SAP requires little effort	*4.051*	**6.369**	Q1.37 SAP includes necessary features and functions	2.148	3.312
Q1.19 The interface looks good	2.499	3.498	Q1.38 The SAP user interface can be easily adapted to ones personal approach	2.783	**4.953**
Q1.20 The SAP system responds quickly	2.300	**4.804**	Q1.39 The SAP system is always up and running as necessary	2.731	**4.366**
Q1.21 SAP utilizes resources efficiently	3.592	**5.925**	Q1.40 The SAP system responds quickly enough	2.237	**4.439**
Q1.22 Faults in SAP can be easily diagnosed	3.824	**5.331**	Q1.41 SAP requires only the minimum number of fields and screens to achieve a task	2.797	**5.609**
Q1.23 SAP can be easily modified, corrected or improved	*4.368*	**5.725**	Q1.42 SAP can be easily modified, corrected or improved.	3.119	**5.312**
Q1.24 SAP can continue functioning if changes are made	2.507	**4.133**			
Q1.25 SAP can be tested easily	2.658	3.826			
Q1.26 SAP can be installed easily	1.843	2.627			
Q1.27 SAP can replace other applications within the organisation	2.395	3.689			

Discriminant Validity of the Formative Measures

We next compared the correlation between each of the formative items and the criterion variables for System Quality. These were compared with the correlation between the formative items and the criterion variables for Organisational Impact. The correlation with the criterion variables for System Quality was expected to be significantly higher. These results were equivocal. More than half the formative items (from both item sets) had weak or non-significant relationships with the criterion variables for System Quality. Further, nearly half the items had stronger correlations with the Organisational Impact criterion variables. This was unexpected, and led us to re-examine our global variables.

Post-hoc Analysis of the Criterion Variables

Following our finding of poor discriminant validity in the global variables, we carried out further analysis on the global variables. We carried out an exploratory factor analysis on the combined global items and the overall summative measure, and we examined the extent to which the global items for System Quality were correlated with the items for Organisational Impact, and with the summative SAP success measure. We found that all the items loaded strongly (above 0.7) on a single factor, which explained 70% of the variance in the items. This suggests that there was poor discriminant validity between the global variables (Straub et al. 2004).

When we examined the correlations between the items, we found that all the combined items were strongly correlated with each other and with the overall system success measure (Table 2). This suggests that respondents were unable to clearly distinguish the global constructs based on our operationalisations, and were responding to all questions based on their general attitude towards the success of the system. This finding also raises a number of interesting questions about the use of criterion measures in formative models.

Table 2: Correlations Between Criterion Variables

	Pearson Correlations				
	Summative	OI_CV_1	OI_CV_2	SQ_CV_1	SQ_CV_2
Summative	1	0.608	0.589	0.516	0.579
OI_CV_1	0.608	1	0.858	0.652	0.42
OI_CV_2	0.589	0.858	1	0.689	0.584
SQ_CV_1	0.516	0.652	0.689	1	0.574
SQ_CV_2	0.579	0.42	0.584	0.574	1
	all correlations were significant at the 0.01 level				

DISCUSSION

Model Comparison

We recall that we posited that if the two models were shown to be equivalent and to have strong predictive power on a global System Quality construct, this would triangulate both measures and enhance their validity claims, provide evidence for the existence of a generalisable formative System Quality construct, and suggest that either set of measures could be used effectively to measure System Quality. Both sets of measures appeared to be valid formative measures, had very similar predictive power on the global variable, and the combined set of measures had a high degree of multicollinearity, suggesting that up to half the measures in the combined set were potentially redundant. Further, the r-square value of the global System Quality variable barely improved when the two sets of measures were combined. This supports the assertion that the two models are equivalent, and likely valid formative measures of System Quality. However, the relatively low predictive power on the global reflective construct, and the issues with the criterion variables, leave some questions remaining.

Issues with Use of Criterion Measures

This study represented an initial attempt to define some criterion measures for System Quality, and a possible limitation of the study was that the criterion measures did not correctly capture the constructs. However, our experience may indicate wider issues with the use of global criterion measures for information System Quality, success, impact and satisfaction. These are all highly abstract and generalised, and it may be that in moving up to a higher level of abstraction, the constructs become very difficult to distinguish. Our post-hoc analysis on the criterion measures suggest that they all blur together into an overall attitude towards the system. Theories of attitude formation from social psychology suggests that this may indeed be the case; as we move away from more specific perceptions and beliefs into more generalised attitudes, measures become less reliable (Fishbein et al. 1975). It has also been suggested that more generalised measures are likely to capture a wide range of unmeasured causes – the respondent's 'experiential residue' from other experiences beyond the context of the study (Fishbein et al. 1975; Tate et al. 2009). This study represents one of the first attempts to validate an information System Quality construct using criterion variables. Clearly more research is needed to understand how respondents answer these questions, and to determine to what extent it is possible to develop suitably generalised criterion measures for information System Quality constructs that nevertheless have discriminant validity.

Issues with Statistical Tools

A further issue with conducting this analysis is the limitations of the statistical tools available. Cenfetelli and Bassellier (2009, p 695) argue that "*formative indicators essentially 'compete' with one another to be explanatory of their targeted construct. In this competition to explain variance, only a limited number of* *indicators will likely be significant while the others will be nonsignificant*". As such, formative constructs with a relatively large number of indicators will generally have many low indicator weights (Cenfetelli et al. 2009). Non-significant weight of a formative indicator may lead one to conclude that an indicator has no relationship with the formative construct it measures, hence permitting its exclusion from the model. However, MacKenzie et al. (2005, p 712) state that "*dropping a measure from a formative-indicator model may omit a unique part of the conceptual domain and change the meaning of the variable, because the construct is a composite of all the indicators*". In this study, we found that multicollinearity is unlikely to be a cause of low indicator weights, suggesting that the low weights may be an artefact of the large number of indicators used to assess System Quality. The large number of formative indicators has "*important implications for the statistical significance and the magnitude of each indicator's weight*" (Cenfetelli et al. 2009, p 694). Since these item sets have a very strong provenance individually and collectively, and System Quality is a complex and multi-faceted concept, it makes intuitive sense that a reasonably large item set would be required to capture all its facets. We suggest that the characteristics of the way formative constructs are calculated, in combination with the difficulty in developing valid generalised criterion measures for System Quality, may mean that is difficult to develop a formative item set and criterion measures for System Quality where all the items are significant and collectively explain a high proportion of the variance in the criterion measures. While we hope to see future studies which will improve significantly on our present results, there may be inherent limitations in this approach.

CONCLUSION AND FURTHER RESEARCH

This study makes a number of contributions to the understanding of Systems Quality. We add additional empirical support to the conceptualisation of System Quality as a formative construct. We empirically validate an operationalisation of the ISO 9126 software quality standard, which has extensive provenance in industry, and find that it is at least as good as a similar measure derived from the academic research community, and comparable in many respects. This comparison provides triangulation, additional confidence, and additional choice for academics and practitioners seeking measures of System Quality.

However, this study also raises a number of issues for researchers wishing to develop formative constructs for dimensions of information System Quality and success, who may be hoping to address the risk of interpretational confounding by the use of global criterion measures. It appears that there may be significant challenges in deriving global measures that are broad enough to cover the scope of the construct, and yet sufficiently reliable and specific. On the one hand, the metrics for our formative measures were borderline according to model quality heuristics. On the other hand, we selected both sets of measures have because they have an extensive, and independent provenance in academic literature, qualitative research and expert opinion. In addition, the ISO 9126 standard in particular has proven efficacy for users. They appear to be good quality measures. This raises the question of to what extent we should privilege the results of statistical techniques with known limitations over other sources of evidence. In our view, the search for valid System Quality measures is continuing.

REFERENCES

Bagozzi, R. "Measurement and Meaning in Information Systems and Organizational Research: Methodological and Philosophical Foundations," *MIS Quarterly* (35:2) 2011, pp 261-292.

Bollen, K. "Evaluating Effect, Composite, and Causal Indicators in Structural Equation Models," *MIS Quarterly* (35:2) 2011, pp 359-372.

Bollen, K., and Lennox, R. "Conventional Wisdom on Measurement: A Structural Equation Perspective," *Psychological Bulletin* (110:2) 1991, pp 305-314.

Borsboom, D. *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics* Cambridge University Press, Cambridge, England, 2005.

Cenfetelli, R. T., and Bassellier, G. "Interpretation of Formative Measurement in Informations Systems Research," *MIS Quarterly* (33:4) 2009, pp 689-707.

Chin, W. W. "Issues and Opinions on Structural Equation Modeling," *MIS Quarterly* (22:1) 1998, pp vii-xvi

Curd, M., and Cover, J. A. *Philosophy Of Science: The Central Issues* W.W. Norton & Company, New York, New York, 1998.

DeLone, W. H., and McLean, E. R. "Information Systems Success: The Quest for the Dependent Variable," *Information Systems Research* (3:1) 1992, pp 60-95.

DeLone, W. H., and McLean, E. R. "The DeLone and McLean Model of Information Systems Success: A Ten-year Update. ," *Journal of Management Information Systems* (19:4) 2003, pp 9-30.

Diamantopoulos, A. "The Error Term in Formative Measurement Models: Interpretation and Modeling Implications," *Journal of Modelling in Management* (1:1) 2006, pp 7-17.

Diamantopoulos, A., Riefler, P., and Roth, K. P. "Advancing Formative Measurement Models," *Journal of Business Research* (61:12) 2008, pp 1203-1218.

Diamantopoulos, A., and Winklhofer, H. M. "Index Construction with Formative Indicators: An Alternative to Scale Development," *Journal of Marketing Research* (38:2) 2001, pp 269-277.
Diamontopoulos, A. "Incorporating Formative Measures into Co-variance-Based Structural Equation Models," *MIS Quarterly* (35:2) 2011, pp 335-358.
Diamontopoulos, A., and Siguaw, J. A. "Formative Versus Reflective Indicators in Organisational Measure Development," *British Journal of Management* (17) 2006, pp 263-282.
Diamontopoulos, A., and Winklhoffer, H. "Index Construction with Formative Indicators: an Alternative to Scale Development," *Journal of Marketing Research* (38:2) 2001, pp 269-277.
Fishbein, M., and Ajzen, I. *Belief, Attitude, Intention and Behaviour: An Introduction to Theory and Research* Addison-Wesley, Reading, MA, 1975.
Gable, G., Sedara, D., and Chan, T. "Reconceptualising Information Systems Success: The IS-Impact Measurement Model," *journal of the Association for Information Systems* (9:7) 2008, pp 377-408.
Garson, G. D. "Regression: StatNotes from North Carolina State University," 2008.
Gefen, D., Rigdon, E., and Straub, D. "EDITOR'S COMMENTS: An Update and Extension to SEM Guidelines for Administrative and Social Science Research," *MIS Quarterly* (35:2) 2011, pp iii-xiv.
Götz, O., Liehr-Gobbers, K., and Krafft, M. "Evaluation of Structural Equation Models using the Partial Least Squares (PLS) Approach.," in: *Handbook of partial least squares: Concepts, methods and applications in marketing and related fields*, V. Esposito Vinzi, H. Wang, J. Henseler and W.W. Chin (eds.), Springer, Heidelberg, 2010.
Henseler, J., Ringle, C. M., and Sinkovics, R. R. "The use of Partial Least Squares Path Modeling in International Marketing," *Advances in International Marketing* (20) 2009, pp 277-320.
MacKenzie, S., Podsakoff, P., and Podsakoff, N. "Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques," *MIS Quarterly* (35:2) 2011, pp 293-334.
MacKenzie, S. B., Podsakoff, P. M., and Jarvis, C. B. "The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and Some Recommended Solutions," *Journal of Applied Psychology* (90:4) 2005, pp 710-730.
Padayachee, I., Kotze, P., and Van Der Merwe, A. "ISO9126 External System Quality Characteristics, Subcharacteristics and Domain Specific Criteria for Evaluating e-Learning Systems." South African Institute for Computer Scientists and Information Technologists, South Africa, 2010.
Petter, S., DeLone, W. H., and McLean, E. R. "Measuring Information Systems Success: Models, Dimensions, Measures, and Interrelationships," *European Journal of Information Systems* (17) 2008, pp 236–263.
Petter, S., Straub, D., and Rai, A. "Specifying Formative Constructs in Information Systems Research," *MIS Quarterly* (31:4) 2007, pp 623-656.
Rai, A., Patnayakuni, R., and Seth, N. "Firm PerformanceIimpacts of Digitally Enabled Supply Chain Integration Capabilities," *MIS Quarterly* (30:2) 2006, pp 225-246.
Straub, D., Boudreau, M.-C., and Gefen, D. "Validation Guidelines for IS Positivist Research," *Communications of the ACM* (13, article 24) 2004, pp 1-79.
Tate, M., and Evermann, J. "Perceptive Users with Attitudes: some Heuristics of Theorising," ICIS, San Francisco, 2009.
Urbach, N., and Ahlemann, F. "Structural Equation Modeling in Information Systems Research Using Partial Least Squares," *Journal of Information Technology Theory and Application* (11:2) 2010, pp 5-40.

COPYRIGHT