JITTA
JOURNAL OF INFORMATION TECHNOLOGY THEORY AND APPLICATION

CONTENT VALIDITY OF INSTRUMENTS IN IS RESEARCH

DEWI ROOSLANI TOJIB, Monash University
Department of Marketing, Building S, Caulfield Campus, Victoria 3145, Australia,
Email: Dewi.Tojib@buseco.monash.edu.au

LY-FIE SUGIANTO, Monash University
Clayton School of Information Technology, Building 63, Clayton Campus, Victoria 3800, Australia,
Tel. (+613) 9905 5803 Fax. (+613) 9905 5159, Email: Lyfie.Sugianto@infotech.monash.edu.au

ABSTRACT

This paper presents a literature review of content validity, summarizes qualitative and quantitative approaches for content validity assessment, and explores the extent to which IS researchers have content validated their developed instruments. An extensive review of scale development papers published in five major IS journals between 1989-2005 revealed that the proportion of published studies reporting content validity had indeed increased; however, qualitative assessment of content validity remains the preferred approach. To encourage the utilization of an alternative approach to content validity assessment, this paper describes a quantitative approach to evaluating the content validity of the B2E portal user satisfaction instrument.

INTRODUCTION

A valid instrument is one which measures what it is supposed to measure (DeVellis 2003). It also enables researchers to interpret variables and the relationships between variables in a more theoretically meaningful fashion (Bagozzi 1980). Therefore, the development of a valid instrument is the most fundamental aim of any instrument developer.

The issue of whether IS researchers sufficiently validate their instruments was initially raised by Straub (1989) who reported that only a few had devoted serious attention to validation. His work was replicated by Boudreau, Gefen, and Straub (2001) who investigated the extent to which IS researchers had responded to Straub's (1989) suggestion. Their findings revealed that the number of empirical studies reporting instrument validation had indeed increased since 1989. However, the number of studies reporting content validity was the lowest among all validity types investigated. This trend is supported by Straub, Boudreau, and Gefen (2004) who affirmed that content validity is indeed infrequently assessed in IS research. This is surprising as it should be the first type of validity to be established prior to examining other types of validity when developing instruments (Ebel 1967).

This paper responds to the call for more IS research on instrument validation, particularly on content validity. It presents theoretical literature on content validity including different approaches for assessing content validity. It also investigates the extent

Mihir A Parikh acted as the senior editor for this paper.

Tojib, D. R. and L.-F. Sugianto, "Content Validity of Instruments in IS Research," *Journal of Information Technology Theory and Application (JITTA)*, 8:3, 2006, 31-56.

to which IS researchers have considered content validity when developing new instruments. Finally, the use of a quantitative approach for assessing content validity of the B2E portal user satisfaction instrument is described.

CONTENT VALIDITY THEORY

Haynes, Richard, and Kubany (1995 p.238) defined content validity as 'the degree to which elements of an assessment instrument are relevant to, and representative of, the targeted construct for a particular assessment purpose'. Content validity can be established through the application of a two-stage process: development and judgement (Lynn 1986). The former is divided into three sequential steps: domain identification, item generation, and instrument formation (Carmines and Zeller 1979). Initially, the construct should be conceptually defined based on literature. A set of items is then generated and these items are arranged in a suitable sequence for the next stage of preparation. The judgement process, the primary goal of content validation, involves asking a specific number of experts to evaluate the validity of individual items and the whole instrument. The aim of this process is to retain the best items which are believed to adequately measure a desired content domain (Grant and Kinney 1992). In assessing experts' feedback, qualitative or quantitative approaches can be utilized.

The main difference between qualitative and quantitative approaches lies in the method used to determine when the finalised items - after modifying, deleting, or adding the original generated items - are adequate to measure the targeted construct. When applying qualitative methods, final decisions are generally obtained after all panel experts arrive at a consensus. Statistical calculation may be involved but it does not carry much weight in determining final decisions. On the other hand, quantitative approaches rely greatly on a particular statistical calculation. Items are eliminated if the statistical results are below the minimum threshold value. Experts may be asked to provide their opinion on the appropriateness and clarity of the items. To a certain extent, instrument developers may consider the experts' advices in revising items.

CONTRIBUTIONS

This study reviews 62 scale development articles reported in five esteemed IS journals and explores the content validity methods that have been utilized by IS researchers. Clearly, it makes a valuable and practical contribution as instrument developers can have exposure to alternatives in selecting methods and approaches to content validate their instruments. Considering the limited use of quantitative approach to assess content validity of a scale in IS research, this study illustrates the practicality of such approach when developing the Business-to-Employee Portal User Satisfaction Scale (B2EPUSS).

As shown in Table 1, a number of qualitative and quantitative approaches to content validation exist. Each approach has its own strengths and weaknesses and thus researchers should select the most appropriate method depending on the nature of their studies. The following paragraphs summarize the main characteristics of the two methods. Qualitative approaches may require content experts to physically meet one another as commonly occurs when the Q-sort technique is employed (Waltz, Strickland and Lenz 1991). The ability to communicate directly among experts may encourage in-depth discussions and any issues that may potentially create misunderstanding can be clarified without hesitation. Some may also argue that with the advances in telecommunication technologies, video conferencing or net meetings can replace the need to have the experts physically meet one another. However, not all experts may have the same compatible facilities, or those who are dispersedly located may be affected by time-zone differences. Thus, it may be difficult to arrange a time that is comfortable for all experts. On the other hand, when adopting quantitative approaches, an instruction document can be delivered through mail or e-mail. Experts can complete the tasks at their own convenience and return the document within the given time frame. Thus, the flexibility offered by quantitative approaches cannot be undervalued.

Content experts are generally expected to participate in the item evaluation process once when utilising quantitative approaches. In certain circumstances, when a second round is needed, quantitative approaches allow researchers to invite a different group of experts. However, researchers should ensure that the new group has the same level of expertise and the conclusion derived from the first round is clearly explained. This is considerably different from the Delphi technique, which requires the same panel of experts to be involved until all the evaluation processes are completed (Grant, Kinney and Guzzetta 1990). Employing the same experts is desirable as the group becomes more familiar with the issue under discussion; however, some experts, particularly those who have other commitments, may find the repeated exercises immensely tedious. As experts mostly participate voluntarily, they may discontinue at any time. Consequently, less exhaustive commitment required in the quantitative approaches may increase the participation rate of invited experts.

Once the statistical result achieves the recommended cut-off score and no significant modification is made to the instrument, the process of content validation of the instrument quantitatively can be concluded. This is quite different from the qualitative approach where consensus is commonly achieved after several rounds, with three rounds being the average when utilising the Delphi technique (Grant, Kinney, and Guzzetta 1990). The time may be extended if experts are asked open-ended questions, as researchers must transcribe their responses. It will take more time if clarifications of the experts' answers are needed. Quantitative approaches subsequently take less time compared to the Delphi or Q-sort technique.

It is apparent that quantitative approaches offer practicality in terms of time and cost. However, these techniques are not without limitations. Some techniques including the index of item-objective congruence and inter-observer agreement do not have a cut-off score, making such methods difficult to apply as the adequacy of content validity may be questionable. It is also

Table 1. Content validity: qualitative and quantitative approaches

Qualitative approach				
Method	**Description**	**Advantages**	**Disadvantages**	**Procedures**
Delphi	A survey method that is designed to structure group opinion and discussion (Goodman 1987).	Flexibility in data collecting process. The anonymity of experts and responses encourage true opinions that are not influenced by peer pressure or other extrinsic factors.	Time consuming.	A panel of anonymous experts is asked to evaluate a set of items in the form of questionnaires. Their anonymous responses are evaluated until a desired level of consensus is achieved. The number of rounds is varied with three rounds as being the average (Grant, Kinney, and Guzzetta 1990).
Q-sort	A technique that is often employed to assess the degree of similarity between different experts of certain issue at a given time (Waltz, Strickland, and Lenz 1991)	Relatively inexpensive. Result is fairly simple to handle and analyse.	Must be present in the sorting procedure.	Each expert is presented with index cards, each of which contains a descriptive statement (i.e. item). Each expert is then asked to read the card and place it into a specified number of different categories. Several sorting rounds are employed until consensus is reflected.

Table 1. Content validity: qualitative and quantitative approaches (Cont'd)

Quantitative approach				
Method	**Description**	**Advantages**	**Disadvantages**	**Procedures**
Content Validity Ratio (CVR)	A method proposed by Lawshe (1975) to assist researchers in forming decisions to retain or delete the item from the instrument.	Offers practicality in terms of time and cost. Quick and easy to perform.	Not reported.	Allows researchers to make forming decisions to retain or delete items from an instrument through the calculation of CVR.
Index of item-objective congruence	Alternative procedure proposed by Rovinelli and Hambleton (1977) for the analysis of judgements of experts.	Applicable regardless of the number of domains measured. Flexible as it only requires minimum of 2 experts. Offers practicality in terms of time and cost. Quick and easy to perform.	Lack of communicable standards for determining the adequacy of content validity. No recommendation on cut- off score.	Allows researchers to make decisions to retain or delete items from an instrument through the calculation of item-objective congruence index.
Content validity index (CVI)	This method is derived from the rating of the content relevance of the items on an instrument using a 4-point ordinal rating scale (Lynn 1986).	Flexible as requires only a minimum of 3 experts. Offers practicality in terms of time and cost. Quick and easy to perform.	The CVI may be inflated by chance. A 4-option scale is not universally used in CVI determinations.	Allows researchers to make decisions to retain or delete items from an instrument through the calculation of CVI.
Weighted mean score	This technique is based on obtaining experts opinion on the degree to which each item is indicative of a given construct/dimensions/sub-dimensions (Fehring 1987).	Offers practicality in terms of time and cost.	Not reported.	Allows researchers to make decisions to retain or delete items from an instrument through the calculation of weighted mean score.
Inter-observer agreement	It is obtained by calculating the proportion of number of experts assigning item to the expected domain over the total number of experts (Thorn and Deitz 1989).	Simple and intuitively obvious method of measuring agreement among experts. Offers practicality in terms of time and cost.	Spuriously high estimates as this method does not account for the contribution of chance agreements.	Allows researchers to make decisions to retain or delete items from an instrument through the calculation of % of agreement among experts.

important to note that either approach may be used to assess content validity. One approach might be better than the other depending on the type of data collected, the purpose of study, and the availability and location of the experts, as well as the time constraint. Therefore, researchers should consider both approaches and carefully select the most

appropriate approach, either qualitative or quantitative, for their research study.

A REVIEW OF CONTENT VALIDITY IN IS RESEARCH

In order to determine the extent to which instruments developed in IS research have been content validated, a review and analysis of the articles from five journals: *MIS Quarterly (MISQ)*, *Information and Management (IM)*, *Journal of Management Information Systems (JMIS)*, *Management Science (MS)*, and *Information Systems Research(ISR)* were conducted. These were the same journals used by Boudreau, Gefen and Straub (2001) in their studies. To a certain extent, this study can be viewed as an expansion of their studies; however, we focus only on content validity and, since the above-mentioned journals are still considered as top tier journals in IS, we decided to base our study on the same journals.

The qualifying criteria used to sample the articles were: 1) Published articles within the period between January 1989 and December 2005. The main objective of this study is to explore the extent to which IS researchers have content validated their instruments; hence, it was necessary to examine articles from the early years. The year 1989 was chosen as the starting period because the issue of whether IS researchers were sufficiently validating their instruments was initially raised in that year (Straub 1989). 2) We chose articles describing the development of an instrument to measure a particular phenomenon in IS or to validate proposed IS model. We decided to restrict the sample articles to instrument development studies for several reasons. Firstly, as content validity is the first validity type to be established when developing a new instrument (Ebel 1967), we expected to obtain required information from instrument development papers. Secondly, IS researchers often use existing instruments that have been previously validated for theoretical and practical reasons (Boudreau, Gefen and Straub 2001). The former involves assessing the extent to which the existing instruments are applicable or comparable to their studies. The latter is concerned with the efficiency issue. Existing instruments, if developed rigorously at first, must have undergone at least one validation cycle. If they were initially found to be reliable and valid, IS researchers may find it trustworthy to use them for other studies. Furthermore, researchers often feel that they cannot afford the time to validate their instruments (Straub 1989, Boudreau, Gefen and Straub 2001). They may also find it unnecessary to perform the same process repeatedly. Hence, in practice, although it is highly recommended to validate any instruments (new or existing ones) utilized in IS studies, it is quite often observed that researchers do not re-validate the instruments, particularly content validity which generally takes an additional substantial amount of time. Therefore, based on this reasoning, we believe that the sample articles captured adequately the purpose of our study. 3) Articles performing the survey research method. This last criterion was imposed because the administration of the instrument is generally taken in the form of a survey (Boudreau, Gefen, and Straub 2001).

A total of 62 articles were used in the literature analysis. Among the articles reviewed, 9 articles originated in ISR, 3 in MS, 12 in JMIS, 10 in MISQ, and 28 in IM. These articles were collected from the period of 1989 – 2005 and were analyzed in three time periods: 1989-1996, 1997-1999, and 2000-2005. The year 1989 was selected as the starting point because it was the milestone when Straub initially raised the issue of instrument validation in IS research. The period of 1997 - 1999 was established as the middle point since Boudreau and his colleagues covered those periods when they replicated Straub's (1989) work in 2001. The period of 2000-2005 attempts to explore what has changed, if anything, in the intervening years.

In keeping with the recommendation of Lynn (1986), the extent to which each article had gone through the development and judgement stages of content validation process was investigated. The development stage is the stage where the domain of construct is identified and items are generated to measure the construct. On the other hand, the judgement stage commonly involves a panel of experts who are required to assess the relevancy and validity of the items. For each

article, the method used to identify the domain of construct, to generate items, and to evaluate the judgement process was identified. Furthermore, as stated by Alreck and Settle (1995), pre-testing the instrument often leads to content validity. A pre-test is a preliminary trial of some or all aspects of the instrument in order to examine some possible difficulties that may be encountered by the potential respondents when filling it out. A pre-test can be given by asking a number of people, who have characteristics similar to those of the prospective respondents, to evaluate the clarity of the instructions and items, to identify any items that may have different interpretations/meanings, to comment on the formatting and questionnaire design, or even to record the time taken to complete the questions. In this study, whether or not the instrument was refined by a pre-test was also evaluated. The content analysis of past literature is presented in Appendix 1 and the key summary of our findings is presented in the following paragraphs.

Our findings show that all sample articles have reported the development stage of the content validity process. A literature review has been the favoured method for identifying the research domain as appeared in all published articles from 1989 to 2005. Items for measuring the research domain have been extensively derived from relevant literature and existing instruments. Only a small portion of past studies reported item generation using other methods. Four articles reported the utilisation of focus groups, two articles used the interview method, and one article reported the use of experts/practitioners. Surprisingly, these articles all appeared between the years 2000 and 2005. This shows that new approaches to item generation are emerging despite the current predominant methods such as literature and existing instruments.

In the case of the judgement stage, 34 articles in the period between 1989 and 2005 reported that the generated items were reviewed by a panel of experts or judges. However, only 13 articles specifically discussed the methods used to retain, delete, or refine the items. Nevertheless, as shown in Table 2, the proportion of published studies reporting the judgement stage increased gradually, mainly in the period of 2000-2005.

It is also apparent that IS researchers have long acknowledged the qualitative and quantitative approaches to content validation. 13% of the sample articles within the period of 1989 - 1996 reported the use of a qualitative approach while 7% reported the use of a quantitative approach. The finding also revealed that the utilisation of the qualitative approach decreased slightly between 1997 and 1999; however, its utilisation increased slightly between 2000 and 2005. Unfortunately, this increase was not mirrored by the quantitative approach, the use of which has, in fact decreased.

A careful investigation of the sample articles reporting specific methods for content validation revealed that the Q-sort technique (cited in eight articles) seems to be in favour compared with the Delphi technique (cited in two articles). On the other hand, the content validity ratio (cited in two articles) and inter-observer reliability (cited in one article) were the methods used in the quantitative approach.

Table 2. Percentage of Studies Reporting Methods Utilized in Judgement Stage

Period	No. of Articles	Judgement Stage				Pre-Test
		Qualitative Approach	Quantitative Approach	Method Not Specified	Not Reported **	
1989-1996	15	13%	7%	20%	60%	53%
1997-1999	8	12.50%	0%	37.50%	50%	37.50%
2000-2005	39	18%	5%	38.50%	38.50%	46%

* Method Not Specified includes articles claiming that expert judgement was performed but methods used were not specified

** Not Reported includes articles not claiming the use of expert judgement to evaluate items

Of these techniques, the Q-sort seems to be the most utilized as it was cited in six articles published between 2000 and 2005. Surprisingly, four of these followed the Q-sort procedure proposed by Moore and Benbasat (1991). It can then be concluded that when developing instruments, IS researchers prefer qualitative to quantitative approaches.

Regarding pre-tests, our findings revealed that 29 studies in the period between 1989 and 2005 conducted pre-tests on their initial instruments. As shown in Table 2 above, compared to the 1989-1996 time period, the proportion of studies reporting pre-tests in the period of 1997-1999 decreased slightly. However, this trend increased somewhat between 2000 and 2005. Nevertheless, it is interesting to see the inverse relationship between the judgement stage and the pre-test. In other words, there is a tendency for studies reporting the judgment stage to not describe pre-tests and vice versa. To further investigate this phenomenon, we explored studies reporting the utilisation of content experts only, pre-tests only, and both methods. The results are presented Table 3.

Although the proportion of published studies reporting both the judgement stage and pre-tests increased considerably in the period of 2000-2005, Table 3 clearly shows the existence of an inverse relationship between the judgement stage and pre-tests. A number of possible explanations for this exception are as follows. Firstly, many researchers often ask experts to provide additional comments on the clarity of the items. Thus, researchers may find it unnecessary to pre-test the instruments. Secondly, researchers may assume that the judgement stage is the same as a pre-test. While both methods involve asking people's opinion about the developed instruments, the invited people and the tasks assigned to the participants are different. The judgement stage requires experts in a particular field of study to evaluate the items. The major aim is to ensure that the developed instrument measures the intended construct appropriately. Ordinary people can be used for the pre-test and they are generally asked to evaluate the clarity of the instructions, the comprehensibility of the items, and the design of the questionnaires. Thus, these two procedures are different in nature and they are not interchangeable. Consequently, studies reporting pre-tests but not the judgement stage should justify the content validity of their instruments.

While Lynn (1986) emphasizes the importance of expert judgment to achieve content validity, a detailed review of sample articles within the period of 1989 and 2005 revealed that there are four studies claiming the accomplishment of content validity based exclusively upon relevant literature reviews and existing scales. The major concern here is the extent to which the researchers feel confident that the generated items capture the essence of the construct domain. The confidence level can be increased with the utilisation of content experts as they will provide useful insights into the completeness and appropriateness of the items. Hence, the judgment stage is necessary to justify the content validity claim of the instruments.

It is also of interest to observe whether the content validity process reported in published studies differs from journal to journal. Table 4 shows the overall proportion of studies reporting the judgement stage and the pre-test in each of the sample journals. It can be noted that for the period 1997 to 1999, both JMIS and MS did not publish any instrument development papers and hence data cannot be projected. There are two possible explanations for this. Firstly, IS researchers are generally encouraged to utilize existing

Table 3. Percentage of Studies Reporting Judgement Stage and Pre-test

Period	No. of Articles	Judgement Stage Only	Pre-Test Only	Both Methods	Not Reported*
1989-1996	15	20%	33%	20%	27%
1997-1999	8	37.50%	25%	12.50%	25%
2000-2005	39	33%	18%	28%	21%

* Not Reported includes articles reporting neither judgment stage nor pre-test

instruments and to develop new ones only if the existing ones are not applicable for their studies. This may inhibit the growth of scale development research in IS and thus not many papers were published in this area. Secondly, although the IS leading journals were sampled, they may not be the best publication outlets for instrument development types of research. For instance, MS may have different missions and objectives from IM as the former published only three instrument development articles while the latter published over 25 articles between 1989 and 2005.

As shown in Table 4, compared with the periods of 1989-1996 and 1997-1999, the proportion of sample articles reporting the judgement stage increased for most journals between 2000 and 2005. Such a positive improvement shows that IS instrument developers have taken the validation issue into greater consideration, particularly in the area of content validity. Table 4 also shows that, compared with the period of 1997-1999, there is a slight decrease in the proportion of studies reporting the judgement stage in MISQ for the period of 2000-2005. This trend is similar to ISR, where a slight decrease was detected in the period of 1997-1999 compared with the period of 1989-1996. This finding is not conclusive as there was only one paper published in each journal within these periods. Nevertheless, it was surprising to find that while IM published the largest number of instrument development papers in the period of 2000 - 2005 (n=20), the proportion of studies reporting the judgement stage was quite low compared with the other journals.

Knowing that the proportion of studies reporting the judgement stage increased in most journals, it is of interest to find out which journal(s) reported the content validity stage in a more rigorous way. As shown in Table 5, most published studies across the five journals have described the judgement stage process; however, only articles published in IM, JMIS, and MISQ reported the utilisation of a specific approach to content validation between 2000 and 2005. Furthermore, Table 6 shows that, compared with the other sample journals, a high proportion of articles published in JMIS reported a thorough content validation process, including the utilisation of the qualitative or quantitative approach to the content expert judgment process, as well as pre-tests. Again, considering that IM published the largest number of instrument development studies, it is quite surprising to see that it has the highest proportion of studies not reporting either the judgement stage or pre-tests more recently between 2000-2005.

When reviewing the sample articles, it was also found that sections describing the expert judgement process and pre-test

Table 4. Percentages of Studies Reporting Judgement Stage and Pre-test by Journal

	Judgement Stage			Pre-Test		
Journal	**89-96**	**97-99**	**00-05**	**89-96**	**97-99**	**00-05**
IM	50%	25%	45%	25%	50%	40%
JMIS	40%	-	86%	60%	-	100%
MISQ	25%	100%	80%	75%	100%	20%
ISR	100%	67%	80%	100%	0%	20%
MS	0%	-	50%	0%	-	50%

Table 5. Percentages of Studies Reporting Specific Content Validation Methods by Journal

	Judgement Stage											
Journal	**Qualitative**			**Quantitative**			**Method Not Specified***			**Not Reported**		
	86-96	**97-99**	**00-05**	**86-96**	**97-99**	**00-05**	**86-96**	**97-99**	**00-05**	**86-96**	**97-99**	**00-05**
IM	0%	0%	5%	0%	0%	5%	50%	25%	35%	50%	75%	55%
JMIS	20%	-	43%	20%	-	14%	0%	-	29%	60%	-	14%
MISQ	0%	100%	60%	0%	0%	0%	25%	0%	20%	75%	0%	20%
ISR	100%	0%	0%	0%	0%	0%	0%	67%	80%	0%	33%	20%
MS	0%	-	0%	0%	0%	0%	0%	0%	50%	0%	0%	50%

* Method Not Specified includes articles claiming expert judgement was performed but methods used were not specified

Table 6. Percentages of Studies Reporting Judgement and Pre-test Methods by Journal

Journal	Judgment Stage Only			Pre-Test Only			Both Methods			Not Reported		
	86-96	97-99	00-05	86-96	97-99	00-05	86-96	97-99	00-05	86-96	97-99	00-05
IM	50%	25%	30%	25%	50%	25%	0%	0%	15%	25%	25%	30%
JMIS	20%	-	0%	40%	-	14%	20%	-	86%	20%	-	0%
MISQ	0%	0%	60%	50%	0%	0%	25%	100%	20%	25%	0%	20%
ISR	0%	67%	60%	0%	0%	0%	100%	0%	20%	0%	33%	20%
MS	0%	-	50%	0%	-	50%	0%	-	0%	100%	-	0%

were generally short and not substantially described. The fact that an increasing number of journals limit the length of submissions may constrain the researchers to shorten this section. This is possibly one of the reasons that IS researchers have not adequately addressed the issue of content validation in instrument development studies. There are two suggestions to address this issue. Firstly, since the development of an instrument requires a number of sequential stages to be performed, journal editors should be flexible about the number of submitted and published pages for papers in this area. Secondly, researchers could write an article focusing on the content validation stage of their instrument development process. While it is not common for such a paper to be published in the IS field, it is widely published in other streams, for instance, Barbara (1997), Beitz and Rijswijk (1999), and Zuzelo, Inverso, and Linkewich (2001).

Based on our findings, it can be concluded that there is advancement in the proportion of studies reporting the content validity stage within the period of 2000-2005 compared with the period of 1989-1999. It is also apparent that only a small proportion of studies reported the use of qualitative and quantitative methods for item refinement in the expert judgement stage. The use of the qualitative technique seems to be more favoured, indicated by more frequent use of the Q-sort technique. The shortage of IS literature discussing the quantitative approach to content validation may be the reason for this trend. Therefore, to encourage IS researchers to consider the quantitative approach as an alternative method for assessing content validity, the next section describes how this approach can be employed when validating a Business-to-Employee (B2E) portal user satisfaction instrument.

CONTENT VALIDATION OF THE B2E PORTAL USER SATISFACTION (B2EPUS) INSTRUMENT

Business-to-Employee (B2E) portals have been widely implemented in various organisations. These portals are specifically developed to support the access and availability of customized and personalized information for employees. While benefits of such portal implementation have been significantly promoted in business literature, there is no theoretical framework that can guide organisations in determining the extent to which their portal implementations are successful. Our proposed approach is to measure user satisfaction with B2E portals. Since existing user satisfaction instruments in IS research cannot be applied to measure the B2EPUS construct, we decided to develop a new one.

When developing the B2EPUS instrument, we found it necessary to assess the content validity of the instrument as it is the first type of validity to be established prior to examining other types of validity (Ebel 1967). In our study, the major purpose of content validity assessment is to ensure that all items generated to measure the B2EPUS construct are relevant to, and representative of, the construct. If the items are proved to be content valid, the researchers should feel confident that the instrument is able to measure the investigated construct.

This section discusses a quantitative approach to content validation assessment for the Business-to-Employee (B2E) portal user satisfaction instrument. As the aim of this section is to demonstrate the use of the quantitative approach in content validation of an instrument, some parts that are considered unnecessary to discuss in detail have been omitted. Detailed explanation of various phases of this research can be found in Tojib

and Sugianto (2006), Tojib and Sugianto (2007), Sugianto and Tojib (2007). Following Lynn (1986), the content validation procedures were divided into two stages: development and judgement.

Stage 1: development stage

The preliminary stage in developing an instrument includes conceptually defining the construct domain to be measured. In our study, the domain of the B2EPUS construct, which consists of nine dimensions, was identified from user satisfaction and B2E portal literature (shown in Figure 1). In keeping with the recommendation of Lewis, Templeton, and Byrd (2005), the sub-parts of each dimension were then determined and each sub-part was then converted into an item on the B2EPUS instrument. Whenever appropriate, items were derived from existing general user satisfaction instruments or relevant IS literature. Initially, 47 items were generated. These items were then refined in the subsequent analysis: the judgement process.

Stage 2: judgement process

As content validity mostly depends on the content experts' judgement, it is very crucial to select the right experts. For the purpose of content validating the B2EPUS instrument, the criteria for selecting the content experts were derived from the guidelines proposed by Grant and Kinney (1992) that is, 1) they must hold a PhD qualification or be PhD candidates; and 2) they should actively conduct research in the domain of interest or have professional experience in B2E portal development. The procedures for recruiting the experts, and collecting and analysing the obtained responses are described in the following section.

First round judgement process

Personalized email invitations were sent to the 32 academics, two doctorate students, and the members of EDUCAUSE portal mailing list. Each email outlined the reasons that they were selected, the purpose of the study, and a request for their participation in the study. For those who agreed to participate, a specific, structured instruction document was emailed to each of them, outlining in detail the tasks to be completed. The response time was one month and within this time, eleven responses were received from the content experts. Five responses were excluded because some important sections in the document were not completed. Hence, only six responses could be included for further analysis and this were considered adequate as it met the minimum requirement recommended by Lynn (1986).

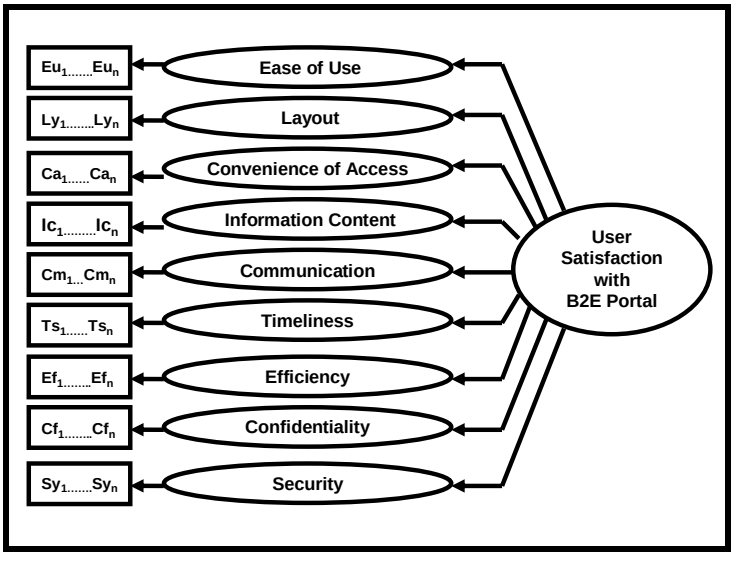


Figure 1. Conceptual Model of User Satisfaction with B2E Portal

They were asked to complete two tasks. In the first section, they were presented with nine dimensions of the B2EPUS construct, each with a definition. They were then asked to rate the importance of each dimension using a common 5-point rating scale (1 represents 'Extremely Unimportant', 2 represents 'Somewhat Important', 3 represents 'Neutral', 4 represents 'Somewhat Important', and 5 represents 'Extremely Important'). All mean values of the importance rating given for each dimension were greater than 3, suggesting that all dimensions were essential for the construct.

In the second section, the experts were presented with 47 items that had been generated to measure the nine dimensions or factors. They were to match each item with its nominated dimension and rank the relevancy of the item to the assigned dimension. When rating the relevancy of the item to each factor, they used a 5-point rating scale (1 represents 'not at all', 2 represents ' very little', 3 represents 'somewhat', 4 represents 'well' and 5 represents 'very well'). The 5-point rating scale was used as Fehring's (1987) weighted score method was adopted to analyze the data. These tasks were to identify the extent to which the experts correctly assigned each item to the expected dimension; and to investigate how well the experts thought that each item fitted the definition of each corresponding dimension.

When employing Fehring's (1987) method, the weighted ratios for each item were calculated. The weights are as follows: 0 weighting for rating 1, 0.25 weighting for rating 2, 0.50 weighting for rating 3, 0.75 rating for weighting 4, and 1.00 weighting for rating 5. The weights are provided to ensure that the maximum value for the total score is 1.0 and an item considered irrelevant (to the tested constructs) by the experts will be discarded (Fehring 1987). Weighted scores for each item were averaged to produce ratios of importance. The cut off value for this method was the average weighted ratios less than or equal to 0.50 (Fehring 1987). The results showed that there were fifteen items with average weighted ratios less than 0.50. Hence, our final decision was to exclude all fifteen items from further analysis.

There were two modifications suggested by the experts. First, *Confidentiality* was combined with *Security* as they thought that these two dimensions are similar. Second, *Efficiency* was renamed *Usefulness*. Furthermore, it was decided to add five more items to the instrument for the following reasons. Firstly, after deleting 15 items, *Timeliness* and *Information Content* were affected. There was only one item left for *Timeliness*. As a single item measuring one dimension may incur threats of unreliability (Hinkin and Schriesheim 1989), it was decided to add two more items. Three new items were added to measure dimension *Information Content* because the remaining items did not adequately tap the sub-dimensions initially determined.

Second round judgement process

Although the quantitative approach to content validity usually does not require second round processes, we believed that the second round process for our study was necessary because some dimensions and items were revised and a number of new items were added. Furthermore, we would like to ensure that there is sufficient rigor in this process, bearing in mind that many scale development practices include measures that lack content validity in the item development stage (Hinkin 1995).

Similar to the first round, a personalized email invitation was sent to two doctorate students, five portal project leaders, one portal researcher, and one portal practitioner. Specific, structured instructions for these reviewers were then emailed to each of them, outlining in detail the tasks that were to be completed. They were asked to complete the tasks using Microsoft Word and to email the completed document to the researcher. Within one month, five returned the responses, two of whom failed to fill out some important sections. Hence, only three responses were retained and these met the minimum number of experts proposed by Lynn (1986).

We decided to simplify the tasks, bearing in mind that content experts from the first round considered the assigned tasks were too exhaustive. This may be the reason that a number of content experts did not complete the tasks as specified in the instruction

document. The modified task was more appropriate to be analyzed by calculating the Content Validity Index (CVI). Thus, the instruction document was modified accordingly.

In the second round, content experts were presented with the eight dimensions along with their individual definitions and corresponding items. They were required to evaluate the extent to which each item is relevant to the assigned factor using a 4-point rating scale (1 represents 'irrelevant', 2 represents 'somewhat relevant if phrasing is profoundly adjusted', 3 represents 'relevant with some adjustment as to phrasing', and 4 represents 'very relevant') As stated by Lynn (1986), the 4-point rating scale is preferable for two reasons: 1) it does not include the ambivalent middle rating; and 2) it provides sufficient delineated information upon which to calculate a meaningful CVI. The aim of this process is to ensure that all the revised items are relevant to the designated dimensions and to identify whether there are items that need to be further revised.

Following Lynn (1986), the CVI for each item and for the whole instrument was calculated. The CVI value for each item was determined by the proportion of experts who rated it as content valid (a rating of 3 or 4). As there were only three responses, all three experts had to give a rating of 3 or 4 in order to retain the item. Three items were rated 1 or 2; thus, these items were removed from the scale. The CVI for the entire instrument was 91.89% (that is, 34 out of 37 items were judged content valid by the content experts). The CVI value clearly exceeded the expected minimum CVI of 0.80 (Davis 1992) and thus showed an adequate content valid instrument.

We performed further item analysis for the remaining 34 items. We deleted redundant items as they did not add to construct validity (Nunnally and Bernstein 1994) and they made the instrument unnecessarily long (Gatignon, Tushman, Smith, and Anderson 2002). This investigation led to the exclusion of six more items. To conclude, after a systematic and thorough two-stage content validation process, the remaining 28 items were considered to be content valid. The scale refinement process is shown in Appendix 2.

DISCUSSIONS

This paper addresses two major topics related to content validity. Firstly, through a review of 62 instrument development articles, we have explored how rigorously IS researchers have content validated their instruments and which approach they most often utilize for the content validity assessment. Secondly, a case example describing the procedure for assessing content validity quantitatively when developing the B2EPUS instrument has also been presented. Further detailed discussion on each topic is described separately in the following paragraphs.

Content Validity in IS Research

A review of 62 sample articles on instrument development studies within the three periods of analysis (1989-1996, 1997-1999, 2000-2005) revealed an increase in the number of IS researchers who content validated their instruments. All studies reported in the sample articles accommodated the development stage of the content validation process. Compared with the period of 1989-1999, there was an increase in the proportion of studies reporting the judgement stage in the period of 2000-2005. However, 38.5% of the sample articles during this period still did not report the utilisation of content experts in validating the instrument items and hence, the assessment of content validity in developing new instruments has not yet reached a satisfactory level. Furthermore, more than 50% of studies reporting the judgement stage within the period of 2000-2005 did not report specific methods; that is, whether the qualitative or quantitative approach was used. The qualitative approach, particularly the Q-sort technique, was the preferred method used for identifying the best items to retain in the instruments. One possible explanation for this trend is the fact that Moore and Benbasat (1991) initially provided a thorough description of the Q-sort method, which may have attracted other researchers to use the same method. This may also explain the fact that the quantitative approach to content validity has not gained wide attention among IS researchers, particularly as none of the sample articles thoroughly reported a quantitative content validation approach.

A summary of content validation approaches presented in Table 1 attempts to compensate for the lack of literature on the methods and means of assessing content validity in IS research. Although this summary demonstrates that content validity can be assessed qualitatively or quantitatively, both approaches can assist researchers to determine the content validity of their instruments. Researchers should evaluate the advantages and disadvantages of each approach and adopt the most appropriate method depending on the nature and purpose of their research.

Content Validating the B2EPUS Instrument

Having realized that the quantitative approach to content validity assessment in IS research is infrequently used, we considered the possibility of utilising this approach when developing the B2EPUS instrument. We initially evaluated the pros and cons of qualitative and quantitative approaches. We concluded that for our study, the quantitative approach was more appropriate than the qualitative approach for several reasons. Firstly, we were dealing with international expert panels and it was impossible to meet them face-to-face. We were not equipped with high technology communication facilities. Furthermore, time-zone differences may affect our ability to perform the structured discussions among experts. Secondly, at the time we commenced our study, there were not many experts in B2E portals; hence, it was crucial for us to attract as many identified prospective experts as possible. The qualitative approach generally requires a higher commitment from the experts as they are expected to be involved in the item evaluation process until a consensus is reached. On the other hand, experts are required to participate only once when the quantitative approach is utilized. We cannot afford to take the risk of receiving a low response rate from the invited prospective experts; hence, the quantitative approach is more practical. Thirdly, all the experts on our panel were strongly engaged in their own work commitments so that they needed flexibility in communicating with the researchers (i.e. completing the required tasks and sending the completed document). Based on these reasons, we concluded that the quantitative approach to content validation was more feasible in terms of practicality than the qualitative approach.

When performing the content validation process of the B2EPUS instrument, a number of unexpected circumstances occurred. We treated these as valuable lessons that other researchers can learn from our experience. Firstly, it is crucial to select the right people as members of the expert panel. They should have extensive knowledge in the area to be investigated as they will assist researchers in determining the most appropriate items to be included in the instruments. For an established discipline, it may be manageable to identify the experts. However, for more recently emerging research areas, it may take additional effort and greater time to identify the experts. It is quite common to find that not many people have expertise in the particular area of interest. In our study, B2E portals are relatively new technologies, and therefore there are not many experts in the field. In addition, we also have limited knowledge about their expertise in B2E portal technology beyond their qualifications and research interests or portal experience. To overcome this issue, we identified academics, researchers, and PhD candidates who have great interest in general portal technology. We invited them to participate in the study and also presented a brief explanation of the scope of B2E portals to them. The provision of relevant background information about the research topic will definitely assist the experts in evaluating the instrument items.

Secondly, we invited more than 30 prospective experts, including academics, portal researchers, portal practitioners, and PhD candidates. In the end, there were only six and three responses used in the first and second round respectively. Although the response rate appeared to be low, this was not a major issue as it is adequate to have a minimum of three in the panel (Lynn 1986). What we experienced is quite common. Therefore, researchers should realise that something similar to this may well occur and, in anticipation of a low response rate, as many experts as possible should be invited to participate.

Thirdly, although researchers can execute the item evaluation process once when

assessing expert feedback quantitatively, additional rounds may be required when the instrument is significantly modified. For instance, sometimes experts may suggest adding new dimensions or new items to the instrument. If they provide a reasonable explanation and justification for this, researchers may need to accommodate their suggestions. At other times, as a result of deleting a number of items, researchers may need to add new items as the remaining items do not adequately capture the construct to be measured. Consequently, these changes will greatly modify the conceptual model of the construct to be measured. Therefore, in such circumstances, an additional round of the judgement process is necessary. Similarly, in our study, new items were added and the conceptual framework was revised, thereby making a second round process inevitable in order to reach a robust conclusion. Based on this experience, researchers should always anticipate the need to conduct additional rounds for the content validation procedure. They need to ensure that they have a sufficient number of experts by: 1) asking for permission to contact those who have been involved in the first round should a second round be required; or 2) having a separate list of experts who can be invited for the second round of the content validation process.

Finally, designing and prescribing the tasks to be performed by the content experts is another major issue to consider. Since there are varieties of procedures that can be used to assess expert feedback, the advantages and disadvantages of each of them should be evaluated and the most appropriate procedure must be selected. Researchers often adopt the procedures that have been utilized by previous researchers with the hope that the familiarity of the given tasks may enhance the integrity and accuracy of the completed tasks. In actual fact, the outcomes may not be according to the anticipated responses. In our study, in the first round, two tasks were given. Firstly, the experts were asked to match each item with its nominated dimension. Secondly, they were asked to rank the relevancy of the item to the assigned dimension. We followed previous research conducted by Head, Maas, and Johnson (2003) and Idyall, Hamrin, Sjostrom, and Unison (2001) in assessing these tasks by calculating the weighted-mean score. We found that some of the experts in the first round did not complete the tasks as specified in the instruction document. We suspected that the tasks might have been too tedious for them or they might not have been familiar with the content validation procedures. Hence, for the second round, we decided to simplify the task, whereby they were asked only to rate the relevancy of each item to the assigned dimension. This task was commonly assessed by calculating CVI as can be seen from the work of Evans (2005), Barbara (1997), Beitz and Rijswijk (1999), and Zuzelo, Inverso, and Linkewich (2001). As different tasks were given in the first and second rounds, the feedback assessment methods needed to be adjusted, resulting in the employment of different methods for each of the two rounds. This shows that no matter how well researchers have prepared the tasks, these may not be well executed by the content experts. The least we can do to overcome this concern is to ensure that the instructions stated in the document are clear and concise. We might also need to separate the tasks into a number of smaller sub-sections to minimize the perception of lengthy and tedium. Finally, we should set the tasks in a simple and well-structured format so that experts can easily comprehend what is required of them.

Nevertheless, this study contains limitations that could be addressed in future studies. Firstly, we included instrument development articles from only five leading IS journals. Future research may replicate and expand upon this study by increasing the size of the sample journals. Secondly, this study mainly focuses on exploring the methods, either the qualitative or quantitative approaches, that have been employed by IS researchers when developing their instruments. Future research may look at the relationship between the methods used for content validity and the research findings. It may also explore whether there are differences in research findings between those studies reporting, and those not reporting, content validity. Thirdly, an inverse relationship between the judgement stage and pre-tests was found and we have explored a number of possibilities explaining this trend. Future research may attempt to

empirically find explanations of this phenomenon.

CONCLUSION

The development of a reliable and valid instrument is a lengthy process. At its most basic level, an instrument must be content valid, in that generated items are representative of the construct to be measured. A review of past literature on instrument development studies revealed that the proportion of studies reporting content validity has indeed increased steadily but is still only 60%. Furthermore, it was also found that assessing content validity qualitatively has been the preferred method utilized by IS researchers when developing instruments. The employment of the weighted-mean score and content validity index for assessing content validity of the B2EPUS instrument has been described in this paper. It is hoped that this study will motivate IS researchers to conduct and report more comprehensive content validation procedures as well as consider the quantitative content validation approach.

Acknowledgement. The authors would like to thank the senior editor and three anonymous referees for the constructive feedback that have greatly increased the lucidity of the paper. This research was supported by the Monash University Postgraduate Publication Award.

APPENDIX 1: A SUMMARY OF CONTENT VALIDATION TECHNIQUE IN IS RESEARCH

			Development Stage		Judgment Stage	Pre-Test	Remark
No.	**Year**	**Author**	**Domain Identification**	**Item Generation**			
1	1989	Joshi	✓ (L)	✓ (L)	✓	✓	Generated items were evaluated by a panel of experts a number of times. Approach used to achieve consensus was not described. Preliminary instrument was pretested prior to pilot /full study.
2	1991	Moore and Benbasat	✓ (L)	✓ (S, NW)	✓	✓	Generated items were refined using Q-sort technique. Pretest is used to comment on its length, wording, and instructions and to test initial reliability of the scale.
3	1991	Goodhue and Straub	✓ (L)	✓ (L)	✓	NR	The initial instrument was reviewed by a group of experts through extensive field interviews. The instrument was also reviewed by an independent group of people through interview and questionnaire responses.
4	1992	Saunders and Jones	✓ (L)	✓ (L)	✓	NR	Dimensions of the targeted construct and items were selected through the use of the Delphi approach.
5	1992	Webster and Martocchio	✓ (L)	✓ (S,NW)	NR	✓	Evidence of content validity was provided by pretesting the scale to check the appropriateness of the items. Not specifically explained how this was achieved.
6	1993	Igbaria and Baroudi	✓ (L)	✓ (S)	NR	NR	Expert judgment or pre-test was not reported.

7	1993	Barki, Rivard, and Talbot	✓ (L)	✓ (S,NW)	NR	✓	A pretest of preliminary instrument was conducted with 10 project managers and 8 users, leading to minor modifications.
8	1993	Ferratt, Short, and Agarwal	✓ (L)	✓ (S,NW)	NR	✓	Content validity reached through extensive review of literature to develop the conceptual domain of the construct. The instrument was pre-tested 3 times.
9	1994	Sethi and King	✓ (L)	✓ (L)	NR	NR	Content validation was claimed based on thorough literature on developing conceptual model. Expert judgment was considered but not used because it raised a number of difficult issues (eg. How to define and choose experts).
10	1994	Barki and Hartwick	✓ (L)	✓ (S)	NR	NR	A thorough review of literature enabled a representative and comprehensive sampling of the construct, providing evidence of content validity.
11	1995	Lewis, Snyder, and Rainer	✓ (L)	✓ (L)	✓	✓	The instrument was developed based on Churchill (1979) paradigm. The initial instrument was pretested and pilot tested. Then, a panel of experts was invited to review the refined instrument. Quantitative content validity method using content validity ratio (CVR) by Lawshe (1975) was used.
12	1995	Abdul-Gader and Kozar	✓ (L)	✓ (L)	NR	✓	The instrument was pre-tested through interview process.
13	1996	Jones and Harrison	✓ (L)	✓ (S)	NR	NR	Expert judgement or pre-test was not reported.
14	1996	Saarinen	✓ (L)	✓ (L)	✓	NR	Control group feedback was used to assess whether the construct and generated items were valid and representative.
15	1996	Palvia	✓ (L)	✓ (L,S)	NR	✓	Generated items were pretested prior to pilot/final study.
16	1997	Chin, Gopal, and Salisbury	✓ (L)	✓ (NW)	✓	NR	Generated items were reviewed by experts but did not specify approach to gain consensus.
17	1997	Davison	✓ (L)	✓ (L,S, NW)	NR	✓	A series of pre-test were conducted to assess the representativeness of items but did not specify approach to gain consensus.
18	1997	Palvia	✓ (L)	✓ (L)	NR	✓	The initial instrument was pre-tested in two rounds. Feedback from these pretests was used to refine the instrument.
19	1998	Agarwal and Prasad	✓ (L)	✓ (S,NW)	NR	NR	Generated items were directly used for data collection.
20	1998	Govindarajulu and Reithel	✓ (L)	✓ (S)	✓	NR	Generated items were presented to researchers who were asked to scrutinise the list and delete items.
21	1998	Doll and Torkzadeh	✓ (L)	✓ (L)	NR	NR	The instrument was developed based on Churchill (1979) paradigm. However, the generated items were not subjected to expert judgment or pretest.
22	1998	Segars & Grover	✓ (L, E, I)	✓ (NS)	✓	✓	Proposed dimensions of the construct domain were reviewed by experts for several rounds. However, methods to generate items were not specified. Q-sort was utilized to ensure the developed items were adequate in capturing the construct domain.
23	1999	Raghunathan, Raghunatan, and Tu	✓ (L)	✓ (S,NW)	✓	NR	Generated items were reviewed by IS researchers to assess for their appropriateness and relevance. Then, IS executives were asked to complete the survey and comment on the clarity and appropriateness of the items.

24	2000	Byrd and Turner	✓ (L)	✓ (L)	NR	✓	The instrument was developed based on Churchill (1979) paradigm. The initial instrument was pretested against IT managers who were asked to comment on the completeness, understandability, terminology, ambiguity of the items.
25	2000	Agarwal and Karahanna	✓ (L)	✓ (S)	NR	NR	Expert judgment or pre-test was not reported.
26	2000	Seyal, Rahim, and Rahman	✓ (L)	✓ (L)	NR	✓	Generated items were pretested against academics who were asked to comment on the format and appropriateness of the items. They were also asked to add items which they believe should be included in the instrument.
27	2000	Chen, Soliman, Mao, and Frolick	✓ (L)	✓ (L)	✓	NR	Generated items were reviewed by a number of academics and practitioners who were asked to rate the relevance of the items in terms of end user satisfaction with data warehouses.
28	2001	Lurey and Raisinghani	✓ (L)	✓ (L)	NR	✓	The initial instrument was pretested prior to final study.
29	2001	D'Ambra and Rice	✓ (L)	✓ (FG)	NR	✓	The initial instrument was pretested against master students who were asked to comment on the items, drop duplicate items, modify double barreled and ambiguous items.
30	2001	Mak and Sockel	✓ (L)	✓ (L,S)	NR	NR	Content validity was established through a comprehensive study of relevant literature and existing instruments.
31	2002	McKnight, Choudhury, and Kacmar	✓ (L)	✓ (S,NW)	NR	NR	Generated items were directly used for data collection.
32	2002	Zhu and Kraemer	✓ (L)	✓ (NW)	✓	✓	Generated items were reviewed by a panel of academic and industry experts but did not specify method to gain consensus. Then, the initial instrument was pretested.
33	2002	Salisbury, Chin, Gopal, and Newsted	✓ (L)	✓ (NW)	✓	NR	Generated items were reviewed by a panel of experts but did not specify approach to gain consensus.
34	2002	McKinney, Yoon, and Zahedi	✓ (L)	✓ (S)	✓	NR	The instrument was developed based on Churchill (1979) paradigm. Generated items were reviewed by a group of experts but did not specify approach to gain consensus.
35	2002	Agarwal and Venkatesh	✓ (L)	✓ (NW)	✓	NR	Generated items were reviewed by different group of experts for 4 rounds. However, approach to gain consensus was not specified.
36	2002	Gatignon, Tushman, Smith, and Anderson	✓ (L)	✓ (NW)	✓	NR	The instrument was developed based on Churchill (1979) paradigm. Generated items were reviewed by a group of experts. Item was eliminated when a majority of experts responded that the item did not reflect the construct.
37	2002	Bhattacherjee	✓ (L)	✓ (S,NW)	✓	✓	Q-sort technique was used to refine the generated items. A measure of inter-rater reliability (Cohen's Kappa) was used to refine the instrument.
38	2002	Templeton, Lewis, and Snyder	✓ (L)	✓ (S,NW)	✓	✓	The initial instrument was pretested to refine the instrument. Then, a panel of experts was invited to review the refined instrument. Quantitative content validity method using content validity ratio (CVR) by Lawshe (1975) was used.

39	2002	Lee,Strong, Kahn, and Wang	✓ (L)	✓ (NS)	✓	NR	Generated items were reviewed by experts to ensure that all items were the representative of the construct. These items were then reviewed by other group of users to ensure that the potential respondents understand the meaning of the items. This review and editing process was repeated until agreement was reached. Approach to gain consensus was not specified.
40	2002	Aladwani and Palvia	✓ (L)	✓ (L)	✓	NR	The instrument was developed based on Churchill (1979) paradigm. Delphi technique was used to refine the instrument. Consensus was achieved after 3 evaluation rounds.
41	2003	Balasubramanian, Konana, and Menon	✓ (L)	✓ (L,I)	NR	✓	Generated items were pretested twice. The instrument was refined based on the feedback from each pretest. Method to retain/delete items was not specified.
42	2003	Peace, Galletta, and Thong	✓ (L)	✓ (S,NW)	✓	✓	Generated items were reviewed by a panel of experts. The refined items were then pretested twice. Method to gain consensus was not specified.
43	2003	Teo, Wei, and Benbasat	✓ (L)	✓ (L)	✓	NR	Generated items were refined through 4 round Q-sort technique described by Moore and Benbasat (1991).
44	2003	Osmundson, Michael, Machniak, and Grossman	✓ (L)	✓ (NS)	NR	NR	The article mentioned the use of interview and focus group meetings. However, it did not specify whether these two techniques were used for generating or refining items.
45	2003	Wang	✓ (L)	✓ (L,S)	✓	NR	A panel of experts was asked to review the generated items and recommend adding 3 items. Method to retain/delete items was not specified.

46	2003	Torkzadeh and Lee	✓ (L)	✓ (L,E)	NR	NR	Practitioners who were involved in item generation felt that the list of items was complete.
47	2004	Bassellier and Benbasat	✓ (L)	✓ (S,NW)	✓	NR	Generated items were refined through Q-sort technique described by Moore and Benbasat (1991).
48	2004	Muylle, Moenaert, and Despontin	✓ (L)	✓ (I, L)	NR	NR	Expert judgment or pre-test was not reported.
49	2004	Van der Heijden and Verhagen	✓ (L)	✓ (L,FG)	NR	NR	The instrument was developed based on Churchill (1979) paradigm. Expert judgement or pre-test was not reported.
50	2004	Chiou	✓ (L)	✓ (S)	NR	✓	Generated items were pretested prior to pilot/final study.
51	2005	Xia and Lee	✓ (L)	✓ (L, I, FG)	✓	✓	Generated items were refined through Q-sort technique. Then, the pre-test was conducted in the form of individual interviews. The purpose of the pre-test was to further refine the instrument.
52	2005	Chang and King	✓ (L)	✓ (S,L,E)	✓	✓	Generated items were refined through 2 round Q-sort technique described by Moore and Benbasat (1991).
53	2005	Kim, Umanath, and Kim	✓ (L)	✓ (S)	✓	✓	Generated items were reviewed through multiple structured interviews but method to gain consensus on items did not specify. The refined items were then pretested.
54	2005	Ko, Kirsch, and King	✓ (L)	✓ (S,NW)	✓	✓	Generated items were reviewed by a group of experts and pretested by a group of consultants and clients. However, method to retain/delete item was not specified.

55	2005	Kankanhalli, Tan, and Wei	✓ (L)	✓ (S,L,NW)	✓	NR	Q-sort technique described by Moore and Benbasat (1991) was employed.
56	2005	Lai and Li	✓ (L)	✓ (S,NW)	✓	✓	The generated items were reviewed by a group of experts but method of retaining/deleting items was not specified. The refined items were then pretested.
57	2005	Kim and Umanath	✓ (L)	✓ (NS)	NR	✓	The initial instrument was pretested through structured interviews but method to retain/delete item was not specified.
58	2005	Huang	✓ (L)	✓ (I,L)	NR	NR	Expert judgement or pre-test was not reported.
59	2005	Molla and Licker	✓ (L)	✓ (NS)	✓	NR	A panel of experts was asked to judge the degree of relevance of each item. They were asked to suggest additional items that were not covered in the instrument. In checking how evaluators agreed in their assessment of variable, inter-observer reliability was evaluated using correlation coefficients.
60	2005	Lee, Lee, and Kang	✓ (L)	✓ (E)	✓	✓	The instrument was developed based on Churchill (1979) paradigm. Generated items were reviewed by a group of experts to determine the quality of the items. Then, pretest was conducted to evaluate the instrument based on its clarity and understandability.
61	2005	Yang, Cai, Zhou, and Zhou	✓ (L)	✓ (FG)	✓	NR	Generated items were reviewed by the managers and users. Some items were refined based on their suggestions. Method to retain/delete item was not specified.
62	2005	Shi, Kunnathur, Ragu-Nathan	✓ (L)	✓ (L)	✓	✓	Generated items were reviewed by the IS and quality directors through field interview. The interview result validated the construct and enhanced the design of measurement items.

* L= derived from literature; S= adapted/adopted from existing instruments; FG= focus group; I = interview; NS = not specified; NW = develop new item; E = derived from experts' or practitioners' opinion; NR = not reported.

APPENDIX 2: THE B2EPUS SCALE REFINEMENT PROCESS

Original Items	Items in CV First Round	Items in CV Second Round	Final Items
The portal is accessible from my office.	✓	x	x
The portal is accessible from my home through internet connection.	✓	✓	✓
The portal is accessible from mobile devices such as mobile phone and PDAs.	✓	x	x
The portal is accessible 24 hours a day, 7 days a week from anywhere.	✓	✓	Revised: The portal is accessible 24 hours a day, 7 days a week.
Gaining access to portal is easy.	✓	✓	✓
Learning to use the portal is easy for me.	✓	✓	✓
The portal is user friendly.	✓	✓	Revised: The staff portal is user friendly with abundant help functions and useful button and links.
The portal is easy to navigate.	✓	✓	Revised: The staff portal is easy to navigate, both forward and backward.
When I am navigating the portal, I feel that I am in control of what I can do.	✓	✓	✓
Training on how to use the portal is not necessary as the portal is easy to use.	✓	✓	Revised: No training on how to use the staff portal is necessary as the portal use is self-explanatory.
When I access the portal, there is very little waiting time between my actions and the web site's response.	x	x	x

The speed of the responses to your request of information is good.	x	x	x
The portal assists me in performing my tasks better.	✓	✓	Revised: The portal assists me in performing my task with a better quality.
Using the portal enables me to accomplish tasks more quickly	✓	✓	✓
The portal has improved my productivity.	✓	✓	✓
The self-service function provided in the portal has successfully streamline work processes.	✓	✓	✓
The portal lets me do more work than was previously possible.	✓	✓	Revised: The portal lets me do more work electronically than was previously possible.
I am concerned that personal information I submit through the portal could be misused.	✓	✓	Revised: I feel confident in submitting personal information through the staff portal because it will be properly used by authorised people
I am concerned about submitting information through the portal because of what others might do with it.	✓	✓	Revised: I am certain that personal information I submit through the staff portal will be properly used by authorised people.
I am concerned about submitting information through the portal because it could be used in a way I did not foresee.	x	x	x
The use of a single sign on procedure (i.e. one password to access all information) increases the security of the portal.	✓	x	x
I feel the portal is secure.	✓	✓	✓
I am concerned that someone else can access my personal information that is available through the portal.	x	x	x
I am concerned that my personal information could be made available to unknown individuals without my knowledge.	x	x	x
The portal enables me to share information with other colleagues.	✓	✓	Revised: The portal enables me to share or exchange project/task information with my team member colleagues.
The portal enables me to share information with the whole organisation.	✓	✓	Revised: The staff portal enables me to share general information via email or on website with other colleagues in the whole organisation.
The portal facilitates me in collaborating with other colleague.	✓	✓	Revised: The staff portal facilitates my collaboration work with all colleagues.
The portal makes it easy for me to discuss issues with other colleagues.	✓	✓	Revised: The staff portal enables me to discuss work or project issues with my immediate work colleagues.
The information provided by the portal is correct.	✓	x	x
I am satisfied with the precision of information presented on the portal.	✓	✓	Revised: I am satisfied with the accuracy of information presented on the portal.
The portal records and processes data without making any errors.	x	x	x

The information provided by the portal is updated regularly.	x	x	x
The portal keeps me informed with the latest information.	x	x	x
The information provided in the portal can be trusted.	✓	✓	✓
The portal always produces dependable information regardless of when and where I access it.	x	x	x
I only have access to information that is related to my roles within the organisation.	x	x	x
I am not overloaded with information as only relevant information is provided.	x	x	x
Information presented on the portal meets my needs.	✓	✓	✓
The information provided by the portal is clear.	x	x	x
The information provided by the portal is understandable.	x	x	x
The text on the portal screen is easy to read.	x	x	x
The portal is aesthetically designed.	✓	✓	✓
The portal screen layout design makes it easy for me to find the content I need.	x	x	x
The design of the portal is attractive.	✓	✓	✓
The portal uses proper font size.	✓	x	x

The portal uses proper colours.	✓	x	x
	Add: It does not take much time to go from one link in the portal to another.	x	x
	Add: While using the portal, there is very little waiting time between my action and portal's response.	x	x
	Add: I can rely on the information provided by the portal.	✓	✓
	Add: Information presented on the portal is dependable.	✓	✓
	Add: The information provided by the portal is always updated.	x	x

REFERENCES

Abdul-Gader, A.H. and K.A. Kozar, "The Impact of Computer Alienation of Information Technology Investment Decisions: An Exploratory Cross-National Analysis," *MIS Quarterly*, 1995, 19:4, pp. 535-559.

Agarwal, R. and E. Karahanna, "Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage," *MIS Quarterly*, 2000, 24:4, pp. 665-694.

Agarwal, R. and J. Prasad, "Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology," *Information Systems Research*, 1998, 9:2, pp. 204-215.

Agarwal, R. and V. Venkatesh, "Assessing a Firm's Web Presence: A Heuristic Evaluation Procedure for the Measurement of Usability," *Information Systems Research*, 2002, 13: 2, pp. 168-186.

Aladwani, A.M. and P.C. Palvia, "Developing and validating an instrument for measuring user-perceived web quality," *Information and Management*, 2002, 39, pp. 467-476.

Alreck, P.L. and R.B. Settle, *The Survey Research Handbook* (2nd Edition), Chicago: Irwin Professional Publishing, 1995.

Bagozzi, R.P, "Evaluating structural equation models with unobservable variables and measurement error: a comment," *Journal of Marketing Research*, 1980, 13, pp. 375-381.

Bailey, J. E. and S. W. Pearson, "Development of a tool for measurement and analyzing computer user satisfaction," *Management Science*, 1983, 29:5, pp. 530-545.

Balasubramanian, S., P. Konana, and N.M. Menon, "Customer Satisfaction in Virtual Environments: A Study of Online Investing," *Management Science*, 2003, 49:7, pp. 871-889.

Barbara, C.A, "Development of the conscious sedation scale: establishing content validity and reliability", *Gastroenterology Nursing*, 1997, 20:1, pp. 2-7.

Barki, H. and J. Hartwick, "Measuring User Participation, User Involvement, and User Attitude," *MIS Quarterly*, 1994, 18:1, pp. 59-82.

Barki, H., S. Rivard, and J. Talbot, "Toward an Assessment of Software Development Risk," *Journal of Management Information Systems*, 1993, 10:2, pp. 203-225.

Bassellier, G and I. Benbasat, "Business Competence of Information Technology Professionals: Conceptual Development and Influence on IT-Business Partnership," *MIS Quarterly*, 2004, 28:4, pp. 673-694.

Beitz, J.M. and L. Rijswijk, 'Wound care algorithms: a content validation study," *Journal of Wound, Ostomy, and Continence Nursing*, 1999, 26:5, pp. 238-249.

Bhattacherjee, A, "Individual Trust in Online Firms: Scale Development and Initial Test," *Journal of Management Information Systems*, 2002, 19:1, pp. 211-241.

Boudreau, M.C., D. Gefen, and D.W. Straub, "Validation in information systems research: a state-of-the-art assessment," *MIS Quarterly*, 2001, 25:1, pp. 1-16.

Byrd, T.A. and D.E. Turner, "Measuring the Flexibility of Information Technology Infrastructure: Exploratory Analysis of a Construct," *Journal of Management Information Systems,* 2000, 17:1, pp. 167-208.

Carmines, E.G. and R.A. Zeller, *Reliability and validity*, Sage Publications: Beverly Hills, 1979

Chang, J.C-J. and W.R. King, "Measuring the Performance of Information Systems: A Functional Scorecard," *Journal of Management Information Systems*, 2005, 22:1, pp. 85-115.

Chen, L., K. S. Soliman, E. Mao, and M. N. Frolick, "Measuring user satisfaction with data warehouses: an exploratory study," *Information and Management***,** 2000, 37:3, pp. 103-110.

Chin, W.W., A. Gopal, and W.D. Salisbury, "Advancing the Theory of Adaptive Structuration: The Development of a Scale to Measure Faithfulness of Appropriation," *Information Systems Research*, 1997, 8:4, pp. 342-366.

Chiou, J-S. "The antecedents of consumers' loyalty toward Internet Service Providers", *Information and Management*, 2004, 41, pp. 685-695.

Cook, J.D., S.J. Hepworth, T.D. Wall, and P.B. Warr, *The experience of work*, San Diego: Academic Press, 1981.

Cronbach, L.J., "Test Validation in Educational Measurement," (2nd Edition) in *American Council on Education*, Thorndike, R.L. (ed.),Washington D.C., 1971, pp. 443-507.

D'Ambra, J. and R.E. Rice, "Emerging factors in user evaluation of the World Wide Web," *Information and Management***,** 2001, 38, pp. 373-384.

Davis, L., "Instrument review: getting the most from your panel of experts," *Applied nursing research*, 1992, 5, pp. 104-107.

Davison, R., "An instrument for measuring meeting success," *Information and Management*, 1997, 32, pp. 163-176.

Davydov, M. M., "*Corporate Portals and e-Business Integration*," New York: McGraw-Hill, 2001

DeVellis, R.F., "Scale development: Theory and Application," (2nd Ed.), United Kingdom: Sage Publications, 2003.

Dillman, D., "*Mail and telephone surveys*", New York: Wiley, 1978.

Doll, W.J. and G. Torkzadeh., "Developing a multidimensional measure of system-use in an organizational context," *Information and Management*, 1998, 33, pp. 171-185.

Ebel, R. L.,'Evaluating content validity'. In *Educational and Psychological Measurement: Contributions to theory and practice*, Payne, D. A. and McMorris, R. F. (Eds), Waltham, Blaisdel, 1967, pp. 85-94.

Evans, B.C., "Content validation of instruments: are the perspectives of Anglo Reviewers different from those of Hispanic/Latino and American Indian Revieweres?," *Journal of Nursing Education*, 2005, 44:5, pp. 216-224.

Fehring, R. J., "Methods to validate nursing diagnoses, "*Heart and Lung*, 1987, 16:6, pp. 625-629.

Ferratt, T.W., L.E. Short, and R. Agarwal., "Measuring the Information Systems Supervisor's Work-Unit Environment and Demonstrated Skill at Supervising," *Journal of Management Information Systems*, 1993, 9:4, pp. 121-144.

Gatignon, H., M.L Tushman., W.Smith, and P. Anderson, "A Structural Approach to Assessing Innovation: Construct Development of Innovation Locus, Type, and Characteristics," *Management Science*, 2002, 48:9, pp. 1103-1122.

Goodhue, D.L. and D.W. Straub, "Security concerns on system users: a study of perceptions of the adequacy of security," *Information and Management*, 1991, 20, pp. 13-27.

Goodman, C.M. , "The Delphi Technique: A Critic," *Journal of Advanced Nursing*, 1987, 12, pp. 729-734.

Govindarajulu, C. and B.J. Reithel, "Beyond the Information Center: An Instrument to Measure End User Computing Support for Multiple Sources," *Information and Management*, 1998, 33, pp. 241-250.

Grant, J.S. and L.L. Davis, "Focus on Quantitative Methods: Selection and use of content experts for instrument development," *Research in Nursing and Health*, 1997, 20, pp. 269-274.

Grant, J. and M. Kinney, "Using the Delphi technique to examine the content validity of nursing diagnoses," *Nursing Diagnosis*, 1992, 3, pp. 12-22.

Grant, J., M. Kinney, and C. Guzzetta, "A methodology for validating nursing diagnoses," *Advanced Nursing Science*, 1990, 12:3, pp. 65-74.

Haynes, S.N., D.C.S., Richard, and E.S. Kubany, "Content validity in psychological assessment: a functional approach to concepts and methods," *Psychological Assessment*, 1995, 7:3, pp. 238-247.

Head, B.J., M. Maas, and M. Johnson, "Validity and community-health-nursing sensitivity of six outcomes for community health nursing with older clients," *Public Health Nursing*, 2003, 20:5, pp. 385-398.

Hinkin, T.R., "A review of scale development practices in the study of organisations," *Journal of Management*, 1995, 21: 5, pp. 967-988.

Hinkin, T.R. and C.A. Schriesheim, "Development and application of new scales to measure the French and Raven (1959) bases of social power," *Journal of Applied Psychology*, 1989, 74:4, pp. 561-567.

Huang, M-H., "Web performance scale," *Information and Management*, 2005, 42, pp. 841-852.

Idvall, E., E. Hamrin, B. Sjostrom, and M. Unosson, "Quality indicators in postoperative pain management: a validation study," *Scandinavia Journal Caring Science*, 2001, 15:4, pp. 331-338.

Igbaria, M. and J.J. Baroudi, "A Short-Form Measure of Career Orientations: A Psychometric Evaluation," *Journal of Management Information Systems*, 1993, 10:2, pp. 131-154.

Jones, M.C. and A.W. Harrison, "IS project team performance: an empirical assessment," *Information and Management*, 1996, 31, pp. 57-65.

Joshi, K., "The Measurement of Fairness or Equity Perceptions of Management Information Systems Users," *MIS Quarterly*, 1989, pp. 343-358.

Kankanhalli, A., B.C.Y Tan, and K.K. Wei, "Contributing Knowledge to Electronic Knowledge Repositories: An Empirical Investigation," *MIS Quarterly*, 2005, 29:1, pp. 113-143.

Kim, K.K. and N.S. Umanath, "Information transfer in B2B procurement: an empirical analysis and measurement," *Information and Management*, 2005, 42, pp. 813-828.

Kim, K.K., N.S. Umanath, and B.H. Kim., "An Assessment of Electronic Information Transfer in B2B Supply-Channel Relationships," *Journal of Management Information Systems*, 2005, 22:3, pp. 293-320.

Ko, D-G., L.J. Kirsch, and W.R. King, "Antecedents of Knowledge Transfer from Consultants to Clients in Enterprise System Implementations," *MIS Quarterly*, 2005, 29:1, pp. 59-85.

Lai, V.S. and H.Li, "Technology acceptance model for internet banking: an invariance analysis," *Information and Management*, 2005, 42, pp. 373-386.

Lawshe, C.H., "A quantitative approach to content validity," *Personnel Psychology*, 1975, 28, pp. 563-575.

Lee, K.C., S. Lee, and I.W. Kang, "KMPI: measuring knowledge management performance," *Information and Management*, 2005, 42, pp. 469-482.

Lee, Y.W., D.M Strong, B.K Kahn., and R.Y. Wang, "AIMQ: a methodology for information quality assessment," *Information and Management*, 2002, 40, pp. 133-146.

Lewis, B.R., C.A. Snyder, and R.K. Rainer, "An Empirical Assessment of the Information Resource Management Construct," *Journal of Management Information Systems*, 1995, 12:1, pp. 199-223.

Lindeman, C.A., "Delphi survey of priorities in clinical nursing research," *Nursing Research*, 1975, 24:6, pp. 434-41.

Lurey, J.S. and M.S. Raisinghani, "An empirical study of best practices in virtual teams," *Information and Management*, 2001, 38, pp.523-544.

Lynn, M., "Determination and quantification of content validity," *Nursing Research*, 1986, 35:6, pp. 382-385.

Mak, B.L. and H.Sockel, "A confirmatory factor analysis of IS employee motivation and retention," *Information and Management*, 2001, 38, pp. 265-276.

McKinney, V., K.Yoon, and F. Zahedi, "The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach," *Information Systems Research*, 2002, 13:3, pp. 296-315.

McKnight, D.H., V.Choudhury, and C. Kacmar., "Developing and Validating Trust Measures for e-Commerce," *Information Systems Research*, 2002, 13:2, pp. 334-359.

Molla and Licker, 2005

Molla, A. and P.S. Licker, "eCommerce adoption in developing countries: a model and instrument," *Information and Management*, 2005, 42, pp. 877-899.

Moore, G.C. and I. Benbasat, "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation," *Information Systems Research*, 1991, 2:3, pp. 192-222.

Muylle, S., R.Moenaert, and M. Despontin,, "The conceptualization and empirical validation of web site user satisfaction," *Information and Management*, 2004, 41, pp. 543-560.

Nunnally, J. C., *Psychometric Theory*. (2nd ed.). New York: McGraw-Hill, 1978

Nunnally, J. C. and Bernstein, I.H. , *Psychometric Theory*. (3rd ed.). New York: McGraw-Hill, 1994

Osmundson, J.S., J.B. Michael, M.J Machniak, and M.A. Grossman, "Quality management metrics for software development," *Information and Management*, 2003, 40, pp. 799-812.

Palvia, P. C., "A model and instrument for measuring small business user satisfaction with information technology," *Information and Management*, 1996, 31, pp. 151-163.

Palvia, P.C., "Developing a model of the global and strategic impact of information technology," *Information and Management*, 1997, 32, pp. 229-244.

Peace, A.G., D.F. Galletta, and J.Y.L. Thong, "Software Piracy in the Workplace: A Model and Empirical Test," *Journal of Management Information Systems*, 2003, 20:1, pp. 153-177.

Raghunathan, B., T.S. Raghunatan, and Q.Tu., "Dimensionality of the Strategic Grid Framework: The Construct and its Measurement," *Information Systems Research*, 1999, 10:4, pp. 343-355.

Rainer, R.K. and A.W. Harrison, "Toward development of the end user computing construct in a university setting," *Decision Sciences*, 1996, 24:6, pp. 1187-1202.

Rogers, T. B. , *The Psychological Testing Enterprise*, Brooks/Cole Publishing Company, Pacific Grove, CA, 1995.

Rovinelli, R.J. and R.K. Hambleton., "On the use of content specialists in the assessment of criterion-referenced test item validity," *Dutch journal for educational research*, 1977, 2, pp. 49-60.

Saarinen, T., "An expanded instrument for evaluation information system success," *Information and Management*, 1996, 31, pp. 103-118.

Salisbury, W.D., W.W. Chin, A. Gopal, and P.R. Newsted, "Research Report: Better Theory Through Measurement-Developing a Scale to Capture Consensus on Appropriation," *Information Systems Research*, 2002, 13:1, pp. 91-103.

Saunders, C.S. and W.J. Jones., "Measuring Performance of the Information Systems Function," *Journal of Management Information Systems*, 1992, 8:4, pp. 63-82.

Schwab, D. P., 'Construct Validity in Organization Behaviour' in *Research in Organizational Behaviour*, Staw, B. M. and L. L. Cummings (Eds), Greenwich: JAI Press, 1980, pp. 3-43.

Segars, A.H. and Grover, V., "Strategic Information Systems Planning Success: An Investigation of the Construct and Its Measurement", *MIS Quarterly*, 1998, 22:2, pp. 139-163.

Sethi, V. and W.R. King., "Development of Measures to Assess the Extent to Which an Information Technology Application Provides Competitive Advantage," *Management Science*, 1994, 40:12, pp. 1601-1627.

Seyal, A.H., M.M. Rahim, and M.N.A. Rahman, "Computer attitudes of non-computing academics: a study of technical colleges in Brunei Darussalam," *Information and Management*, 2000, 37: 169-180.

Shi, Z., A.S. Kunnathur, and T.S. Ragu-Nathan, "IS outsourcing management competence dimensions: instrument development and relationship exploration," *Information and Management*, 2005, 42, pp. 901-919.

Slocumb, E.M. and F.L. Cole, "A practical approach to content validation," *Applied Nursing Research*, 1991, 4:4, pp. 192-200.

Straub, D.W., "Validating Instruments in MIS Research," *MIS Quarterly*, 1989, pp. 147-169.

Straub, D., M.C. Boudreau, and D. Gefen, "Validation Guidelines of IS Positivist Research," *Communication of the Association for Information Systems*, 2004, 14, pp. 380-426.

Sugianto, L and D.R. Tojib, "Modelling User Satisfaction with an Employee Portal", *International Journal of Business and Information*, 2007, --:--, pp. – (to appear).

Templeton, G.F., B.R. Lewis, and C.A. Snyder, "Development of a Measure for the Organisational Learning Construct," *Journal of Management Information Systems*, 2002, 19:2, pp. 175-218.

Teo, H.H., K.K. Wei, and I. Benbasat, "Predicting Intention to Adopt Interorganisational Linkages: An Institutional Perspective," *MIS Quarterly*, 2003, 27:1, pp. 19-49.

Thorn, D.W. and J.C. Deitz, "Examining content validity through the use of content experts," *The Occupational Therapy Journal of Research*, 1989, 9:6, pp. 334-346.

Tojib, D.R. and L. Sugianto, "Content Validating the B2E Portal User Satisfaction Instrument," in Proceedings of the 5th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2006), 10-12 July, Hawaii.

Tojib, D.R. and L.Sugianto, "The Development and Empirical Validation of B2E Portal User Satisfaction (B2EPUS) Scale", *Journal of End User Computing*, 2007, --:--, pp. – (to appear).

Torkzadeh, G. and J. Lee, "Measures of perceived end-user computing skills," *Information and Management*, 2003, 40, pp. 607-615.

Van der Heijden, H. and T. Verhagen, "Online store image: conceptual foundations and empirical measurement," *Information and Management*, 2004, 41, pp. 609-617.

Waltz, C.W. and R.B. Bausell, *Nursing research: Design, statistics and computer analysis*, Philadelphia: F.A. Davis, 1981.

Waltz, C.F., O.L. Strickland, and E.R. Lenz, *Measurement in nursing research (2nd ed)*. Philadelphia: F.A.Davis, 1991.

Wang, Y-S., "Assessment of learner satisfaction with asynchronous electronic learning systems," *Information and Management*, 2003, 41, pp. 75-86.

Watkins, M.W. and P.A. McDermott., "A computer program for measuring overall and partial congruence among multiple observers on nominal scales," *Educational and Psychological Measurement* , 1979, 39, 235-239

Webster, J. and J.J. Martocchio, "Microcomputer Playfulness: Development of a Measure With Workplace Implications," *MIS Quarterly*, 1992, pp. 201-226.

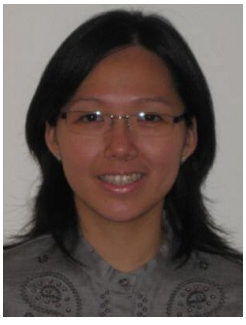
Xia, W and G. Lee, "Complexity of Information Systems Development Projects: Conceptualization and Measurement Development," *Journal of Management Information Systems*, 2005, 22:1, pp. 45-83.

Yang, Z., S. Cai, , Z. Zhou, and N. Zhou, "Development and validation of an instrument to measure user perceived service quality of information presenting web portals," *Information and Management*, 2005, 42, pp. 575-589.

Zhu, K. and K.L. Kraemer, "E-commerce Metrics for Net-Enhanced Organisations: Assessing the Value of e-Commerce to Firm Performance in the Manufacturing Sector," *Information Systems Research*, 2002, 13:3, pp. 275-295.

Zuzelo, P.G., T. Inverso, and K.M. Linkewich, "Content validation of the medication error worksheet," *Clinical Nurse Specialist*, 2001, 15:6, pp. 253-259.

AUTHORS



Dr. Dewi Rooslani Tojib is currently a Research Fellow in the Department of Marketing, Monash University, Australia. She received her Bachelor of Business Systems (Honours) and Doctor of Philosophy degree from Monash University. Her research interests include portal technology, electronic business, mobile commerce, quantitative and qualitative research method, instrument development and validation, and user satisfaction studies.

Dr. Ly Fie Sugianto is currently a Senior Lecturer at the Faculty of Information Technology, Monash University, Australia. She holds Bachelor of Engineering (H1) degree from Curtin University and Doctor of Philosophy from Monash University. Dr Sugianto has published 60+ research papers in optimization technique, Fuzzy mathematics, Decision Support Systems, eCommerce and Information Systems. She has received several grants to conduct research in electricity markets and Information Systems. She has also been nominated as an expert of international standing by the Australian Research Council College of Experts for her work in the electricity market.