:: **Refereed Article A1:**

A content validity approach to creating an end-user computer skill assessment tool

Shirley Gibbs
Informatics & Enabling Technologies, Lincoln University, Christchurch, New Zealand
shirley.gibbs@lincoln.ac.nz

Gary Steel
Tourism, Sport and Society, Lincoln University, Christchurch, New Zealand
gary.steel@lincoln.ac.nz

Alan McKinnon
Informatics & Enabling Technologies, Lincoln University, Christchurch, New Zealand

Abstract

Practical assessment instruments are commonly used in the workplace and educational environments to assess a person's level of digital literacy and end-user computer skill. However, it is often difficult to find statistical evidence of the actual validity of instruments being used. To ensure that the correct factors are being assessed for a particular purpose it is necessary to undertake some type of psychometric testing, and the first step is to study the content relevance of the measure. The purpose of this paper is to report on the rigorous judgment-quantification process using panels of experts in order to establish inter-rater reliability and agreement in the development of end-user instruments developed to measure workplace skills using spreadsheet and word-processing applications.

Keywords

End-user computer skill, spreadsheet skill, word-processing skill, content validity, content relevance, digital literacy, Inter-rater reliability, productivity software

1. Introduction

Typically, people employed in office-based roles are required to use office software, such as word-processing editors and spreadsheet applications (Holtzman & Kraft, 2010). These applications are two of the most commonly used in many workplaces (Holtzman & Kraft, 2010). While the requirement for familiarity with these applications is common, the specific types of use can vary a great deal. In some types of employment, for example, spreadsheets may simply be used to record and tabulate data, while, in others, chart editors and other visualisation tools are the most common feature used (Chambers & Scaffidi, 2010; Lawson, Baker, Powell, & Foster-Johnson, 2009). Due to these differences, it can be difficult to classify a general skill

level, as in one job, a person may be regarded as highly competent and, in another, she or he would appear to be much less qualified. Another difficulty in establishing a person's skill level is that, quite often, end-user computing skills are self-taught. Because such skill acquisition is outside a recognised educational system, this usually means that no formal benchmark has been achieved unless the learner has elected to undertake such an assessment.

2. Literature Review

There are a number of computing learning and testing systems available through educational institutions or accessible via the Internet. Some of these offer industry relevant certification to students. These systems include the product specific SAM (Skill Assessment Manager) and MOS (Microsoft Office Specialist) testing systems. Both of these are tied to Microsoft products. The ECDL and ICDL learning and testing systems are non-product specific.

Vakhitova and Bollinger (2011), say that some employers value computing certification more highly than they value some degree qualifications. This may be because some certifications arm the recipients with very specific skills, whereas the skills gained in a degree, may be regarded by some employers as general. Some employers also believe that employees with certification will require far less workplace training than those without industry type certification (McGill & Dixon, 2004).

Microsoft Certification includes the MOS suite of tests and training modules which Microsoft say will help to validate a person's computing skill (Microsoft learning, 2013). These tests focus on the Microsoft Office suite of applications and involve testing in each. Autrey, Tarver, Myers and Tarver (2004) found that students who had worked their way successfully through the MOS certification were of more value to employers than people without any computing certification. Pascoe (2003) found that gaining MOS certification demonstrates expertise in using the Microsoft Office suite of software. This certification can provide employers with a useful and reliable measure of technical ability and understanding of this particular suite of software applications.

The European Computer Driving license (ECDL), established in 1994 has been expanded across Europe and the rest of the world with the introduction of the ICDL (International Computer Driver's License). The ICDL/ECDL is non-product specific and is able to be delivered in a flexible manner with students learning at their own pace either in a classroom situation or in their own environment (Davis & Cleere, 2003; McLay & Brown, 2006; Calzarossa, Ciancarini, Maresca, Mich & Scarabotto, 2007; Panicos & Sotiris, 2010). To maintain its integrity, the validity of the ECDL/ICDL syllabus is frequently audited as outlined in detail in Davis and Cleere (2003). They note that the validity of the syllabus is overseen by a panel of Subject Matter Experts (SMEs) who undertake a series of core item identification exercises.

Although the aforementioned tests and others are widely available, many employers choose not to use them when assessing computing skills. Instead, many employers rely solely on a person's self-assessment of their ability, which in many cases is over-estimated (Gibbs, Steel & Kuiper, 2011).

As part of a larger study, two instruments were developed to assess workplace skill level in word processing and the use of spreadsheets. These applications were chosen as they are the most widely used end-user applications in many employment situations (Grant, Malloy & Murphy, 2009) The instruments contain a number of practical tasks designed to assess a participant's skill level. Often, tests designed to assess ICT knowledge either involve multiple-choice type questions or consist of self-assessment type instruments, rather than a practical, task-based approach. The problem with multiple-choice assessment is, quite clearly, the do not test actual skill but memory. Self-assessment is prone to both over- and underestimation of one's own capability, with inflation of self-assessed level most likely where the person is seeking employment (Ballantine, McCourt Larres & Ovelere, 2007; Gibbs et al., 2011, Grant et al., 2009). For these reasons, the practical assessment method was chosen. This approach allows participants to demonstrate their knowledge and avoid the traps of self-assessment and the abstract nature of a multiple-choice response.

When creating a new instrument, it is vital that it be constructed in such a way that the content accurately matches the aims and purposes of the test. The content analysis of these instruments began with the formation of the skill areas to be tested and continued with the formation of the questions. Each of the instruments was scrutinised by two panels of end-user experts and ranked on content suitability and difficulty level.

Conventionally, three types of validity can be established: construct, criterion and content (McGartland Rubio, Berg-Weger, Tebb, Lee & Rauch, 2003). In this study emphasis was placed on content validity.

Content validity, also known as logical validity, refers to the whether the items in an instrument adequately capture the entire domain that is intended to be represented in a test's score. In the present study, the concern is whether or not the items represent enough of the domain of skills in word processing and the use of spreadsheets. As McGartland Rubio et al., (2003) have suggested, and as it applies to the current study, content validity is the extent to which the items in a test adequately reflect a particular skill.

Although the term content validity is widely used, Beckstead (2009) argues that generally it is used incorrectly. He asserts that it would be more correct to regard content validity as content relevance. In a rebuttal, Squires (2009) claims that the criticism of the content validity statistic levelled by Beckstead is due mainly to content validity analysis being undertaken at the time the instrument is being developed, rather than after it has been used. Squires (2009) states that this practice is likely due to time constraints and financial imperatives. The latter, particularly, leaves researchers wishing to be sure of an instrument's robustness prior to its use in the field. While this may well be the reason driving this practice, it is not enough to say that an instrument's content is well validated based solely on an expert analysis; evidence of a need for further improvement may come to light once it is released for use with an actual study population (Beckstead, 2009). Thus, validation of content is just one step in the process of instrument validation, often undertaken prior to an instrument being used (Squires, 2009).

Many studies that discuss the process of validating content are studies from medical, nursing and social science disciplines however this process is no less import in the field of Information Technology. Sharp (2010) describes the process taken to assess the content of an instrument created to measure student's perceptions of their IT fluency. In their study Sharp (2010) assessed the content using a panel of experts to rate the relevance of items in a self-assessment instrument. While the testing of actual skill was not included in Sharp's study they concluded that the using an expert panel helped to define and structure their instrument and provided them with the information to make suitable changes.

Typically, the initial approach to evaluation of content is a non-statistical, such as peer or expert panel reviews (McGartland Rubio et al., 2003). The analysis is often broken into two parts. The first part involves the formation of the assessment questions or tasks to ensure that the assessment will cover all the important areas of the test area. Once the questions are formed, they are assessed as to how broadly they cover the subject area. The second part of the analysis involves the formation of expert panels to judge the validity of an instrument. The experts rate each item in an instrument as essential or non-essential for testing the particular skill. Ideally, a panel should comprise a combination of academic experts as well as those users who routinely use the software as a part of their daily work and are considered, by their peers, to be "experts" (McGartland Rubio et al., 2003). Squires (2009) noted that studies that use a panel of expert raters often give no explicit definition of the composition of the panel, which may affect perceptions of the content validity.

The optimal number of panel members has been debated in the literature. Debate about the number in a panel is, in part, due to the subjective nature of the data being collected. McGartland Rubio el al., (2003) state that a panel should be made up of between five and ten members, with an equal or near-equal number of academic experts and expert users. They suggest that a panel size of at least ten will go some

way toward countering the effect of individual subjectivity.

Several different methods for quantifying the level of expert agreement about content are outlined in the literature. These range from the relatively simplistic averaging of expert ratings of item relevance and comparing it to a pre-established acceptance criterion (Beck and Gable 2001) to the use of a multi-rater kappa coefficient to establish the level of agreement between ratings (McGartland Rubio et al., 2003).

One widely used method is the Content Validity Index (C_{VI}). The popularity of the C_{VI} is due, in part, to the ease with which it can be calculated and understood, and because of its emphasis on the assessment of relevance. A C_{VI} can be calculated for each individual item in an instrument or for the instrument as a whole. However, some authors have expressed concern that the C_{VI} focuses on item relevance but does not take into consideration whether or not the instrument consists of items that comprehensively measure what the instrument is intended to measure (Polit & Beck, 2006). Another frequent criticism of C_{VI} is that no consideration is made for chance agreement (Watkins & Pacheco, 2000, McGartland Rubio et al., 2003; Polit & Beck 2006), which can result in an inflated view of content relevance (Beckstead, 2009).

Measures of inter-rater agreement that do take into account the probability of chance agreement include the widely used Cohen's kappa coefficient calculation (Viera & Garrett, 2005), however this statistic does not discriminate between ratings of relevance and non-relevance (Polit, Beck & Owen, 2007).

Polit el al., (2007) proposed a modified kappa that incorporated chance agreement on relevance alone. Because Polit et al.'s (2007) approach accounted for chance agreement on the quality of most concern in this study (relevance); it was decided to use their method in this study. At the same time, items were also assessed for difficulty level. This step was considered important in the development of this instrument in order to ensure that a range of skill levels could be assessed.

3. Method

Two skill assessment instruments were created consisting of a 16 task spreadsheet skill assessment and a 12 task word processing skill assessment. This process involved two panels of user experts.

3.1 Expert panel formation

Two separate and independent panels of end-user computer experts were formed for the purposes of the study. The first panel (Expert Panel 1) was used to test the clarity of the instructions and tasks. This panel was asked to complete the instruments as a user would. The panels were composed of ten users for the word processing component and eleven for the spreadsheet component. It included a mixture of academic experts and expert users (see Table 1). No members were on both panels.

Table 1. Composition of Panel 1

Spreadsheet Experts	**11 users**
Academic Experts	3
Expert Users	8
Word Processing Experts	**10 Users**
Academic users	3
Expert Users	7

The academic panel members were people involved in the teaching of end-user application software at a tertiary level. Expert users were people who used the software to a high level in their employment. They were identified as the "go-to people" in their organization; i.e., those who others would ask for help.

A second panel of end-user specialists (Expert Panel 2) was formed to assess the degree to which the objectives of each instrument were met by each task in that

instrument; i.e., the relevance of the items. This panel consisted of seven members, all of whom were involved in the teaching or workplace training of end-user computer skills and none of whom was involved in Panel1. (Table 2)

Table 2. Composition of Panel 2

Expert User Panel 2	7 Users
Corporate computer trainers	5
Academic End-user teacher	2

3.2 Instrument development process

A development exercise was undertaken to ensure that the content of the test instruments was valid for the intended purpose. A two-part process was used to validate the test instruments. For each part, a panel of expert users was formed.

Instrument development was an iterative process outlined in Figure 1. Each step in the process involved consultation with domain experts.
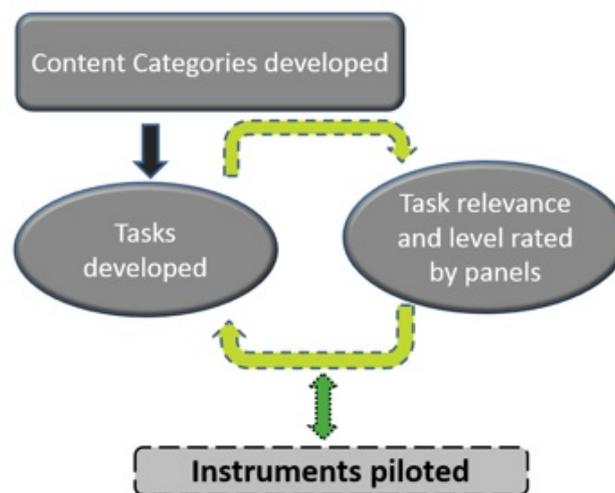

Figure 1. Iterative instrument development process

Step one involved the defining of category content for each application (Table 1).

Table 3. Skill Classifications

Spreadsheet Classification	Word-Processing Classifications
Cells and formatting	Text and paragraph formatting
Functions and formulas	Styles
Charting	Numbered lists
Sorting and Filtering	Tables
Validation and protection	Page Layout

Step two involved the definition of tasks to fit each category. In step three the instrument tasks were rated by the panels and piloted by typical end-users. Each of these steps went through a number of iterations until agreement was reached by the expert panels.The inclusion of the development stage plus the latter two stages of content relevance help to ensure that this instrument has been assessed as thoroughly as is possible at this stage of development. Although content relevance is subjective, the method used in this study has added a level of objectivity. Panels of experts can provide researchers with valuable information to revise a measure.

3.3 Panel tasks

Panel 1 was divided into two sub panels: spreadsheet experts and word-processing experts. These sub panels were each asked to work through the instruments under the same conditions a user would. Each panel member undertook the task in isolation ensuring that no collaboration between panel members took place. Panel members

were asked to rate each task as essential or not and basic or moderately advanced.

Panel 2 members were not required to complete the tests but were asked to rate each of the skills as either essential to test or not as well as judging the difficulty level of each skill. (Table 4)

Table 4. Relevance and difficulty ratings

Essential to test	Difficulty
1 Yes	1 Moderately advanced
2 No	2 Basic

Members of both panels were asked to suggest skills that they thought were essential to test but had been omitted from the instruments.

After each iteration, non-relevant tasks were removed and those suggested by the panels were added. The final instruments consisted of a fifteen task spreadsheet instrument and an eleven task word-processing instrument.

3.4 Content validity analysis

The results from all panels were assessed using the modified kappa (k*) described in Polit et al., (2007). The method of calculation can be found in that article. The variation to the usual kappa formula is to substitute a probability for agreement on relevance alone, instead of chance agreement regardless of the direction of the decision. The strength of the modified kappa index were compared to the values used by Polit et al., (2007). And are displayed in Table 5.

Table 5. Modified Kappa evaluation (Polit et al., 2007).

κ	Evaluation
< 0	Poor agreement
0.01 – 0.39	Slight agreement
0.40 – 0.59	Fair agreement
0.60 – 0.74	Good agreement
> 0.74	Excellent agreement

4. Results

The results of the content relevance analysis are discussed in the sections that follow.

4.1 Spreadsheet skill relevance assessment

The content validity analysis comparison between Panel 1 and Panel 2 revealed some differences amongst the ratings, both between and within the panels. While there was near uniform agreement amongst Panel 1, Panel 2's ratings varied. This lower proportion of agreement from Panel 2 on some of the items is reflected in the range of k* scores. Overall, however, the suggestion was that the test content was close to meeting the requirements of both panels of experts and could be considered fit for the purpose for which is was created. These results are displayed in.

Table 6. Relevance, Task evaluation and Difficulty results for spreadsheet assessment (first iteration)

Task	Relevance	Panel 1 Evaluation	Difficulty	Relevance	Panel 1 Evaluation	Difficulty
1. Basic Cell formatting	1	excellent	1	0.85	excellent	0.9
2. Use simple SUM function	1	excellent	1	0.85	excellent	0.9
3. Use COUNT function	0.9	excellent	1	0.85	excellent	0.9
4. Create a conditional formula to count values	0.9	excellent	0.2	0.66	good	0.1
5. Use TEXT functions	1	excellent	0.3	0.26	slight	0.3
6. Used fixed cell referencing in a formula	0.9	excellent	0.4	1	excellent	0.8
7. Apply cell formatting according to a set criteria	0.8	excellent	0.2	0.66	Good	0.6
8. Use a conditional formula to return values	1	excellent	0.3	0.41	fair	0.3
9. Name a cell or range of cells	0.9	excellent	0.9	0.41	fair	0.3
10. Create a simple column chart	1	excellent	0.8	0.85	excellent	0.8
11. Create a one series column chart with a numeric X-axis	1	excellent	0.5	0.66	good	0.6
12. Sort Multi column data	1	excellent	0.4	0.85	excellent	0.7
13. Use a Lookup function in a formula	1	excellent	0.2	0.66	good	0.3
14. Apply a simple filter	0.9	excellent	0.7	0.66	good	0.4
15. Create a validation rule	0.9	excellent	0.2	0.41	fair	0.1
16. Apply cell protection	0.9	excellent	0.2	0.41	fair	0.2

Although the results for spreadsheet items from each panel in this study were internally consistency with a Cronbach's Alpha score of 0 .75 there was some disagreement between panels as to both the relevance and difficulty of items in this first iteration.

Members of Panel 1 were in complete agreement that all of the tasks presented were relevant, however this view was not shared by members of panel 2. Of the sixteen tasks presented five tasks received less than majority support from members of Panel 2. This results shows the value of having two panels, with two differing points of view and gives the chance for task modification or removal in order to finalise an instrument where both panels have similar agreement.

There was similar disagreement between panels regarding the difficulty levels of the sixteen tasks present. Panels were in agreement on the difficulty level of ten of the fifteen spreadsheet tasks. Of the remaining tasks, panel 2 considered five tasks to be at a lower level than panel one members did. There is, however, a relatively even spread of difficulty within the tests, which supports the notion that the test is construct valid.

Table 7 . Spreadsheet assessment tasks for final instrument

Spread-sheet skills test tasks

1. Basic Cell formatting.
2. Using a Sum function.
3. Using a Count function.
4. Using a conditional function to count values.
5. Using Text Functions.
6. Using a fixed cell reference in a formula.
7. Create a pivot table with one summary value
8. Using a conditional function to return value.
9. Applying a name to a cell or range of cells.
10. Creating a simple Column chart using adjacent ranges
11. Creating a one series column chart with numeric x axis.
12. Sorting multi column data.
13. Using a Lookup function in a formula.
14. Using a simple filter.
15. Creating a validation rule

For the final iteration of the instrument in this content validity exercise the scores from panels were combined. The results are shown in Table 8.

Table 8. Relevance, Evaluation and Dif"ulty results for final spreadsheet assessment

Task	Relevance $k*$	Evaluation	Difficulty
1	0.93	Excellent	Basic
2	0.93	Excellent	Basic
3	0.88	Excellent	Basic
4	0.78	Excellent	Moderately Advanced
5	0.63	Good	Moderately Advanced
6	0.95	Excellent	Basic
7	0.73	Good	Basic
8	0.71	Good	Moderately Advanced
9	0.66	Good	Moderately Advanced
10	0.93	Excellent	Basic
11	0.83	Excellent	Moderately Advanced
12	0.93	Excellent	Basic
13	0.75	Good	Moderately Advanced
14	0.82	Excellent	Basic
15	0.75	Good	Moderately Advanced

4.2 Word-processing skill assessment

The same process followed for the spreadsheet skill assessment was used for the word processing skill assessment. Based on the relevance scores, Table 9 shows that all panelists agreed that each of the tasks was relevant to testing a person's word-processing skills.

Table 9 . Relevance of tasks, evaluation and difficulty for all word processing task

Task	Panel 1 Relevance	Panel 1 Evaluation	Panel 1 Difficulty	Panel 2 Relevance	Panel 2 Evaluation	Panel 2 Difficulty
1. Apply basic paragraph formatting	1.00	excellent	1	0.85	excellent	1
2. Change a paragraph's alignment	1.00	excellent	1	0.85	excellent	1
3. Copy formatting between paragraphs	1.00	excellent	1	0.85	excellent	0.8
4. Indent a paragraph	1.00	excellent	0.9	0.85	excellent	0.9
5. Create multi-level lists	1.00	excellent	0.7	0.63	good	0.7
6. Apply pre-set styles to text	0.90	excellent	0.9	0.84	excellent	0.7
7. Modify pre-set styles	0.90	excellent	0.4	0.84	excellent	0.4
8. Create new styles	0.70	good	0.3	0.84	excellent	0.4
9. Create tables	1.00	excellent	0.7	1	excellent	0.9
10. Sort table data	0.80	excellent	0.8	0.84	excellent	0.9
11. Position images within text	0.90	excellent	0.4	1	excellent	0.6
12. Work with section within a document	1.00	excellent	0.4	0.84	excellent	0.4

The results for word-processing items from each panel in this study were internally consistency with a Cronbach's Alpha scores of 0 .70. As with the spreadsheet skills assessment, there was some disagreement between the panels, although in this case this was minor. Some members of Panel 1 thought that it was not necessary for a person to be able to create a new style (Task 8) while the members of Panel 2 rated this skill as absolutely essential. Panel 2, on the other hand, were in less agreement about the relevance of creating multi-level lists.

The majority of member's panel1 thought that eight of the twelve tasks were things that everyone should know while the majority of panel two indicated that nine tasks fitted this category. The closeness in results indicated that instrument did contain word-processing tasks that varied in level of tasks from basic to difficult. After further iterations of this process eleven tasks were defined as being essential to test. These tasks are shown in Table 10.

Table 10. Word-processing assessment tasks for final instrument

Word processing tasks

1. Applying basic paragraph formatting.
2. Changing a paragraph's alignment
3. Copying formatting between paragraphs.
4. Indent a paragraph.
5. Creating multi-level lists.
6. Applying pre-set styles to text.
7. Modifying a pre-set style.
8. Inserting text and page number into page footer
9. Updating a table
10. Positioning an image within a block of text.
11. Working with sections within a document.

For the final iteration of the instrument in this content validity exercise the scores from panels were combined. The results are shown Table 11.

Table 11. Content Validity results for final Word- processing assessment

Task	Relevance (k)	Evaluation	Level
1	0.93	Excellent	Basic
2	0.93	Excellent	Basic
3	0.93	Excellent	Basic
4	0.93	Excellent	Basic
5	0.82	Excellent	Moderately Advanced
6	0.87	Excellent	Basic
7	0.87	Excellent	Moderately Advanced
8	0.77	Excellent	Basic
9	1.00	Excellent	Basic
10	0.82	Excellent	Moderately Advanced
11	0.95	Excellent	Moderately Advanced

5. Conclusion

This paper demonstrates how to conduct a content relevance assessment for skills tests for two of the most common used end-user computing tools, spreadsheets and word-processing software. The method used was a multi-step iterative approach consisting of a development stage and a judgment-quantification stage using panels of end-user computing experts.

The development stage consisted of the test areas being formulated and the questions created. Once the questions were formed they were assessed as to how broadly they covered the subject area. The second part of the analysis involved the judgment quantification stage, which involved the formation of expert panels to judge the relevance of each instrument. Panel members were also asked to rate the difficulty level for each skill being tested. This step was added to help ensure that all levels of skill could be tested in order to give a more accurate level of skill. Panel members were also asked to contribute any skill area from either test instrument that had been omitted from the original instruments. This stage in the process helps to ensure that vital items are not omitted from an instrument.

The inclusion of the development stage plus the latter two stages of content relevance help to ensure that this instrument has been assessed as thoroughly as is possible at this stage of development. Although content relevance is subjective, the method used in this study has added a level of objectivity. Panels of experts can provide researchers with valuable information to revise a measure. Certainly the panels used in this study have performed this function well. The process of assessing content relevance is often carried out in the period prior to an instrument being used on a test population however the process of validating a measure should be treated as a never-ending process.

The process of analysing panel feedback using a kappa index modified to account for chance agreement on relevance, allowed for an iterative development system that allowed for the creation of instruments that contained relevant tasks aim at users with varying skill levels. There is a risk, when using non-verified instruments to assess skill that the results will not accurately reflect the level of knowledge a user or group of users has. Mixed results can be expected with this type of review and the level of agreement between the experts in this study is considered good and has substantiated the use of this process as being robust and thorough.

Although this paper concentrated on spreadsheet and word-processing assessments it would be useful to broaden this process to take into account a number of other applications and areas now common for many workplaces, such as social media, databases or web applications.

Overall, the results demonstrate the importance of undertaking a rigorous process in order to establish instruments that meet the purpose for which they are being designed. Without such a process, it can be difficult to determine the validity and therefore reliability of instruments being used to assess a person's skill level in important situations such as part of the employment process. Although literature offers guidelines and recommendations for validating test instruments, the tests developed and discussed in this paper have contributed to existing knowledge on this particular subject and have given users of the instrument a reliable method of assessing and assigning end-user skill level.

References

Autrey, K., Tarver, R., Myers, L. A., & Tarver, M. B. (2004). Using Microsoft office specialist certification to enhance employment opportunities for college students. In World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (Vol. 2004, No. 1, pp. 1068-1069).

Ballantine, J., McCourt Larres, P., & Oyelere, P. (2007). "Computer Usage and the Validity of Self- Assessed Computer Competence Among First Year Business Students". Computers and Education, 49, pp. 976 - 990.

Beck, C.T., & Gable, R.K. (2001). Ensuring content validity: An illustration of the process. Journal of Nursing Measurement, 9, pp. 201-215.

Beckstead, J. 2009. Content Validity is naught. International Journal of Nursing Studies, 46, pp. 1274 - 1283.

Calzarossa, M. C., Ciancarini, P., Maresca, P., Mich, L., & Scarabottolo, N. 2007. The ECDL programme in Italian universities. Computers & Education, 49(2), pp. 514-529.

Chambers, C., & Scaffidi C. (2010). "Struggling to Excel: A Field Study of Challenges Faced by Spreadsheet Users". IEEE Symposium on Visual Languages and Human-Centric Computing VL/HCC, Leganes, Madrid, Spain, IEEE, pp. 187 -194.

Davis, P. V., & Cleere, G. (2003).The ECDL Test Development and Validation Process. Retrieved from: http://e-assessmentlive2009.org.uk/pastConferences/2003/procedings/Davis.pdf on 20/08/2011

ECDL: European computer Drivers Licence http://www.ecdl.com

Gibbs, S., Steel, G., & Kuiper, A. (2011). "Do New Business Graduates Have the Computing Skills Expected by Employers? Proceedings: The 2nd International Conference on Society and Information Technologies, Orlando, Florida, March 2011.

Grant, D., Malloy, A., & Murphy, M. (2009). A Comparison of Student Perceptions of their Computer Skills to their Actual Abilities. Journal of Information Technology Education, 8, pp. 141 - 160.

Holtzman, D.M., & Kraft, E.M. (2010). Skills Required of Business Graduates: Evidence from Undergraduate Alumni and Employers. Business Education and Administration, 2 1, pp. 49-59.

McGartland Rubio D., Berg-Weger M., Tebb S., Lee, E., & Rauch S. (2003). Objectifying content validity: Conducting a content validity study in Social Work Research. Social Work Research, 27 2, pp. 94-104.

Lawson, B.R., Baker, K.R., Powell, S.G., & Foster-Johnson, L. (2009). A comparison of spreadsheet users with different levels of experience. Omega, 37 3, pp. 579-590.

McGill, T., & Dixon, M. (2004). Information technology certification: A student perspective. Proceedings of International Resource Management Association. pp. 302-306.

McLay, A., & Brown, K. (2006).A Look into the Integration of the ICDL Program into the Workplace: It's a Team Thing! New Zealand Association for Cooperative Education Conference, Queenstown. pp. 22 -31.

Microsoft Office Specialist (MOS) http://www.microsoft.com/learning/en-us/mos-certification.aspx

Panicos M., & Sotiris A. 2010. The design and implementation of an in-application automated testing and evaluation system for computer literacy skills based on the European and international computer driving license (ECDL/ICDL). International Conference on Education and New Learning Technologies, Barcelona, Spain

Pascoe, R. (2003, July). Is there a MOUS in your house. In th annual NACCQ Conference, Christchurch.

Skill Assessment Manager (SAM : http://www.cengage.com/samoffice2013/

Townley, S.A. (2004) European Computer Driving License. pp 1145, In Caesar, C.G & Scott, D.H.T (2004). Strength of disposable laryngoscopes. Anesthesia, 59(11) pp. 1144 -1145.

Polit, D.F., & Beck, C.T. (2006). The Content Validity Index: Are You Sure You Know What's Being Reported? Critique and Recommendations. Research in Nursing and Health, 29, pp. 489 -497.

Polit, D.F., Beck, C.T., & Owen, S.V. (2007). Focus on Research Methods. Is the C_{VI} an Acceptable Indicator of Content Validity? Appraisal and Recommendations. Research in Nursing and Health, 30, pp. 459-467.

Sharp, M. (2010). Development of an instrument to measure student's perceptions of information technology fluency skills: Establishing content Validity. Perspectives in Health Information Management, 1-10. Accessed 26/05/2012 http://search.proquest.com/docview/746603195?accountid=27890

Squires, A. (2009). A valid step in the process: A commentary on Beckstead 2009. International Journal of Nursing Studies, 46, pp. 1284-1285.

Townley, S.A. (2004). European Computer Driving Licence. pp 1145, In Caesar, C.G & Scott, D.H.T (2004). Strength of disposable laryngoscopes. Anesthesia, 59(11) pp. 1144 -1145.

Vakhitova, G., & Bollinger, C. R. (2011). Labor market return to computer skills: Using Microsoft certification to measure computer skills (No. 46).

Viera, A. and Garrett, J. (2005). Understanding Interobserver Agreement: The Kappa Statistic. Family Medicine Research Series, 37 5, pp. 360 -363.

Watkins, M, and Pacheco, M.E. (2000). Interobserver Agreement in Behavioral Research: Importance and Calculation. Journal of Behavioral Education, 10 4, pp. 205-212.