Refinement of the TELT survey instrument

Stage 3, iteration 1

Prepared by Mariya Pachman, Anatoly Logunov & Stephen Quinton

29 October 2010
Introduction

The UNSW Technology Enhanced Learning and Teaching evaluative framework, introduced late 2009 (see – Evaluation of the TELT platform: essential elements and methodologies doc), emphasised the importance of embedding iterative processes when evaluating educational technologies. In its final format, the evaluative process was determined to be tri-fold – pedagogical, technological, and business processes where the iterative components are embedded into the pedagogical layer (sub-layer 1) of Stage 3.

The first iterative component involves the use of a survey instrument that is designed to reflect students’ opinions about educational technologies in holistic context (that is, accounting for their previous experience with technology; preconceived notions of learning; resistance to implied authorities in relation to learning and teaching; preconceived notions about eLearning; and the flexibility of the technology application in relation to learning and teaching).

Refinement of the instrument begins by initiating the iterative process. Second, such an instrument should be applicable to a broader range of educational technologies particularly in studies where for example, a comparison of different technologies (as opposed to the evaluation of a single technology) is required. This aspect addresses the need for valid and reliable measurement tools for evaluating educational technologies (a point that is further discussed in the Theoretical considerations subsection to follow). Finally, the instrument allows the synthesis of the contextual factors (mentioned above) in order to provide a clearer picture of the student population in relation to their perceptions of educational technologies as well as their overall ideas about learning and teaching.

One of the practical considerations made during the initial planning stage included the size of the survey: it was limited to approximately thirty items given that participation was voluntary and the majority of students were not receiving additional course credits for participating in the evaluative activities.

Theoretical considerations

Locating valid and reliable measurement instruments appear to be problematic for the field of educational technology. First, part of the difficulty may be attributable to the fact that CBI (computer-based instruction) presents a different environment in comparison with face-to-face learning and teaching. To be specific, multimedia capabilities and interaction effects inherent to the CBI environment could to similar degrees be instrumental in influencing instructional and motivational factors (Newby et al, 2000; Relan, 1992). Moreover, the learner controls described by Klein and Keller (1990) are yet another factor that makes the measurement of different variables in CBI delivery a challenge.
A second part of the problem is to be found in the application versus the research focus of educational technology studies. Many popular survey instruments are applied by researchers without establishing or examining validation data. Anecdotal evidence that supports this statement includes the measurement of learning motivation, affective experiences in distance learning courses, and validation of evaluative scales for Course Management Systems (CMS).

In the case of motivational surveys, a well-used instrument (Instructional Material Motivational Survey, Keller, 1983) was validated by a group of researchers and involved 875 subjects (Huang, Huang, Diefes-Dux, & Imbrie, 2006). Out of the 36 original items (brief version), only 20 proved to be statistically significant with one construct reflected by only one survey item. Two factors were highly inter-correlated and although the measurement model was intended to account for these construct-interaction effects, the overall conclusion suggested by the authors was “that it is necessary to revisit each item in the original IMMS to apply it to a computer-based tutorial setting” (p. 254). This result suggests the need to devise new survey instrument.

In an earlier case, (Thomerson & Smith, 1996) verified that researchers can fail to attempt to validate their own instruments in terms of the latent variables. Their scales comprised statements taken from different research scales and they simply interpreted the results (findings) and mentioned the high reliability of their own scale (Cronbach alpha = .8938). It should be noted that as a field of study, educational technology progressed substantially from 1996 when this study was first published. However, despite the lengthy interval since, the evaluation of different educational technologies still presents difficulties for educational administrators and decision-makers.

There is also no clarity as to which factors influence the evaluation and perception of technologies. In the main, the factors that researchers seek to establish in the evaluation of educational technologies include: student–instructor interactions during the course and students’ scholastic aptitude (Hong, 2002); instructor quality; information content quality; system quality; service quality; learner’s attitude and supportive issues (Ozkan, & Koseler, 2009), or computer self-efficacy; system functionality; content feature; interaction; performance expectations; learning climate and learning satisfaction (Wu, 2010). Around the same time, another recent research study that focused on CMS as a technological application lists a different set of factors: user satisfaction, feedback, personalised environment, response time, and ease of use (Ioannou & Hannafin, 2008). Moreover, these studies show that some include contextual factors such as learner’s attitude, while there are studies that do not. The generic

---

1 Overall, the construct subscales are developed to include at least three items (Clark & Watson, 1995). Thus, this subscale needs to be further developed, i.e. more items assessing this specific construct should be included (DeVellis, 2003).
evaluative framework developed at UNSW has addressed this deficiency. Thus, along with a growing need to provide guidelines on which educational tool to use, and when and how to use it (Quillen, 2010), there is a need to recognise that the development of a survey instrument is a crucial step in achieving the institution’s educational and pedagogical goals.

The Scale

Nielsen’s heuristics (1994) and Benson, Elliot, Grant, Holschuh, Kim, and Lauber’s work (2002) served as the basis for development of the usability survey items incorporated into the UNSW TELT Evaluation framework. Nielsen’s heuristics (1994) have been widely employed in usability evaluation, including evaluating educational applications. For example, Peng, Ramaiach and Foo (2004) performed a Nielsen’s heuristic-based user interface evaluation (N=88) of the electronic media services gateway at Nanyang Technological University (Singapore). Sim, Read and Cockton (2009) also adapted Nielsen’s model (1994) as a framework for “synthesising a set of domain specific heuristics for evaluating CAA applications” (Sim, Read & Cockton, 2009, p. 204). For the assessment of the emotional aspects of interaction with the system, we used some of Benson et al. (2002); Saadé, He and Kira (2007); and Venkatesh, Morris, Hall, Davis, Davis and Walton’s developments (2003). Benson et al. (2002) used an expert evaluation with eight panel members to apply usability heuristics to learning software.

Saadé et al. (2007) studied the dimensions of online learning and system usage through the four constructs of student attitude, affect, motivation and perception with undergraduate students (N=105). They concluded that the affect and perception dimensions in their model have strong measurement capabilities. In addition therefore, we designed four items to assess student resistance to implied authorities in relation to one’s teaching and learning on a 7-point Likert response scale (from 1 = strongly disagree to 7 = strongly agree). The inclusion of the preconceived notions about eLearning items were influenced by the work of Ainley, Banks and Fleming (2002) and Baylor and Ritchie’s (2002) educational computer usage categories. Finally, two other items were designed to assess flexibility of application in relation to teaching and learning. Two content experts were then asked to review the scale separately. Based on their feedback, the scale was revised and finalised.

In completing the survey, participants were required to express their degree of agreement or disagreement with each statement by selecting one of the following seven options: 1 – Strongly Disagree; 2 – Disagree; 3 – Somewhat Disagree; 4 – Neither Agree or Disagree; 5 – Somewhat Agree; 6 – Agree; and 7 – Strongly Agree. Participants were also invited to provide non-mandatory short qualitative feedback alongside any particular statement. A seven-point Likert scale was chosen as the most appropriate response tool. Many respondents were deemed to be familiar with the Likert scale,
which was selected for its traditional usefulness, clarity of understanding, and statistical robustness. Measurement studies are quite unanimous in claiming the seven-point scale as an optimal scale in relation to the number of response options, and as the scale with maximum reliability (Cox, 1980; Preston & Colman, 2000).

A total of 152 items was originally developed for survey purposes. However, as we were limited to a suggested range of 30 questions to be included in this pilot, the total number of questions related to the contextual factors was assumed to be around 22 items, and 8 demographic items (learners’ previous experience with technology) served as a basis for a division on learners groups (see TELT educational technologies pilots: brief summary of findings doc).

**Procedures**

The participants involved staff and students from four different faculties (COFA, FASS, Science and Engineering). Completion of the survey was voluntary. A total of 273 students and 4 staff members participated in the study (36.4% of the original 760 invitations). During the semester (Semester 1, 2010), participants completed an anonymous online survey during the final four weeks of classes. The survey remained available two weeks after the end of the semester.

Exploratory Factor Analysis (EFA) is particularly appropriate when the researcher does not know how many factors are necessary to explain the interrelationships among the measured variables (Pett, Lackey, & Sullivan, 2003). Thus, an EFA was carried out to determine the factor structure of the data, and to decide which items should be eliminated from the different subscales and which items should be refined.

Prior to conducting the EFA, descriptive statistics was run to get a feel for the data. The means and standard deviations looked good. Means were in the range of 3.2 to 5.26 on a 7-point Likert scale, and standard deviations were generally larger than 1.2, indicating that we had achieved acceptable variability in the way people responded. The correlation matrix was also closely examined for consistency of the survey items, and to identify items that were too highly correlated ($r \geq .80$; Pett et. al., 2003). No considerable redundancies were found among the survey items.

EFA with an oblique rotation - Promax with Kaiser Normalisation ($k = 4$) was conducted with an original survey scale. Evaluation of the correlation matrix indicated that it was factorable. The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy value was .891, which closely corresponds to “marvellous” ($\geq .90$) according to Kaiser’s criteria (Pett et. al., 2003). The MSAs values on the diagonal of the anti-image correlation matrix were generally in 80s ($\geq .80$) and 90s ($\geq .90$), which corresponds to “meritorious” and “marvellous” according to Kaiser’s (1974) criteria. The Bartlett’s Test of Sphericity ($\chi^2 = 2569, df = 231, p < .000$) was significant. This test showed that the data
used in this study did not produce an identity matrix and thus were multivariate normal and acceptable for applying factor analysis.

The number of factors to extract was determined based on Eigenvalues (Eigenvalue > 1). As a result, five factors were extracted that accounted for 60% of the total variance in the items with a weak loading of items # 11, 27, and 14, and a marginally weak loading of #8 (see Table 1). By weak loadings we refer to loadings of .30 to .45 on one or two factors. By marginally weak loading we mean loadings of .30 to .51 on one or two factors. In this case, the advised cut-off score for interpretation purposes is often about .40 or greater (Stevens, 2002). Thus, although we retained all the above items, the refinement process described below included re-wording these items or replacing them with similar items from the originally developed question pool. The five factors\(^2\) that were extracted include:

1. Usability (items 13, 15, 16, 19, 10, 25, 28, 11, 27)
2. Feelings toward an application (emotional aspects of technology evaluation) – items 18, 26, 21, 12, 14
3. Resistance to implied authorities in relation to one’s teaching and learning (items 29, 30, 23, 8)
4. Preconceived notions about eLearning (items 6, 7)
5. Flexibility of application in relation to the teaching and learning (items 24, 22).

Since certain questions did not clearly and strongly load on one factor, we re-worded them taking into consideration students’ qualitative data (from their comments).

**Validity considerations**

Question #11 was originally intended as a usability item (factor 1), but was equally (and weakly) loading on both factor 1 and factor 2 (Feelings toward an application). The original version of this question – “I think the application is not consistent with my usual usability experience with other software and websites” was changed to “Based on your overall experience with the use of other software and websites, how would you rate this application” with a Likert scale from 1 (Very easy) to 7 (Very difficult). It is anticipated that this refinement will load stronger on factor 1.

Question #27 was originally intended as a reflection of participants’ opinions about the system. However, it was weakly loading on factor 1 (Usability) and less consistently on factor 2 (Feelings toward an application). The original version of this question – “Having the application as part of teaching this course, or learning this course, frustrates

---

\(^2\) In this document the words factor and subscale are used interchangeably.
Table 1: Pattern matrix generated from EFA solution with oblique rotation - Promax with Kaiser Normalisation (κ = 4)

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
<th>Factor 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. The way the information and resources are organised within the application easily integrates with the way I naturally organise or access my information and resources.</td>
<td>.910</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. The process of searching for, finding and accessing the information I needed was clear and intuitive</td>
<td>.873</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. The application allows me to perform my tasks quickly.</td>
<td>.837</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. In considering the usability and usefulness of the features and functions of this application, how strongly would you agree that you thought there were NO problems that you particularly identified?</td>
<td>.791</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. I think the application has a pleasant, appealing look and feel.</td>
<td>.789</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. The application integrates well into my learning or teaching approach and processes without me having to try look for alternative ways of integrating it and making use of it in my own individual way.</td>
<td>.732</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28. I think other students/staff/users would find this application usable and useful.</td>
<td>.606</td>
<td>.360</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. I think the application is consistent with my usual usability experiences with other software and websites.</td>
<td>.454</td>
<td>.310</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. Having the application as part of teaching this course, or learning this course, frustrates me</td>
<td>.335</td>
<td>.310</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. I thought about referring to help and documentation during my use of application.</td>
<td>-1.031</td>
<td>.310</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. I felt somewhat apprehensive about using the application</td>
<td>.719</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. I am not sure how this application, its features and its content are meant to integrate with my teaching or learning</td>
<td>-.583</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. The application makes me complete irrelevant, superfluous or distracting tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Items in italics are discussed in the Validity subsection while items in bold and italics are discussed in the Reliability subsection.  
* Reverse items
that are not related to my learning (or teaching) processes / I think there is too much irrelevant information and features

14. It is a tedious process to access information and resources using this application

29. I do not believe that based on my opinions any serious improvements will be applied to the application

30. I can conclude that the institution does not understand my unique learning or teaching approach and learning or teaching processes

23. If an application doesn't integrate well into my learning and teaching approach, I don't try to look for alternative ways of integrating it and making use of it in my own individual way, because that is not what I should be required to do

8. I did not really explore or use any application resources or features that were suggested to me through training resources or course guides and by other people

6. Some people say that eLearning applications are just tools that present existing information, and are of little value in student construction of new knowledge, collaboration and deep learning. Based on your experiences with this application, would you agree?

7. Some people say that eLearning applications are valuable tools that encourage the active collaboration between students and staff, exploration of different opinions and promote a culture of social learning. Based on your experiences with this application, would you agree?

24. Even if an application does not integrate well into my learning or teaching approach and learning or teaching processes, it is important that it is flexible enough so that I can integrate and use it in my own individual way

22. I believe I understand, and am well informed of the overall way this application, its features and its content should integrate with student learning and teaching
“me” was changed to “Using this application unreasonably complicates my learning or/and teaching in this course” in order to load stronger on factor 1.

Question # 14 was also originally intended as a usability item (factor 1), but was weakly, more consistently loading on factor 2 (Feelings toward an application) than on factor 1. The original version of this question – “It is a tedious process to access information and resources using the application” was changed to “It is somewhat frustrating to access information and download resources using this application” so that it would load stronger on factor 2.

Question # 8 that was initially intended to reflect to what degree students explore the application, was in fact loading on factor 3 (Resistance to implied authorities in relation to one’s teaching and learning). From students’ comments it was evident that the question was perceived as an item referring to the course design rather than features of a technological application. The original version of this question – “I did not really explore of use any application resources or features that were suggested to me through training manuals, online resources or course guides and by other people” was changed to “My use of the application was guided by training aids and technical documentation”.

Reliability of the scale

Following the factor analysis, Cronbach’s alpha reliability analysis was conducted on each subscale (see Table 2). Cronbach’s coefficient alpha (Cronbach, 1951, 1984) is an important and widely used measure of internal consistency in a set of items (Pett, et al., 2003), and represents the proportion of a scale’s total variance that is attributable to the true score of the latent construct being measured (Netemeyer, Bearden, & Sharma, 2003). An acceptable reliability coefficient used in social sciences is .70 (Netemeyer, et al., 2003; Nunnaly, 1978). Thus, caution needs to be taken in interpreting the Resistance to implied authorities in relation to one’s teaching and learning. 

Table 2: Cronbach’s alphas reliabilities based on retained items (N = 277)

<table>
<thead>
<tr>
<th>Scale</th>
<th># Items</th>
<th>Alpha</th>
<th>Means Inter-Item Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Usability evaluation</td>
<td>9</td>
<td>.89</td>
<td>.49</td>
</tr>
<tr>
<td>2. Feelings toward an application</td>
<td>4</td>
<td>.76</td>
<td>.45</td>
</tr>
<tr>
<td>3. Resistance to implied authorities in relation to one’s teaching and learning</td>
<td>3</td>
<td>.58</td>
<td>.32</td>
</tr>
<tr>
<td>4. Preconceived notions about eLearning</td>
<td>2</td>
<td>.52</td>
<td>.37</td>
</tr>
<tr>
<td>5. Flexibility of application in relation to the teaching and learning</td>
<td>2</td>
<td>.32</td>
<td>.19</td>
</tr>
</tbody>
</table>
learning, the Preconceived notions about eLearning, and the Flexibility of application in relation to the teaching and learning subscales. Moreover, Clark and Watson (1995) advised against subscales that include only two items.

The other measures used for reliability analysis:

1. We inspected the inter-item correlation matrix for possible highly correlated items. Correlations larger than .8 between two items suggest redundancy, and therefore the researcher should consider deleting one of the items.
2. We inspected the mean and standard deviation of the inter-item correlations (Table 2). Ideally the average inter-item correlations should be in the range of .4 -.7 (Pett et. al, 2003).

Again, the last three subscales had slightly lower than recommended inter-item correlations averages although no scale items appeared redundant.

Question # 23 that was negatively (but strongly) loading on factor 3 (see Table 1), actually decreased the reliability of the subscale. It could be considered a “bad item” (see Santos, 1999 for a discussion). Thus, the number shown in Table 2 actually refers to the reliability of the other three items on this subscale with the exception of question #23. Question #23 was replaced with “I should not be asked to modify my learning and/or teaching approach in order to work effectively with this technology application” question from the originally developed question bank in order to increase the reliability of this subscale.

Following the recommendations of Clark and Watson (1995), we also included one additional question for each of the following subscales: Preconceived notions about eLearning and Flexibility of application in relation to the teaching and learning. The decision to straighten those subscales instead of excluding them from the questionnaire was dictated by the theoretical and practical importance of measuring those constructs. Firstly, students are often considered tabula rasa in relation to educational technologies, and their previous background is not taken into consideration. This might lead to a situation where their “objective” estimates reflect prior experiences with different technologies as opposed to their actual evaluation of these technologies. Secondly, the long term goal of educational technology units at the institutional level may well be to change these preconceived notions and to improve students’ understanding of technologies. With this goal in mind, the new items for subscales 4 and 5 were: “Some people say that eLearning applications enable the effective acquisition of practical knowledge and skills that students can then apply in real life situations. Based on your experiences with this application, would you agree?” (subscale 4) and “This application is not flexible enough to integrate with my learning and teaching approach”(subscale 5).
Interpretation of the Resulting Subscales

The 9 items in the Usability subscale measure the extent to which the given technology application is perceived to be usable. A high score on this subscale means that the technological characteristics of the application (such as user interface etc.) are in line with the ever-changing demands of modern learners. A low score means the application is perceived as outdated or poorly designed.

The four items in the Feelings toward an application subscale measure the emotional component of the learners’ perception of a given application. A high score on this subscale means that with all the other factors being equal, learners have a strong emotional appeal in using this application. A low score means that with all the other factors being equal, learners do not like using the application.

The three items on the Resistance to implied authorities subscale in relation to one’s teaching and learning subscale measure the extent to which students differentiate between their own learning style, and the institutional regimen and affordances in relation to a particular technology. A high score on this subscale means the individual has a good understanding of his/her own learning style and of how the institution and the technology could simplify achievement of his or her learning goals. A low score on this subscale means either the students’ understanding of their individual learning styles is deficient or the institution and the technology in question have contributed to attainment of the learning goals and therefore resistance to authority is low.

The two items in the Preconceived notions about eLearning subscale measures how high or low learners’ opinions are about eLearning based on their previous experiences. Therefore, a high score on this subscale means that the individual has a positive opinion about eLearning. A low score means that the student’s eLearning experiences were not so positive, thus a degree of mistrust could be present in their opinion on the characteristics of a given technology application.

The two items in the Flexibility of application subscale in relation to the teaching and learning subscale measure how flexible the given application is and how much the application can be adjusted for individual preferences. A high score on this subscale means the application is flexible or is aligned with the individual’s learning and teaching approach. A low score means the application is not adjustable or flexible.

Error experience

The notion of an error experience that serves as a significant factor in usability evaluation was further elaborated on for the next survey iteration. In particular, error experience was operationalised as an if...then... cycle with a score of >4 on 7 items
scale in turn prompting participants to answer an additional question regarding the nature of the error: “What was the nature of the error?” Three response options were developed based on qualitative data from the original survey are then made available to the participants: “a) Technical (files not uploading, browser problems etc.); b) Course design (improper use of calendar, announcements, task/assignments deadlines, not showing a particular week, forum access problems); c) Implementation (Access/login problems)”. This elaboration allows us to distinguish why the experiencing errors were so influential on the usability items that originally were designed to reflect an objective rating of the application. The answers to question # 18 that negatively and strongly loaded onto factor 2 (see Table 1) were related to the error experience for all participant groups (see TELT educational technologies pilots: brief summary of findings doc). Thus, this question is recommended for a separate analysis in the context of an error experiences rather than as a standard component of the survey instrument subscale.

Summary

Contextual factors are thought to be an important aspect of educational technologies evaluation at the higher education level. For the most part university students possess a reasonable level of experience with eLearning along with explicit notions about their individual learning preferences that should be accounted for during the evaluative activities. The refined survey instrument will allow inclusion of these contextual factors within the broader evaluative context.

The next evaluative cycle will assist to identify potential flaws (if any) in the refined version of the survey instrument and perhaps allow for finalisation of the instrument design.

The research significance of this new scale is outlined above in the Theoretical Considerations section. Furthermore, in light of the shortage of valid and reliable instruments for evaluating the effectiveness of educational technologies, the inclusion of contextual factors in the now refined scale may contribute greater value for researchers and evaluators in accomplishing their research/evaluative goals.

The practical significance of this recent iteration of the survey instrument refinement process lies in the effects of market influences that have driven many educational institutions to the point where choosing the right technology for the right purpose is viewed as a time-wasting luxury, and as a result, a “one size fits all” approach prevails. However, access to a valid and reliable evaluative scale could help institutions to optimise the decision-making process in selecting a range of educational technologies appropriate to their needs and circumstances.
References:


