Computers & Education 58 (2012) 1247–1259

Contents lists available at SciVerse ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

# Measuring integration of information and communication technology in education: An item response modeling approach

Jef Peeraer [a,*], Peter Van Petegem [b]

[a] Flemish Association for Development Cooperation and Technical Assistance, Thao Vien guesthouse A3 rm 307, 1B Bac Son - Ngoc Ha, Hanoi, Viet Nam
[b] University of Antwerp, Institute for Education and Information Sciences, Venusstraat 35, 2000 Antwerp, Belgium

ARTICLE INFO

*Article history:*
Received 28 August 2011
Received in revised form
25 November 2011
Accepted 13 December 2011

*Keywords:*
Evaluation methodologies
Improving classroom teaching
Integration of ICT
Pedagogical issues
Teaching/learning strategies

ABSTRACT

This research describes the development and validation of an instrument to measure integration of Information and Communication Technology (ICT) in education. After literature research on definitions of integration of ICT in education, a comparison is made between the classical test theory and the item response modeling approach for the development and validation of a questionnaire. Following the last approach, a construct on integration of ICT is developed, items are generated and an outcome space of Likert type answering categories is defined. The resulting questionnaire has been administered to 933 teacher educators. In this study the collected data are tested for fit to the Rasch model of measurement. It is concluded that the instrument can be used for fundamental measurement of perceived use of ICT for teaching and support of student learning of the reference population, allowing for identification of stages of innovation of ICT integration. We reflect on the critical value of the item response modeling approach and the Rasch measurement model for measurement of integration of ICT in education and discuss some limitations of the study.



## 1. Introduction

Information and Communication Technology (ICT) is high on the education reform agenda of developed and developing countries. Policies for education reform are built around the premise and promise of effective ICT integration (Richards, 2004). Even though much is expected from integration of ICT in education, not much research can be found on measuring effective integration of ICT in teaching practice or on added value of ICT for teaching and learning in general. Studies and measurement tools to investigate the ICT integration level of higher education institutions are scarce (Akbulut, 2009). Cox (2008) recommends to identify how to monitor specific ICT types being used and which data collection methods can provide the most reliable and robust results. Proctor, Watson, and Finger (2003) argue that unless more sophisticated notions of defining ICT curriculum integration are developed, researchers run the risk of promulgating severely restrictive ways of measuring it. In what follows we therefore give an overview of how integration of ICT in education is defined in research literature, after which we reflect on how we can measure it for our own research purposes.

### *1.1. Integration of ICT in education: enhancing, innovating or transforming?*

Plomp, Anderson, Law, and Quale (2003) differentiate between learning about ICT, learning with ICT, and learning through ICT. Zhang (2007) distinguishes between an approach where ICT is seen as the object of education with as purpose to learn about ICT and to get technically skilled, an approach where ICT is used to strengthen expositive teaching, and an approach where is strived for innovative teaching practice, harnessing the full potentials of ICT. Also capability theory refers to the potentials of ICT for educational change and understands ICT as tools to reach an end (Alampay, 2006). Another relevant categorization on use of ICT in education is that of Maddux and Johnson (2005), who differentiate between ICT applications of type I and of type II: Type I applications are those educational applications that simply make it easier, quicker, or more convenient to continue teaching or learning in traditional ways; type II applications are those

* Corresponding author. Tel.: +84 0 909 824 889; fax: +84 4 734 72 90.
*E-mail addresses:* jef.peeraer@gmail.com (J. Peeraer), peter.vanpetegem@ua.ac.be (P. Van Petegem).

0360-1315/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compedu.2011.12.015

educational applications that make available new and better ways of teaching or learning. Others see the potential of ICT not only to innovate teaching practice, but also to change the curriculum. Bull, Bell, and Kajder (In Jamieson-Proctor, Watson, Finger, Grimbeek, & Burnett, 2007) identify two approaches to the use of technology that derive from employing the technology to deliver the existing content more efficiently or alternately to employ the innovation to re-conceptualize aspects of the existing curriculum. Gareis and Hüsing (2009) argue that the transformational potential of ICT is rooted in its effect in terms of empowerment of users, by opening up new, more effective ways for achievement of goals rather than simply making existing structures and processes more efficient.

In education, most agree that the purpose of technology integration is to achieve learning goals and enhance learning – not to use fancy technology tools (Liu and Velasquez in Jamieson-Proctor et al., 2007). It is argued that what counts is not the ICT type but its implementation process (Tubin, 2006). Bowes (2003) argues that effective use of ICT in classroom practice depends on teachers explicitly addressing the question in what way, if at all, the use of ICT can value, given a student learning outcome. ICT ideally supports both teachers' professionalism and students' ability to become independent learners. This means using ICT for enhancing inquiry and data-based decisions, the freedom to make mistakes, the opportunity to work with experts out of school, and assuming responsibility for the outcomes (Tubin, 2006).

In much research on integration of ICT in education, different stages or phases are identified. It has also been suggested to analyze ICT-based innovations on a continuum ranging from the assimilation level through the transition level and up to the transformation level (Mioduser, Nachmias, Tubin, & Forkosh-Baruch, 2003). UNESCO identifies four categories or stages of development concerning ICT use in education: emerging, applying, infusing and transforming (UNESCO, 2005). At the transforming stage of ICT-mediated teaching and learning pedagogies, students' thinking processes are supported by ICT (SEAMO, 2010). The pedagogies adopted by educators at this stage are situated in the constructivist paradigm where learning is perceived as an active construction and reconstruction of knowledge, and teaching is a process of guiding and facilitating students in the process of knowledge construction individually and collaboratively (SEAMO, 2010; Steffe & Gale, 1995). Mills and Tincher (2003) formulated and validated a developmental model for technology integrating, based on stages, standards and indicators of their technology professional development initiative. They organized standards into phases to reflect a development approach "from novice technology facilitators who use technology as a tool for the delivery of instruction to expert technology integrators who are being the technology – augmenting student learning with technology" (Mills & Tincher, 2003).

### *1.2. Measuring integration of ICT in education: an item response modeling approach*

In the context of a capacity building program for teacher educators on ICT integration in education, we aim to measure teacher educators' use of ICT in education over the course of the three-year program. We pursue to assess the use of ICT for teaching and support of student learning, ranging from more traditional to more innovative approaches, reflecting stages of development concerning ICT use in education as described in Section 1.1. The capacity building program involves around 1000 teacher educators, who participate in a panel research throughout the program as well. We opt to develop and validate a self-report questionnaire instrument that can be administered at different points in time of the capacity building program.

Cox (2008) argues that the ways in which ICT has evolved, have influenced the focus and scope of research. A large element of the current research agenda is to measure the uptake of ICT in schools by teachers, pupils, types of computers and so on (Cox, 2008). In the last two decades, researchers have also recognized the need to investigate the effects of ICT on students' generic and specific skills and knowledge, the effects of group and collaborative learning, taking account of human–computer interfaces, the changing nature of knowledge presented and the role of the teacher. The most robust evidence of ICT use enhancing students' learning is from studies which focused on specific uses of ICT, and clearly identifying the range and type of ICT use (Cox and Abbot in Marshall & Cox, 2008). Typically, research conducted within a behaviorist perspective will use quantitative methods and questionnaires, designed to provide evidence at a point in time of program practices, features and outcomes (Marshall & Cox, 2008). Christensen and Knezek (2008) argue that competencies, defined in terms of behaviors, could be reasonably assessed by observation as well as by self-report.

Different questionnaire instruments to measure the use of ICT for teaching and learning have been developed and tested (Christensen & Knezek, 2008) following principles of the classical test theory. Validation in classical test theory mostly focuses on models at the test-score level and links test scores to true scores. On the other hand, both person parameters (i.e., true scores) and item parameters (i.e., item difficulty and item discrimination) are dependent on the test and the respondent sample, respectively, and these dependencies can limit the utility of the person and item statistics in practical test development work and complicate any analyses (Hambleton & Jones, 1993). Advantages of classical test theory models are that they are based on relatively weak assumptions (i.e., they are easy to meet in real test data) and they are well-known and have a long track record. Modern test theories are considered superior to the classical test theory as it makes stronger assumptions and provides stronger findings. A good test theory or model like item response theory models can provide a frame of reference for doing test design work. A good test model might specify the precise relationships among test items and ability scores (Hambleton & Jones, 1993). In scale development, both the traditional statistics and item response theory models like the Rasch model can enhance the measurement capacity of a scale (Cavanagh, Romanoski, Giddings, Harris, & Dellar, 2003a).

To validate a self-report questionnaire instrument and to measure integration of ICT in teaching and learning, allowing for identification of stages of innovation of ICT integration, we follow in this study the Rasch measurement model and methodology. The Rasch measurement model and methodology is a modern test theory that involves rigorous and extensive analysis of the data and provides additional psychometric information that cannot be obtained through the classical test theory approach. Items on use of ICT have so far rarely been ordered from "easy" to "hard" by calibration against the distribution of educators' perception on ICT use. Usually questionnaire scales on use of ICT for teaching and support of student learning are not constructed with the items being selected to fit a measurement model and form a one-dimensional scale in which the items can be said to be affected by one dominant trait. Classical test theory tries to have all items of "similar difficulty" and does not have a conceptual measurement design in the preparation of the items (Cavanagh et al., 2003a; Cavanagh, Romanoski, Giddings, Harris, & Dellar, 2003b). Following an item response modeling approach, data are tested for fit into the Rasch model, which allows for a detailed examination of the internal construct validity of the scale, including properties such as reliability and ordering of categories. When data fit the Rasch model, requirements of 'fundamental measurements', as defined by Bond and Fox (2007) are met:

measurements allow an order of ranking, calculations as adding up and subtracting are possible, and calibration of the items is independent of the respondents and vice versa (objectivity).

## 2. Development of an instrument measuring the use of ICT for teaching and support of student learning

In what follows we describe the development of an instrument to measure integration of ICT in teacher education, tested in the context of Vietnam. The development was carried out before the start of the capacity building program and took place in different building blocks, as prescribed by Wilson (2005). In the first building block a comprehensive literature study on integration of ICT in education informed the development of a construct map on integration of ICT in teaching and learning. This led in the second building block to the item design and in the third building block to the development of an outcome space. The result is a self-report questionnaire scale on teacher educators' use of ICT for teaching and support of student learning.

The instrument has been integrated in a questionnaire addressing different aspects of integration of ICT in teacher education. This questionnaire is being used for longitudinal panel research on the effectiveness and impact of the capacity building program on integration of ICT in education for teacher educators. In the course of the capacity building program, data on the questionnaire has been collected at different points in time.

### *2.1. Construct map on integration of ICT in teaching and learning*

A construct map is a visual representation of a continuum, manifested in an ordering of the respondents and/or an ordering of item responses (Wilson, 2005). For the intended use of the instrument, the measurer wants to array the respondents from high to low, or left to right, in some context. A construct is always an ideal and used to suit a theoretical approach. Therefore it is only sensible to use a construct map if the theoretical approach is consistent with the idea of mapping a construct (Wilson, 2005).

We followed the "Technology Mapping" methodology as developed by Angeli and Valanides (2009), which is a situated methodology where educators establish connections among the affordances of a tool (software), content and pedagogy. Based on the review of the literature on integration of ICT in education (see Section 1.1), we listed a series of ICT applications and their educational purposes. In discussion with a group of Vietnamese educators and experts in the field of ICT integration, we ordered them from less innovative to more innovative, reflecting stages of development concerning ICT use in education. This led to a construct map representing teacher educators on a continuum from use of ICT to replace and enhance teaching and learning through use of ICT to innovate teaching and learning and up to the use of ICT to transform teaching and learning (see Fig. 1).

***Increasing innovativeness of use of ICT for teaching and learning***

| **Educators** | **Use of ICT applications …** |
|---|---|
| Educators who use ICT to **transform** teaching and learning practice | Classroom management systems for coaching and evaluation, web search assignments for problem based learning;<br>Electronic communication with students and students communicating with others |
| Educators who use ICT to **innovate** teaching towards more student centred learning, integrated into existing curriculum | Integration into subject teaching (Word processing for shared writing exercises, simulations and data processing by students);<br>Student presentations and students creating multi-media products, students constructing and synthesizing knowledge |
| Educators who use ICT to **enhance** teaching practice | Presentation software for enhancing, multi-media presentations<br>Accessing offline and online databases and information to prepare resources<br>Electronic practice and drill exercises for revision |
| Educators who use ICT to **replace** traditional teaching practice | Word processing for production of documents (preparation of lesson plans, handouts, slides, …)<br>Presentation software for lecturing |

***Decreasing innovativeness of use of ICT for teaching and learning***

**Fig. 1.** Construct map integration of ICT in teaching and learning.

### 2.2. Item design

For this research, we aimed to develop an instrument to measure the use of ICT in teaching and learning practice in education, tested in the context of teacher education in Vietnam. We used behavioral frequency as an indicator of proficiency. As argued in Section 1.2 we opted for a self-report questionnaire. Items have to stimulate responses that can constitute observations about the construct that the measurer wishes to measure. Typically the item is but one of many that could be used to measure the construct (Wilson, 2005). To select a finite set of items for the self-report questionnaire that represents the construct, we started from a review of the research literature on measurement of integration of ICT. ICT resources are necessarily employed within pre-existing contexts of educational and social activity, and the outcome in terms of both pattern of use and learning depends on how they fit with these (Tolmie, 2001). Therefore we especially looked for literature on ICT resources in the context of emerging developing countries (UNESCO, 2003; Wagner et al., 2005). Concerning use of ICT, UNESCO Bangkok (2003) differentiates indicators concerning teaching professionals use and teaching, and student use and learning. We opted to develop a scale consisting of two sets: educators' use of ICT for teaching purposes (see Section 2.2.1) and for support of student learning (see Section 2.2.2). Both sets are mutually influential and collectively comprise use of ICT for teaching and support of student learning.

#### 2.2.1. Educators' use of ICT for teaching purposes

A series of tools can be used by teaching professionals for different educational purposes. The SITES M2 Case Studies, provide a detailed picture of ICT that is being used to support innovative teaching practices around the globe (Quellmalz & Kozma, 2003): productivity tools, web resources, e-mail, multi-media software, web design tools, specialized educational software; to create products and presentations, to search for information, to support communication, to organize instruction, to monitor and assess student work, to support student collaboration. In an overview of teaching professionals' use of computers for teaching purposes, UNESCO Bangkok lists the use of computers for lesson preparation, teaching, reporting, recording, communicating, and assessing (UNESCO, 2003). Van Braak, Tondeur, and Valcke (2004) distinguish between use of computers for support in educational practice and integration of computers as a teaching tool or learning device. Ward and Parr (2010) identify five usage factors or types of use for educators: generic pedagogical use, specific pedagogical use, preparation and presentation of lesson material, core professional work and personal use. In a review of research literature, Pilkington (2008) describes that as teachers begin to use IT they may do so at first in ways that reinforce traditional practice. Later they may use IT to make modest enhancements, e.g. exploiting properties of multi-media to improve resources in ways that impact on the understanding of concepts. Later, as they continue to integrate their subject knowledge with the use of IT, they may include more subject specific software to improve and extend the curriculum. As they continue to evolve their practice, perhaps adopting more social and collaborative ways of working, they may also use ICT for extended collaborative projects (Pilkington, 2008).

For our research and in this set, we focused on items representing educators' use of ICT for preparation and presentation of lesson material and for pedagogical use (Table 1). The final set of items represents a selection of tools which are commonly used in education in Vietnam and which are typically used for a range of teaching purposes. Exemplary software which is commonly known and used in teacher education in Vietnam was added to the items to clarify what kinds of tools are meant.

**Table 1**
Questionnaire items on educators' use of ICT for teaching.

| |
|---|
| T.1 I use word processing for production of documents (e.g. overhead transparencies and handouts). |
| T.2 I use communication and presentation software for lecturing (e.g. PowerPoint, Mind-mapping software, Digital picture story software, …). |
| T.3 I use subject specific software for integration into lectures (e.g. Graphics software, Statistical packages, Simulation software, …). |
| T.4 I use information accessed through CD-ROM/DVD as resource materials for lecturing. |
| T.5 I use electronic communication with students (E-mail, World Wide Web (WWW), ...). |
| T.6 I use internet/WWW as a source of information for lecturing. |
| T.7 I use classroom management software in a computer classroom setting (e.g. NetOp). |

The different items gradually address more innovative use of ICT for teaching practice. Item T.1, T.4 and T.6 represent the use of ICT to replace and/or enhance traditional teaching practice as these tools make it more easy to prepare teaching practice (type I applications). Item T.2 and especially item T.3 represent an innovative approach to teaching: the presented tools make available new and better ways of teaching or learning (type II applications). Item T.5 represents transformative use of ICT: electronic communication with students potentially brings education outside the classroom. Item T.7 represents a re-conceptualization of student assessment as it allows for real time monitoring and evaluation of student work.

#### 2.2.2. Students' use of ICT for learning purposes

Twenty-first century assessments should incorporate the explicit examination of technologies in supporting, extending, and transforming student learning (Quellmalz & Kozma, 2003; Wagner et al., 2005). As an indicator on student use of ICT, UNESCO (2003) proposes the frequency of use of ICT for the following purposes: informative (acquisition and use of information), functional (use and manipulation of existing information), creating (composition, compilation and production of new information), and communication (exchange of information). Quellmalz and Kozma_(2003) introduce the concept of "ICT strategies", used by learners to integrate technology into problem solving activities. The ICT strategies include strategies to taking advantage of the capabilities of technologies (Quellmalz & Kozma, 2003), e.g. to "access and organise information and relevant data", "represent and transform data and information", "analyse and interpret information and data", "communicate ideas, findings and outputs". The focus is on generalizable ICT strategies, rather than discrete, often changing features of technology tools. Jamieson-Proctor et al. (2007) developed a measurement instrument which contains two theoretically underpinned factors. The instrument measures student use of ICT, as perceived by the teacher. The items comprising the first factor define ICT as a tool for the development of ICT-related skills and the enhancement of curriculum learning outcomes (integration). The items comprising the second factor define ICT as an integral component of reforms that change what students learn and how school is structured and organized. As Jamieson-Proctor et al. (2007) conclude, the instrument measures both curriculum enhancement and transformational dimensions in relation to ICT use by students. Drent and Meelissen (2008) argue that the use of ICT can only be regarded as “innovative” if the ICT application facilitates the student-centered learning, and there is a variation in ICT use (Drent, Voogt and Odenthal in Drent & Meelissen, 2008).

For our research we adopted a set of items (Table 2) representing educators’ perceived use of ICT for support of student learning. The selection of items focused on innovative use of ICT by students as defined in Drent and Meelissen (2008), addressing generalizable ICT strategies (Quellmalz & Kozma, 2003) and different educational purposes (UNESCO, 2003): orientation, acquisition and use of information (item L.1 and L.2); use and manipulation of existing information (item L.3); problem solving (item L.4), composition, compilation and production of new information (item L.5, L.6 and L.7); and exchange of information (item L.8). Following the argumentation of Jamieson-Proctor et al. (2007), most of these items define ICTs as a tool for the development of ICT-related skills and the enhancement of learning outcomes, suggesting the use of ICTs to support and improve teaching and learning. Item L.8 furthermore defines ICTs as an integral component of reforms that change what students learn and how education is structured and organized, implying a transformative ICT function.

**Table 2**
Questionnaire items on educators’ use of ICT for support of student learning.

| |
|---|
| L.1 I let students work with the computer to orientate themselves to a new subject. |
| L.2 I let students gather information from electronic databases. |
| L.3 I let students use the computer to process collected data. |
| L.4 I let students work with a computer program in which a problem is given that they have to solve, supported by the computer. |
| L.5 I give students an assignment to give a presentation supported by a computer. |
| L.6 I let students integrate different media to create appropriate products. |
| L.7 I let students use the computer to synthesize their knowledge. |
| L.8 I let students use the computer to communicate with others (locally and/or globally). |

### *2.3. Outcome space*

An outcome space is a set of qualitatively described categories for recording and/or judging how respondents have responded to items. An outcome space is a set of categories that are well defined, finite and exhaustive, ordered, context-specific, and research based (Wilson, 2005). Scoring of answering categories relates the outcome space back to the construct map. Likert type answering categories in questionnaires are usually scored according to the number of response categories allowed (Wilson, 2005).

For the proposed indicators on use of computers for teaching purposes, UNESCO provides the following answering categories: “rarely”, “sometimes”, “regularly”, “always” (UNESCO, 2003). Ward and Parr (2010) defined answering categories starting from “not at all”, followed by “rarely”, “sometimes” and “often”. Others (Van Braak et al., 2004) work with more definite categories where respondents indicate to what extent they use the computer for various tasks: “never”, “every term”, “monthly”, “weekly”, “on a daily basis”. As suggested by Proctor et al. (2003) a four-point ordinal response scale to gauge frequency of use ensures a recognizable separation for respondents between the frequencies from “never” to “regularly” or “very often”. A four-point scale avoids the selection of a default median frequency. For our instrument we opted for a generic approach, going from “never” (0), over “rarely” (1) and “sometimes” (2) to “regularly” (3), which, as argued by Proctor et al. (2003), enables responses from educators across a span of different teaching grades, and for educators of all curriculum areas.

## 3. Research objective: validation of the measurement instrument

In Section 2 of this research paper, we reported on the development of the self-report questionnaire. In what follows, we aim to validate the developed measurement instrument, following the principles of the Rasch measurement model and methodology.

Our research aims to apply a scale development and validation process that can:

1. Produce a scale to measure teacher educators’ self-reported use of ICT for teaching and support of student learning;
2. Produce a scale with item difficulties and measures of use of ICT for teaching and support of student learning calibrated on the same scale; and
3. Produce a scale that elicits data to fit the theoretical model on the use of ICT to enhance, innovate and/or transform education, reflecting stages of development concerning ICT use in education.

## 4. Methodology

### *4.1. Data collection*

Data collection for this validation study took place in the beginning of 2010. The questionnaire was presented to all teacher educators working in five teacher education institutions participating in the capacity building program. The five provincial institutions are from different regions in the north and center of Vietnam and were selected by the Ministry of Education and Training of Vietnam for participation in the program. On a total population of 1021 teacher educators, 933 completed the questionnaire (91%). This data is now evaluated for fit to the Rasch measurement model.

### 4.2. Rasch measurement model

After a traditional principal components factor analysis to analyze the underlying structure of the items (Cohen, Lawrence, & Morrison, 2007), we apply Rasch statistics to further validate the scale. The Rasch model is a measurement model expressed at the item level and the instrument level, focusing on modeling probability of the observed responses. In the Rasch model the form of the relationship is that the probability of the item response for item $i$, $X_i$, is modeled as a function of the respondent location $\theta$ and the item location $\delta_i$. The logic of the Rasch model is that the respondent has a certain amount of the construct, indicated by the arbitrary symbol $\theta$, and an item also has a certain amount, indicated by the arbitrary symbol $\delta_i$. The way the amounts work is in opposite directions – hence, the difference between them is what counts. The 'distance' between the person and item location determines the probability. An item response function describes how a respondent responds to an item. Related to the construct map, the difference between a respondent's location and the item location will govern the probability that the respondent will make that response (Wilson, 2005).

The following analyses were performed on the collected data:

1) Constructing the Wright map
2) Evaluation of the measurement model: Item fit
3) Reliability: measurement error
4) Validity: item consistency with instrument.

#### 4.2.1. Wright map

A Wright map combines the idea of the construct map with the Rasch model. A central scale in logits determines the relationship of the construct to the probability of response. The logit is the log of the "odds" of an event (the proportion of times that event occurs compared to the times it does not occur). Locations of respondents are indicated by an "X" on the left-hand side, in the shape of an on-the-side histogram. Item locations are indicated on the right-hand side. Items are shown on the construct map at the point where the probability of choosing a response of "1" is 0.5. When the person and item locations are the same, the probability of answering 'correct' is 0.5. As the person location moves above, the probability increases above 0.5; as the person location moves below, the probability decreases below 0.5. Wilson (2005) provides a convenient table (see Table 3) of Logit differences between item and person locations and probabilities for the Rasch model which allow for a rough estimation.

**Table 3**
Logit differences and probabilities for the Rasch model (Wilson, 2005).

| $\theta - \delta$ | Probability |
|---|---|
| −4.0 | 0.02 |
| −3.0 | 0.05 |
| −2.0 | 0.12 |
| −1.0 | 0.27 |
| 0.0 | 0.50 |
| 1.0 | 0.73 |
| 2.0 | 0.88 |
| 3.0 | 0.95 |
| 4.0 | 0.98 |

When there are more than two score categories (polytomous data), the Rasch model can be generalized. Step parameters $\delta_{ik}$ govern the probability of making the "step" from score $k - 1$ to score $k$ (Wright & Masters in Wilson, 2005). Thurstone thresholds are mapped onto a Wright map, where the $k$th Thurstone threshold can be interpreted as the point at which the probability of the scores below $k$ is equal to the probability of the scores $k$ and above (and that probability is 0.5). In other words, a Thurstone threshold is the point on the variable (in the context of a particular item) at which the probability of being observed below a given category is equal to that of being observed in or above that category. The Thurstone thresholds in general are not the item parameters, but the relative locations can be used for interpretive purposes.

Statistical estimation approaches are used to estimate item and person locations. The software "ConstructMap version 4.5.0" is used to calculate the estimations in this research (Wilson, 2005).

#### 4.2.2. Evaluation of the measurement model: item fit

Item locations are estimates and this is subject to a degree of uncertainty. Item locations have a standard error. This item standard error is required to calculate the item fit statistic. With respect to items, investigation of fit emphasizes on consideration of how well the shapes of the empirical item characteristic curves are captured by the curves generated by the estimated item parameters. The mean square fit statistic (Wright & Masters in Wilson, 2005) compares how much the actual residuals vary from how much we would expect them to vary if the data fit the model. Adams & Khoo (in Wilson, 2005) have indicated that 0.75 (=3/4) is a reasonable lower bound and 1.33 (=4/3) is a reasonable upper bound. Marginal Maximum Likelihood (MML) estimation method is used to supply the weighted values for the mean square statistics.

#### 4.2.3. Reliability: measurement error

In creating a construct and realizing it through an instrument, the measurer has assumed that each respondent who might be measured has some amount of that construct and the amount is sufficiently measurable to be useful (Wilson, 2005). Each location is an estimate, subject to a degree of uncertainty. This uncertainty is usually characterized using the standard error of the location – the so-called standard error of measurement. The standard error of measurement ($sem(\theta)$) is an index of measurement error for respondents which tells the measurer how accurate each estimate is. The standard error of measurement varies depending on the respondent's location. The relationship is typically a "U" shape, with the minimum near the mean of the item thresholds and the value increasing toward the extremes. The closer the respondent is to an item, the more the item can contribute to the estimation of the respondent's location. The formula for the standard error of measurement is given in Adams, Wilson, and Wang (1997).

#### 4.2.4. Validity: item consistency with instrument

Construct validity issues are built into the item design as well as into the construct map. One requirement is that the items are consistent with the instrument as a whole. Respondents higher on the construct would, in general, also score higher on each item. In terms of the Wright map, one could consider the locations of the respondents within each score group for a given item: if the mean location of each group tends to increase as the scores increase, it seems reasonable to say that this particular expectation that comes from the item design has been fulfilled.

## 5. Results

### 5.1. A Wright map on use of ICT for teaching and support of student learning

Factor analysis on the items of both sets (Extraction method: PCA) retains two factors. Nevertheless, all items load higher on the first retained factor with factor loadings from 0.498 to 0.848. For our research on integration of ICT in teacher education, we combine the item sets on teacher educators' perceived use of ICT for teaching and support of student learning. In theory, both sets are mutually influential and collectively comprise a single, one-dimensional scale. Semantically, the items for each set were written to elicit data on a similar construct. Separation reliability $r$ is calculated and can be interpreted as a conventional reliability coefficient, although it is calculated in the metric of the respondent locations rather than in the traditional score metric (Wilson, 2005). Proportion of variance accounted for by the estimator of a respondent's location is used as a basis for calculating the separation reliability, $r$. The reliability coefficient is 0.93 (Cronbach's Alpha).

In our research the single trait we aim to measure is teacher educators' use of ICT for teaching and support of student learning. The combined set is mapped on a Wright Map (see Fig. 2).

```
Lo-  Raw                                                  Item responses
git  score  Respondents         0 vs. 1, 2 & 3            0 & 1 vs. 2 & 3               0, 1 & 2 vs. 3
------------------------------------------------------------------------------------------------------------
3    |                         |                                                        T.7.3
     |                         |
     |                         |
     |45          X            |                                                        L.1.3
     |                         |
2    |            X            |                                                        L.3.3  L.7.3
     |            XX           |                          T.7.2                         T.3.3  L.4.3  L.6.3
     |42        XXXXX          |                                                        T.5.3  L.5.3
     |           XXX           |                                                        T.4.3  L.2.3  L.8.3
     |            XX           |                                                        T.2.3
1    |        XXXXXXXXX        |
     |       XXXXXXXXXXX       |
     |      XXXXXXXXXXXXX      | T.7.1                    L.1.2                         T.1.3
     |   XXXXXXXXXXXXXXXXXX    |                                                        T.6.3
     |       XXXXXXXXXXXX      |                          L.4.2  L.8.2
0    |     XXXXXXXXXXXXXXXX    |                          T.3.2  L.3.2  L.5.2  L.7.2
     |      XXXXXXXXXXXX       |                          T.5.2  L.6.2
     |      XXXXXXXXXXXX       |                          L.2.2
     |        XXXXXXXX         | L.1.1                    T.4.2
-1   |          XXXX           | L.4.1  L.8.1             T.2.2
     |           XX            | T.5.1  L.3.1  L.5.1  L.6.1  L.7.1   T.1.2
     |          XXXX           | T.3.1
     |3          XX            | L.2.1                    T.6.2
     |2          X             |
-2   |                         | T.2.1
     |1          X             |
     |                         | T.1.1  T.4.1
     |0          XX            | T.6.1
     |                         |
-3   |                         |
============================================================================================================
     Each X represents 7 respondents, each row is 0.213 logits
```

**Fig. 2.** Wright Map on use of ICT for teaching and support of student learning.

The respondents range from those that are "less innovative" at the bottom to those that are "more innovative" at the top. Each location (i.e., each histogram bar) corresponds to a score on the instrument, ranging from 0 to 45. The right-hand side of the map shows the Thurstone thresholds. Each item of the instrument has three Thurstone thresholds: one between each of three ordered response categories. The first threshold for each item, governing the transition from "never" to "rarely", "sometimes" or "regularly", is shown in the first column on the right-hand side, and so forth. The first two digits are the item number and the third digit after the decimal point is the threshold.

The relative locations of the items can be examined to review how well the Wright map reflects the construct map. The thresholds should indicate a monotonic (one direction) response process (Hagquist & Andrich, 2004). The Rasch model tests that Likert type categories are answered consistently and logically; that is, that they provide an ordered set of responses from 'easy' to 'hard' (Cavanagh et al., 2003a).

Discrimination for the item thresholds of our questionnaire is satisfactory as thresholds are ordered in correspondence with the ordering of the response categories. The relative location of the Thurstone thresholds shows that the Wright map is a good reflection of the construct map (see Fig. 1), especially for the items on the use of ICT for teaching purposes (T-items). In each of the three columns, item thresholds on the use of ICT for teaching purposes reflect the construct map. The use of word processing (item T.1) for preparation of lesson plans, but also the Internet as source of information for lecturing (item T.6), is mapped as less innovative, replacing traditional teaching practice. Items T.2 and T.4 are representing the use of ICT to enhance teaching practice, while item T.3 and T.5, represent the use of ICT to innovate teaching and learning. Item T.7, of which item thresholds are consistently on the top of the Wright Map, could represent the use of ICT to transform teaching and learning. Thurstone thresholds of items on the use of ICT to support student learning (L-items) are concentrated around the level of educators' innovative use of ICT (thresholds of T.3 and T.5). This is consistent with the item design as only items were selected which represent innovative use of ICT, facilitating student-centered learning, as suggested by Drent and Meelissen (2008). Specifications related to the L-items in the construct map are less exact: most item thresholds are all located around the threshold level of innovative use of ICT. Therefore an exact comparison for these items is limited. Only item thresholds for item L.1 are consistently higher up in the Wright Map, which implies that the item is a more 'difficult' one to agree with.

Logit differences can be translated into probability statements about expectations for respondents giving a certain response. When relating T-items to L-items, it is clear from the Wright Map that when respondents regularly integrate specific software into lectures (item threshold T.3.3) or use applications to communicate with students (item threshold T.5.3, location around 1.5 Logits) they will more probably let their students use ICT in a student-centered fashion from time to time (item threshold L3.1, L5.1, L6.1 and L7.1, location around −1.2 Logits) than that they would never let them do that (Logit difference of around 3 Logits, Probability around 0.95). Even the lowest scoring respondents have about 50% probability of responding above "never" to "Using the internet/www as a source of information for lecturing". But the same respondents have a small probability of responding the same way on "Letting students work with the computer to orientate them to the subject" (about 12%), and even less to "Using classroom management software" (less than 5%).

Respondent distribution is somehow bell shaped, with a lower tail and an upper tail, but with most respondents located between −2 and 2 Logits. Respondents at a location of 0 Logits have a probability of 0.50 to integrate subject specific software in their teaching practice (item threshold T.3.2), and to let students use a computer to collect data, to give a presentation or to synthesize their knowledge (item thresholds L.3.2, L.5.2, L.7.2).

Linacre (2002) states that step difficulties have to advance with minimum 1.4 logits and maximum 5.0 logits. For some items, the difference is less than 1.4 logits in between the first and the second threshold. When the advance is less than 1.4 logits, redefining the categories to have wider substantive meaning or combining categories may be indicated (Linacre, 2002).

### 5.2. Evaluation of the measurement model: item fit

The weighted mean square (see Fig. 3) indicates that all but one of the items (L.7) fit within reasonable bounds. The overall finding is that the data fit the model reasonably well. Item L.7 does not seem to be a crucial item: it is not particularly different from the content of other L-items, and it has item threshold locations similar as item L.3. The factor analysis on all items of the combined scale shows for item L.7 a factor loading of 0.830 on the dominant factor. Therefore, it could be considered to delete the item, as the construct will not lose any of its definition.

```
Weighted Mean Square
============================================================================
                  0.58 0.67 0.75 0.83 0.92 1.00 1.08 1.17 1.25 1.33 1.42 1.5
--------------------+----+----+----+----+----+----+----+----+----+----+----+
T.1                           .              |                   *
T.2                           .              |               *   .
T.3                           .              |               *   .
T.4                           .              |                *  .
T.5                           .              *                   .
T.6                           .              |       *           .
T.7                           .              |        *          .
L.1                           .      *       |                   .
L.2                           .  *           |                   .
L.3                           . *            |                   .
L.4                           . *            |                   .
L.5                           .       *      |                   .
L.6                           .   *          |                   .
L.7               <*          .              |                   .
L.8                           .              *                   .
============================================================================
```

**Fig. 3.** Item fit statistics.

### 5.3. Reliability: measurement error

For the measurement instrument on the use of ICT for teaching and support of student learning, the respondents in the middle will always have more items near them than those at the extremes, hence the ($sem(\theta)$) will be smaller in the middle than at the end.

The graph (see Fig. 4) shows that for our instrument, the most sensitive part is from approximately −2.0 to +2.0 logits. Looking back at the Wright map (Fig. 1), this corresponds to approximately the range of most thresholds for all the items expect for the first threshold of

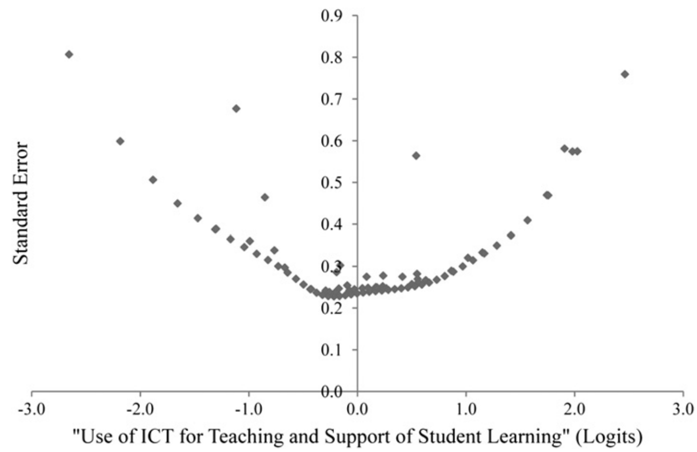


**Fig. 4.** The standard error of measurement (each dot represents a score).

three items (i.e., T1.1, T4.1, T6.1) and the last threshold of two items (i.e., L1.3, T7.3). Thus, the instrument's range of maximum sensitivity makes general sense with respect to the item response categories. The distribution of the respondents shows moreover that almost all respondents are located between −2.0 and +2.0 logits.

### 5.4. Validity: item consistency with instrument

In Table 4, the mean location of each score group is shown for all the items of the instrument. In every case, the mean increases as the score rises.

**Table 4**
Item statistics.

| Statistics | Response categories | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | 3 |
| Item T.1 | | | | |
| Count | 68 | 143 | 342 | 363 |
| Percent (%) | 7.42 | 15.61 | 37.34 | 39.63 |
| Pt-Biserial | −0.28 | −0.26 | −0.15 | 0.49 |
| Mean location | −0.82 | −0.37 | −0.03 | 0.58 |
| Std. Dev. of locations | 0.37 | 0.28 | 0.26 | 0.30 |
| Item T.2 | | | | |
| Count | 104 | 173 | 367 | 273 |
| Percent (%) | 11.34 | 18.87 | 40.02 | 29.77 |
| Pt-Biserial | −0.44 | −0.30 | 0.04 | 0.52 |
| Mean location | −0.98 | −0.34 | 0.13 | 0.74 |
| Std. Dev. of locations | 0.37 | 0.27 | 0.25 | 0.31 |
| Item T.3 | | | | |
| Count | 186 | 251 | 312 | 167 |
| Percent (%) | 20.31 | 27.40 | 34.06 | 18.23 |
| Pt-Biserial | −0.51 | −0.20 | 0.21 | 0.50 |
| Mean location | −0.75 | −0.11 | 0.30 | 0.95 |
| Std. Dev. of locations | 0.34 | 0.25 | 0.26 | 0.33 |
| Item T.4 | | | | |
| Count | 89 | 231 | 369 | 228 |
| Percent (%) | 9.71 | 25.19 | 40.24 | 24.86 |
| Pt-Biserial | −0.44 | −0.23 | −0.00 | 0.54 |
| Mean location | −1.08 | −0.18 | 0.10 | 0.85 |
| Std. Dev. of locations | 0.39 | 0.27 | 0.26 | 0.32 |
| Item T.5 | | | | |
| Count | 209 | 188 | 316 | 205 |
| Percent (%) | 22.77 | 20.48 | 34.42 | 22.33 |
| Pt-Biserial | −0.55 | −0.20 | 0.19 | 0.53 |
| Mean location | −0.70 | −0.18 | 0.27 | 0.89 |
| Std. Dev. of locations | 0.32 | 0.26 | 0.26 | 0.32 |
| Item T.6 | | | | |
| Count | 54 | 96 | 356 | 409 |
| Percent (%) | 5.90 | 10.49 | 38.91 | 44.70 |
| Pt-Biserial | −0.38 | −0.27 | −0.20 | 0.54 |
| Mean location | −1.35 | −0.47 | −0.07 | 0.57 |
| Std. Dev. of locations | 0.44 | 0.29 | 0.26 | 0.29 |

(*continued on next page*)

**Table 4** (*continued*)

| Statistics | Response categories | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | 3 |
| Item T.7 | | | | |
| Count | 524 | 231 | 109 | 49 |
| Percent (%) | 57.39 | 25.30 | 11.94 | 5.37 |
| Pt-Biserial | −0.56 | 0.24 | 0.32 | 0.30 |
| Mean location | −0.27 | 0.41 | 0.74 | 1.18 |
| Std. Dev. of locations | 0.28 | 0.27 | 0.29 | 0.39 |
| Item L.1 | | | | |
| Count | 279 | 251 | 287 | 100 |
| Percent (%) | 30.43 | 27.37 | 31.30 | 10.91 |
| Pt-Biserial | −0.69 | −0.06 | 0.44 | 0.43 |
| Mean location | −0.70 | 0.04 | 0.57 | 1.12 |
| Std. Dev. of locations | 0.31 | 0.25 | 0.27 | 0.36 |
| Item L.2 | | | | |
| Count | 144 | 176 | 377 | 218 |
| Percent (%) | 15.74 | 19.23 | 41.20 | 23.83 |
| Pt-Biserial | −0.59 | −0.28 | 0.18 | 0.56 |
| Mean location | −1.04 | −0.26 | 0.24 | 0.89 |
| Std. Dev. of locations | 0.37 | 0.25 | 0.25 | 0.32 |
| Item L.3 | | | | |
| Count | 219 | 182 | 351 | 161 |
| Percent (%) | 23.99 | 19.93 | 28.44 | 17.63 |
| Pt-Biserial | −0.67 | −0.18 | 0.32 | 0.54 |
| Mean location | −0.84 | −0.13 | 0.36 | 1.06 |
| Std. Dev. of locations | 0.33 | 0.25 | 0.25 | 0.35 |
| Item L.4 | | | | |
| Count | 259 | 212 | 271 | 170 |
| Percent (%) | 28.40 | 23.25 | 29.71 | 18.64 |
| Pt-Biserial | −0.69 | −0.14 | 0.33 | 0.57 |
| Mean location | −0.75 | −0.06 | 0.44 | 1.05 |
| Std. Dev. of locations | 0.32 | 0.24 | 0.26 | 0.34 |
| Item L.5 | | | | |
| Count | 211 | 195 | 324 | 187 |
| Percent (%) | 23.01 | 21.26 | 35.33 | 20.39 |
| Pt-Biserial | −0.66 | −0.18 | 0.24 | 0.58 |
| Mean location | −0.85 | −0.11 | 0.32 | 1.02 |
| Std. Dev. of locations | 0.33 | 0.24 | 0.25 | 0.34 |
| Item L.6 | | | | |
| Count | 203 | 185 | 342 | 185 |
| Percent (%) | 22.19 | 20.22 | 37.38 | 20.22 |
| Pt-Biserial | −0.61 | −0.18 | 0.28 | 0.47 |
| Mean location | −0.80 | −0.14 | 0.34 | 0.86 |
| Std. Dev. of locations | 0.33 | 0.26 | 0.26 | 0.32 |
| Item L.7 | | | | |
| Count | 206 | 199 | 349 | 161 |
| Percent (%) | 22.51 | 21.75 | 38.14 | 17.60 |
| Pt-Biserial | −0.65 | −0.21 | 0.31 | 0.54 |
| Mean location | −0.85 | −0.16 | 0.36 | 1.05 |
| Std. Dev. of locations | 0.33 | 0.25 | 0.26 | 0.34 |
| Item L.8 | | | | |
| Count | 250 | 203 | 257 | 202 |
| Percent (%) | 27.41 | 22.26 | 28.18 | 22.15 |
| Pt-Biserial | −0.64 | −0.14 | 0.25 | 0.55 |
| Mean location | −0.71 | −0.08 | 0.36 | 0.93 |
| Std. Dev. of locations | 0.32 | 0.25 | 0.26 | 0.33 |

## 6. Conclusions

Testing data for fit to the Rasch model is rarely done so far for questionnaire scales on use of ICT for teaching and support of student learning. Complementing classical test theory, the item response modeling approach can add value to the measurement capacity of a scale. It can enhance the development of measurement scales in the field of integration of ICT in education. Most theoretical models for technology integration are based on stages, standards and indicators. Analysis of ICT-based innovation is taking place on a continuum ranging from "lower" levels of assimilation to "higher" levels of transformation. Researchers are endeavoring to measure these stages of integration. The Partnership on Measuring ICT for Development (2010) states that to reap the benefits of the rapidly changing information society, governments need to monitor and benchmark progress in ICT indicators with a view to designing and reviewing national policies and strategies. The item response modeling approach is a valuable approach to develop measurement instruments which allow for measurement of use of ICT for teaching and learning. In this study we defined the construct of ICT integration in education, reflecting these stages of development concerning ICT use for teaching and support of student learning. Following the item response modeling approach, we presented the construct in a construct map. We described two sets of items that make up a self-report measurement instrument and we described a strategy for coding responses on a questionnaire into an outcome space. After data collection from 933 Vietnamese teacher educators, we used ConstructMap 4.5.0 to calibrate the construct map in a Wright map. A factor analysis showed that even though two factors are retained, there is a dominant factor on which all items load most. On the Wright map, item difficulties and measures of use of ICT for teaching and support of student learning are calibrated on the same scale. The scale has a reliability coefficient (Cronbach's alpha) of 0.93. With a Rasch model analysis we tested whether items on teacher educators' use of ICT for teaching and support of student learning link together on the same scale. Investigation of item fit showed that the data fit the Rasch model reasonably well. Only item L.7 does not fit within reasonable bounds. We suggest to delete the item for future use of the measurement instrument as the construct will not lose any of its definition without it. Item parameters cover the region of the scale where the respondents lay. Generally, the instrument is better when the item locations span the full range of the respondents (Wilson, 2005). Large gaps in the coverage indicate areas where the respondents are being measured less well than elsewhere. As shown in the Wright map, the item thresholds do indeed span the range of the respondents fairly well. For some items, step difficulties of the first and second threshold advance with less than 1.4 logits. For future use of the measurement instrument, combining the answering categories "Rarely" and "Sometimes" can be suggested.

The main objective of the instrument is to measure the use of ICT by the respondents enabling statements about the level of integration of ICT in education. In 2010 the Southeast Asian Ministers of Education Organization (SEAMO) published a report on the status of ICT integration in Education in eleven member countries. SEAMO stated that in Vietnam some schools may be able to transform their ICT-mediated teaching and learning practices. But no measurements on use of ICT are reported in the findings. Administration of the current questionnaire across teacher educators in Vietnam, can lead to valuable information about the actual use of ICT for teaching and learning. As the data collected from 933 teacher educators fit the Rasch measurement model, the precise relationship among test items and ability scores can be specified. The internal structure as is expected in the construct map can be observed to a large extent in the Wright map. There is empirical evidence that there is a ranking of item locations, going from the use of ICT to replace and enhance teaching practice, over innovative use of ICT, toward the use of ICT to transform education. Items on use of ICT for teaching and support of student learning can now be ordered from "easy" to "hard" by calibration against the distribution of teacher educators, relating back to the construct map on integration of ICT in education. As such, the instrument answers the need to analyse ICT-based innovations on a continuum ranging from the assimilation level through the transition level and up to the transformation level, reflecting stages of development concerning ICT use in education (Mills & Tincher, 2003; Mioduser et al., 2003; SEAMO, 2010; UNESCO, 2005). The item thresholds addressing the use of ICT to support student learning are all located around the level of teacher educators' innovative use of ICT for teaching, as defined in the theoretical framework (Drent & Meelissen, 2008; Miller & Ewing, 2000; Mills & Tincher, 2003; Quellmalz & Kozma, 2003; UNESCO, 2003) and as described in the construct map. Therefore, the research objective to produce a scale to measure teacher educators self-reported use of ICT for teaching and support of student learning is achieved. As the data fit the Rasch model, requirements of 'fundamental measurements', as defined by Bond and Fox (2007) are met. We calculated measurement error to conclude whether the instrument does measure self-reported use of ICT for teaching and support of student learning with sufficient consistency over individuals. The relationship between the $sem(\theta)$ and the respondent location shows that for the developed instrument the most sensitive part is from approximately $-2.0$ logits to $+2.0$ logits. This corresponds with the range of most thresholds for all items of the instrument and makes that the instrument can be used for similar samples of teacher educators. If we expect on the other hand that after participation in the capacity building program teacher educators have improved their skills to integrate ICT in teaching and learning, the instrument could be augmented with more items up at the transforming end of the scale. Another validity requirement is item consistency. We calculated the mean location of respondents within each score group of all items. As for all items the mean location of each group increases as each score increases, we can conclude that all items are consistent with the construct as a whole.

In our findings we gave in detail meaning to the item thresholds. The main benefit of a questionnaire scale with item difficulties is that on the Wright map, each respondent can be located at a particular point on the logit-scale, allowing for a probability interpretation. The framework for making location estimates meaningful is one of the most important features of the construct modeling approach to measurement (Wilson, 2005). The Wright map in this study shows that teacher educators who are not using ICT much and who integrate ICT in less innovative ways are covered by the first threshold of four items (T.6.1, T.1.1, T.4.1 and T.2.1). Teacher educators who regularly use ICT and who use ICT in more innovative ways are covered by the upper item thresholds of three items (L.3.3, L.7.3 and L.1.3). If we look at the distribution of respondents on the logit-scale, we can conclude that the majority of teacher educators have a high probability to use ICT to replace their existing teaching practice or to enhance student learning from time to time. But regular, innovative use of ICT in support of student learning is still very improbable for most. None of the respondents reach the 50% probability level of the third threshold of item T.7.

## 7. Discussion

The question remains whether a model fits the data well enough to be useful in guiding the measurement process. Statistical evidence and also judgment play important roles in answering this question (Hambleton & Jones, 1993). The statistical evidence provided in our research led us to the conclusion that with the developed measurement instrument we can fundamentally measure the use of ICT for teaching and support of student learning in teacher education. On the other hand, when the Rasch model is employed, the objective is to obtain data which fit the model (Andrich, 2004). This makes a lot of sense in an item response modeling approach on constructing measures. Test items are merely a sample of all possible test items (Bond, 2003). When fit statistics show that items behave differently than expected, the researcher has to judge whether or not to include the item in the instrument. The development of a measurement instrument therefore always takes place in a series of iterations and in a dialectical relation between theory and test practice, which is in line with Bond's (2003) statement that measurement in the human sciences must be theoretically driven. In creating a solid and integrated argument for the validity of an instrument, different forms of evidence should be gathered. The forms are not merely a list of types of evidence – they bear important relationships to one another. In this research paper we gathered evidence regarding content and regarding internal structure, showing that the construct – described in the content validity evidence is reflected in observations of the instrument in operation. Gathering this evidence is not an end in itself, but a means to achieve the goal of a sound instrument. This process is not straightforward and depends also on the maturity of the theoretical construct. Different sorts of conclusions can be drawn from negative evidence, for example that the original idea of the construct was in some way wrong, that the items that were developed to match the construct are not working as intended, or that the scores that have been developed for the items are incorrect. Any sound instrument is the result of several iterations (Wilson, 2005).

Measurement instruments are typically used across a wide spectrum of respondents. An important requirement of the items design is that, across important subgroups, items function in a similar way for respondents who are at the same location – that is, they should exhibit no evidence of differential item functioning (DIF) (Wilson, 2005). Bond (2003) states that the long-term aim of a genuine measurement system in the human sciences should be access to a series of co-calibrated testing situations, such that the raw score or locally calculated person measure could be expressed on the single scale used world-wide. For the measurement instrument described in this research paper, we followed the "Technology Mapping" methodology (Angeli & Valanides, 2009), keeping in mind that ICT resources are necessarily employed within pre-existing contexts of educational and social activity (Tolmie, 2001). We developed the measurement instrument based on literature research on integration of ICT in education and input from Vietnamese practitioners and experts in the field. The selected items are therefore a reflection of the staged model of integration of ICT in education in the context of education in Vietnam. Currently, the intended use for the instrument is therefore in education in Vietnam. We can make no claims on potential universal use as we did not investigate item functioning amongst different subgroups, for example in different countries.

Data collected with self-report questionnaires should be complemented with other data collection methodologies (Jamieson-Proctor et al., 2007). Other ways to give a more detailed picture of how ICT is being used to support innovative teaching practices are case studies, or performance assessments of student learning with technology (Quellmalz & Kozma, 2003). It appears that well-done self reporting (no high stakes, valid and reliable instruments) on a large scale, complemented by well-selected (nonintrusive, randomly assigned, systematically reported) observations of the same environment, may be the most effective approach to gaining a true picture in most ICT in education environments (Christensen & Knezek, 2008). In that framework, the development of a questionnaire instrument on the use of ICT for teaching and support of student learning, following an item response modeling approach, is relatively straightforward; and also an essential element for valid and reliable benchmarking of integration of ICT in education.

## Appendix. Supplementary material

Supplementary material related to this article can be found online at doi:10.1016/j.compedu.2011.12.015.

## References


Adams, R. J., Wilson, M., & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1–23.

Akbulut, Y. (2009). Investigating underlying components of the ICT indicators measurement scale: the extended version. *Journal of Educational Computing Research, 40*(4), 405–427.

Alampay, E. A. (2006). Beyond access to ICTs: measuring capabilities in the information society. *International Journal of Education and Development Using Information and Communication Technology (IJEDICT), 2*(3), 4–22.

Andrich, D. (2004). Controversy and the Rasch model: a characteristic of incompatible paradigms? *Medical Care, 42*(1), I–7.

Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT-TPCK: advances in technological pedagogical content knowledge (TPCK). *Computers & Education, 52*(1), 154–168.

Bond, T. (2003). Validity and assessment: a Rasch measurement perspective. *Metodología de las Ciencias del Comportamiento, 5*(2), 179–194.

Bond, T., & Fox, C. (2007). *Applying the Rasch model fundamental measurement in the human sciences* (2nd ed.). USA: Psychology Press.

Bowes, J. (2003). The emerging repertoire demanded for of teachers of the future: surviving the transition. *Paper presented at the ICT and the teacher of the future (IFIP)*.

Cavanagh, R., Romanoski, J., Giddings, G., Harris, M., & Dellar, G. (2003a). Application of the Rasch model and traditional statistics to develop a measure of primary school classroom learning culture. *Paper presented at the annual conference of the Australian association for research in education*. From http://www.aare.edu.au/03pap/cav03396.pdf.

Cavanagh, R., Romanoski, J., Giddings, G., Harris, M., & Dellar, G. (2003b). Development of a Rasch model scale to measure student information and communication technology learning self-reported outcomes, behaviours and attitudes. *Paper presented at the annual conference of the Australian association for research in education*. From http://www.aare.edu.au/03pap/cav03396.pdf.

Christensen, R., & Knezek, G. (2008). Self-report measures and findings for information technology attitudes and competencies. InVoogt, J., & Knezek, G. (Eds.). (2008). *International handbook of information technology in primary and secondary education, Vol. 20* (pp. 349–365). US: Springer.

Cohen, L., Lawrence, M., & Morrison, K. (2007). *Research methods in education* (6th ed.). London and New York: Routledge.

Cox, M. J. (2008). Researching IT in education. InVoogt, J., & Knezek, G. (Eds.). (2008). *International handbook of information technology in primary and secondary education, Vol. 20* (pp. 965–981). US: Springer.

Drent, M., & Meelissen, M. (2008). Which factors obstruct or stimulate teacher educators to use ICT innovatively? *Computers & Education, 51*(1), 187–199.

Gareis, K., & Hüsing, T. (2009). Measuring transformational use of ICTs at regional level. In V. Weerakkody, M. Janssen, & M. Y. Dwivedi (Eds.), *Handbook of research on ICT-enabled transformational government: A global perspective* (pp. 351–378).

Hagquist, C., & Andrich, D. (2004). Is the sense of coherence-instrument applicable on adolescents? A latent trait analysis using Rasch-modelling. *Personality and Individual Differences, 36*(4), 955–968.

Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*(3), 38–47.

Jamieson-Proctor, R., Watson, G., Finger, G., Grimbeek, P., & Burnett, P. C. (2007). Measuring the use of information and communication technologies (ICTs) in the classroom. *Computers in the Schools, 24*(1), 167–184.

Linacre, J. M. (2002). Understanding Rasch measurement: optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85–106.

Maddux, C. D., & Johnson, D. L. (2005). Type II applications of technology in education: new and better ways of teaching and learning. *Computers in the Schools, 22*(1/2), 1–5.

Marshall, G., & Cox, M. J. (2008). Research methods: their design, applicability and reliability. InVoogt, J., & Knezek, G. (Eds.). (2008). *International handbook of information technology in primary and secondary education, Vol. 20* (pp. 983–1002). US: Springer.

Miller, D., & Ewing, J. (2000). Beyond knowledge transmission? Computer-supported learning in teacher education: some benefits in terms of stress, control and self-belief. *Journal of Information Technology for Teacher Education, 9*(3), 363–376.

Mills, S. C., & Tincher, R. C. (2003). Be the technology: a developmental model for evaluating technology integration. *Journal of Research on Technology in Education, 35*(3), 382.

Mioduser, D., Nachmias, R., Tubin, D., & Forkosh-Baruch, A. (2003). Analysis schema for the study of domains and levels of pedagogical innovation in schools using ICT. *Education and Information Technologies, 8*(1), 23–36.

Partnership on Measuring ICT for Development. (2010). *Core ICT indicators, 2010*. Geneva: United Nations Partnership on Measuring ICT for Development.

Pilkington, R. M. (2008). Measuring the impact of information technology on students' learning. InVoogt, J., & Knezek, G. (Eds.). (2008). *International handbook of information technology in primary and secondary education, Vol. 20* (pp. 1003–1018). US: Springer.

Plomp, T., Anderson, R. E., Law, N., & Quale, A. (Eds.). (2003). *Cross-national policies and practices on information and communication technology in education*. Information Age Publishing Inc.

Proctor, R. M. J., Watson, G., & Finger, G. (2003). Measuring information and communication technology (ICT) curriculum integration. *Computers in the Schools, 20*(4), 67–87.

Quellmalz, E. S., & Kozma, R. (2003). Designing assessment of learning with technology. *Assessment in Education: Principles, Policy & Practice, 10*(3), 389–408.

Richards, C. (2004). From old to new learning: global imperatives, exemplary Asian dilemmas and ICT as a key to cultural change in education. *Globalisation, Societies & Education, 2*(3), 337–353.

SEAMO. (2010). *Status of ICT integration in education in Southeast Asian countries*. Bangkok: The Southeast Asian Ministers of Education Organization (SEAMO).

Steffe, L., & Gale, J. (Eds.). (1995). *Constructivism in education*. New Jersey: Lawrence Erlbaum Associates Inc.

Tolmie, A. (2001). Examining learning in relation to the contexts of use of ICT. *Journal of Computer Assisted Learning, 17*(3), 235–241.

Tubin, D. (2006). Typology of ICT implementation and technology applications. *Computers in the Schools, 23*(1/2), 85–98.
UNESCO. (2003). *Performance indicators for ICT in education – Consultative workshop*. Bangkok: UNESCO Asia and Pacific Regional Bureau for Education.
UNESCO. (2005). *Regional guidelines on teacher development for pedagogy-technology integration*. Bangkok: UNESCO Asia and Pacific Regional Bureau for Education.
Van Braak, J., Tondeur, J., & Valcke, M. (2004). Explaining different types of computer use among primary school teachers. *European Journal of Psychology of Education - EJPE, 19*(4), 407–422.
Wagner, D. A., Day, B., James, T., Kozma, R. B., Miller, J., & Unwin, T. (2005). *Monitoring and evaluation of ICT in education projects: A handbook for developing countries*. Washington, DC: infoDev, International Bank for Reconstruction and Development/ The World Bank.
Ward, L., & Parr, J. M. (2010). Revisiting and reframing use: implications for the integration of ICT. *Computers & Education, 54*(1), 113–122.
Wilson, M. (2005). *Constructing measures: An item response modelling approach*. Mahway: Erlbaum.
Zhang, J. (2007). A cultural look at information and communication technologies in eastern education. *Educational Technology Research and Development, 55*(3), 301–314.