Time for a new approach to prepare future teachers for educational technology use
— **Source link** ↗

Jo Tondeur, Johan van Braak, Fazilat Siddiq, Ronny Scherer

Institutions: Research Foundation - Flanders, Ghent University, University of Oslo

Related papers:

- Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence

- Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge

- Technological pedagogical content knowledge – a review of the literature

- Student teachers' thinking processes and ICT integration: Predictors of prospective teaching behaviors with educational technology

- The impact of authentic learning experiences with ICT on pre-service teachers' intentions to use ICT for teaching and learning

Share this paper: 🔵 🐦 in ✉

View more about this paper here: https://typeset.io/papers/time-for-a-new-approach-to-prepare-future-teachers-for-5absshsogl

Time for a new approach to prepare future teachers for educational technology use:

Its meaning and measurement

Abstract

The main objective of this study is to develop a self-report instrument to measure pre-service teachers' perceptions of the extent to which they experience the necessary support and training in order to integrate technology into classroom activities. The questionnaire items of this instrument were drawn up on the basis of a synthesis of 19 qualitative studies (Authors et al., 2012) and were reviewed by experts in the field. In order to study its reliability and aspects of validity, data were collected and analysed consisting of a sample of 688 pre-service teachers in Flanders (Belgium). The resulting scale showed highly satisfactory psychometric properties. Item response theory revealed a good fit of the measurement to a Rating Scale Model for 22 out of 24 items. The results also indicate that the items differ in their degree of difficulty. It seems that helping pre-service teachers to design ICT-rich lessons and providing adequate feedback can be considered more challenging for teacher training institutions. Recommendations are given regarding how the new scale can be useful for both teacher training institutions and schools in developing approaches to equip pre-service teachers with the competencies needed to integrate technology in teaching and learning processes.

Keywords: Pre-service teachers; Teacher training; Technology use; Rasch analysis; Scale development

1. Introduction

In this digital age, teacher training institutions (TTI) are expected to prepare pre-service teachers to adequately use technology in their educational practice (e.g., Brun & Hinostroza, 2014; ISTE, 2012; Kaufman, 2015). TTIs around the world have therefore engaged in various efforts to re-shape their curriculum (Ottenbreit-Leftwich, Glazewski, Newby, & Ertmer, 2010; Tømte, Enochsson, Buskqvist, & Kårstein, 2015). More specifically, they have infused technology into the entire curriculum, giving pre-service teachers the opportunity to understand the pedagogical reasons for using technology by experiencing first-hand how it can support teaching and learning across different subjects (Authors, 2012; Polly Mims, Shepherd, & Inan, 2010). A number of approaches that may develop the competences that future teachers require in this context have already been identified by previous research (Chien, Chang, Yeh, & Chang, 2012; Lee & Lee, 2014; Valtonen et al., 2015). However, promoting pre-service teachers' competencies for educational technology use in an integrated and cross-curricular manner is a complex process that demands multiple strategies (Kay, 2006; Mouza, Karchmer-Klein, Nandakumar, Ozden, & Hu, 2014; Polly Mims, Shepherd, & Inan, 2010).

These strategies were listed and reviewed by Author (2012b) and an overarching SQD-model (Synthesis of Qualitative Evidence) was developed to present how they relate to each other/to present their complex interrelations. After this model had been introduced, it became clear that the next step would be to assess to what extent these strategies are actually implemented by TTIs. However, no comprehensive instruments that assess effective strategies to prepare future teachers for technology integration have been described in the research literature. This study intends to address this gap in the existing literature by providing an instrument that does precisely that. More concretely, the current research has three specific aims: 1) to develop a self-report instrument based on a theoretical model to measure pre-

service teachers' perceptions of the extent to which they experience the necessary support and training in order to integrate technology into classroom activities; 2) to use IRT (Item Response Theory) to establish a reliable scale; and 3) to explore the item difficulties of strategies to prepare pre-service teachers for technology use. This will lead to a better understanding of the support future teachers need for the use of technology in education.

Before presenting the results of the study, we first examine the literature upon the complexity of technology in teacher education and the choice of IRT as a method of data collection. This will be followed by a description of the design study and conclude with a discussion on the implications for practice: how the proposed instrument could be used 1) to frame and improve the training of future teachers to effectively integrate technology in their teaching and learning practices and 2) in structuring the professionalization of the teacher educators.

2. Research Overview

2.1 Teacher education and technology use

The new generation of teachers faces considerable demands with respect to the use of new technologies in education (OECD, 2010). As a result, TTIs are compelled to respond to a society driven by technology and to seize the opportunities it creates for education (Chien et al., 2012; Kaufman, 2015). According to the standards set for teachers, TTIs should facilitate such learning events, design and develop learning experiences and assessments fit for the digital age, model 21st century learning, promote digital citizenship, and engage in ongoing professional development and leadership of technology use (ISTE, 2012; Kennisnet, 2012; UNESCO, 2008).

In this respect, TTIs are expected to provide new teachers with the competencies to integrate technology into education and to educate pupils to become ICT competent (see e.g., Brun & Hinostroza, 2014; Tømte, Enochsson, Buskqvist, & Kårstein, 2015). To respond to

this expectation, many institutions have included introductory technology courses, primarily focused on the development of technological knowledge and skills (Polly et al., 2010). However, several studies suggest that technology is often under-used by pre-service teachers and beginning teachers (Agyei & Voogt, 2011; Chien et al., 2012; Sadaf, Newby, & Ermer, 2012). Moreover, it seems that only a small number of pre-service and beginning teachers are able to use technology in diverse and flexible ways (Authors, 2012; Bate, 2010; Gao, Wong, Choy, & Wu, 2011).

According to Ottenbreit-Leftwich et al. (2010), there is a clear discrepancy between what pre-service teachers are taught in their courses and how teachers actually use technology in a real classroom. Therefore, Koehler and Mishra (2009) argued that TTIs should not only focus on how to use technology but also how technology intersects with pedagogical and content knowledge, as directed by the concept of Technological Pedagogical Content Knowledge (TPACK). Further research on the development of ICT capabilities for pre-service teachers reveals that technology should be infused into the entire curriculum (Authors, 2014). Without such integrated approaches, Polly et al. (2010) predict that the competencies pre-service teachers gain are likely to remain isolated and unexploited. Practical experiences worldwide have corroborated the potential of such integrated approaches as well as the difficulties associated with their implementation (Drent & Meelissen, 2008; Mouza et al., 2014).

2.2 Strategies to prepare future teachers for technology use

Numerous strategies have been proposed in the literature to facilitate integrated approaches to prepare pre-service teachers for technology use (e.g. Agyei & Voogt, 2011; Banas, & York, 2014; Jang, 2008; Niess, 2005). To illustrate, the "Preparing Tomorrow's Teachers to Use Technology" (PT3) programme provided funds to support the development of teacher technology learning experiences (see Authors, 2012). This programme emphasized

technology training in authentic teaching situations (based on Kay, 2006; Ottenbreit-Leftwich et al., 2010; Polly et al., 2010): hands-on technology skill-building activities (e.g., workshops), collaboration among pre-service teachers (e.g., design teams), practicing technology in the field (e.g., field experiences), and technology integration reflections (e.g., electronic portfolios).

It is clear that effective preparation of pre-service teachers for technology integration does not only require attention to the separate strategies, but also the combination of different approaches (Chien et al., 2012; Kay, 2006; Mouza et al., 2014). To illustrate, the study of Chien et al. (2012) is based on a combination of four steps (i.e., Modelled Analysis, Guided Development, Articulated Implementation, and Reflected Evaluation) for "assisting science teacher educators in closing the gap between instructional design and technology design while teaching technology integration" (p. 579). Specifically, the aim of such interventions is to transform pre-service teachers from passive users of technology into active designers of technology (see also Koehler & Mishra, 2009; Lee & Lee, 2014; Sang et al., 2010).

2.3 Towards an evidence-based model to inform teacher education

As illustrated above, there are different strategies for the content and delivery methods to prepare pre-service teachers for technology use (e.g., Angeli & Valanides, 2009; Mouza et al., 2014; Niess, 2005). Still, the question remains how TTIs can get a clear overview of effective strategies. According to Chien et al. (2012) there is a need for evidence-based models to inform teacher educators globally. In his review of quantitative studies on previous pre-service teacher technology training, Kay (2006) concluded that the jury is still out on what strategies work best. To tackle this uncertainty, Authors (2012) reviewed 19 qualitative studies in order to deliver an SQD-model on which content and delivery methods best prepare pre-service teachers to integrate technology into their future classrooms.

The current study adopted the output of Authors' (2012) review as a theoretical

foundation to build an instrument. In their review, Authors suggest that twelve key themes

need to be in place in the teachers' education programme to prepare student teachers for

technology integration (see Fig. 1). These key themes are either related to the preparation of

pre-service teachers at the micro level (e.g., using teacher educators as role models, learning

technology by design, scaffolding authentic technology experiences), or to conditions

necessary at the institutional level (e.g., technology planning and leadership, co-operation

within and between institutions, training staff).

<center>< Insert Figure 1 about here ></center>

Six of these twelve themes concern strategies at the micro level, which will be

examined in the present study. The first of these six strategies (Strategy 1) involves teacher

educators acting as role models. Although this has proven to be an important motivator for

future technology integration in the classroom (Kaufman, 2015; Tearle & Golder, 2008),

simply having pre-service teachers watch examples of technological applications is helpful

but not sufficient. As pre-service teachers should also be able to interpret these examples in a

specific educational context, a second strategy could consist of observing, discussing and

reflecting upon successful uses of technology (Strategy 2). This might help them see the

utility, value and feasibility of using a particular technology and/or teaching strategy (cf.

Goktas, Yilderim, & Yildirim, 2009; Mouza et al., 2014), hence furthering their ability to

differentiate between action and purpose and enabling deeper and more critical thinking about

technology integration.

Besides the importance of role models and reflection, research suggests that providing

the opportunity to learn about technology integration by (re-)designing curriculum materials

(Strategy 3) can also be a promising strategy (Angeli & Valanides, 2009; Lee & Lee, 2014).

In several studies, pre-service teachers have stated that technology integration required

additional planning and preparation because they had no prior knowledge about or experience with the design of ICT-supported learning activities (e.g., Polly et al., 2010). Many studies have demonstrated that group work (Strategy 4) might mitigate these feelings of insecurity when teachers need to design technology-related curriculum materials (Tearle & Golder, 2008). The usefulness of this fourth strategy has been demonstrated by Angeli and Valanides (2009), who have confirmed that collaboration with peers provides a less threatening learning environment for pre-service teachers, thereby reducing anxiety and failure avoidance (see also Jang, 2008).

As a fifth strategy, pre-service teachers may also apply their knowledge of educational technology in authentic settings (Strategy 5). These types of engaging experiences will give them a better understanding of the link between theory and teaching practices (Sang et al., 2010; Valtonen et al., 2015). Finally, the sixth strategy involves on-going and process-oriented feedback (Strategy 6), which has been proven to be beneficial for pre-service teachers' abilities to use technology in the classroom (Banas & York, 2014; Chien et al., 2012). All these strategies have contributed to defining the questionnaire items in the self-report instrument, which was the purpose of the current study.

2.4 Purpose of the study: Measuring strategies to prepare future teachers for technology use

There are already many instruments available to quantitatively measure pre- or in-service teachers' technology competency (e.g., Chien et al., 2012; Christensen & Knezek, 2008; Sang et al., 2015). However, it is more difficult to measure the efforts of TTIs to develop the competencies pre-service teachers need to successfully integrate technology into teaching. Therefore, the present study aims to provide an instrument to measure pre-service teachers' perceptions of the extent to which they experience the necessary support and training in order to integrate technology into classroom activities. Specifically, this study has three goals: 1) to develop a self-report instrument based on a theoretical model; 2) to use IRT

to establish a reliable scale; and 3) to explore the item difficulties of strategies to prepare pre-service teachers for technology use. Ultimately, the results will lead to a better understanding for reliable benchmarking of the support future teachers need for the use of technology in education in TTIs.

This self-report instrument is based on a large-scale study in Flanders (Belgium). It measures to what extent TTIs are integrating the six strategies included in the SQD-model (see Figure 1): 1) using teacher educators as role models, 2) reflecting on the role of technology in education, 3) learning how to use technology by design, 4) collaboration with peers, 5) scaffolding authentic technology experiences, 6) Providing continuous feedback. To create the instrument, we decided to apply Rasch analysis for several reasons. First, it is an acclaimed method to develop new scales, as it makes it possible to design the item set so that it meets model expectations from the very beginning. Second, as argued by Boone, Staver, and Yale (2014), as well as Wilson (2005), it allows for a unified approach to all measurement issues required for the transformation to interval scaling. Finally, the items of the scale can be ordered from "easy" to "hard" (Research goal 3). Nevertheless, testing data for fit to the Rasch model is rarely done for questionnaire scales on use of technology (cf. Aesaert, Van Nijlen, Vanderlinde, & van Braak, 2014) and, to our knowledge, no comprehensive instruments that assess support to prepare future teachers for technology integration have been described in the research literature.

3. Research Design

3.1 Procedure and Sample

The questionnaire items were developed based on the strategies in the inner circle of the SQD-model (Synthesis of Qualitative Evidence). This guiding model was developed as a result of a review of qualitative evidence (Authors, 2012). It appears that these strategies have a positive effect on pre-service teachers' competencies to integrate technology into their

teaching and learning processes. A first version of the questionnaire was reviewed by 15 different stakeholders (e.g., pre-service teachers, teacher trainers, ICT-coordinators and researchers). This assessment focused on identifying relevant items for the six different strategies. The output was used for exploring content validity of the instrument and for providing experts with the chance to suggest re-phrasing or supply new items, fine-tuning the wording of the items and reducing item complexity (cf. Hyrkäs, Appelqvist-Schmidlechner, & Oksa, 2003). This step resulted in the refinement of the instrument, narrowing the items down to 24. After consulting the stakeholders one more time, we were able to optimize the wording of the resulting items. This new version of the instrument is presented in Table 1.

< Insert Table 1 about here >

The final item set was surveyed online in 2014 with a sample of 688 last-year pre-service teachers. The heads of department from 21 teacher training institutions in Flanders (the Dutch speaking part of Belgium) were contacted to contribute to this research. Twenty department heads were willing to participate in this study, representing a high response rate at institutional level. The pre-service teachers completed the questionnaire anonymously. Four participants provided extreme responses at the lowest and highest categories of the SQD-scale, and were thus excluded from the dataset. The final sample comprised $N = 684$ pre-service teachers (females: 74.1%). The representativeness of the sample was given for gender, as the under-representation of males is a common phenomenon in teacher education in Flanders. The average age was 25.0 years ($SD = 7.3$ years). Of these pre-service teachers, 57.7% had obtained a Bachelor's degree in higher education, whereas 42.3% had obtained a specific teacher degree from universities, colleges, or centres for adult learning. Concerning the minimum sample size in IRT analyses, there is no gold standard that can be proposed. According to Thorpe and Favia (2012), IRT models can be estimated with 250 respondents, but around 500 are recommended for accurate parameter estimates.

3.2 Measure

In line with the strategies in the SQD-model, it was decided not to stress the technical aspects required for the training of pre-service teachers in the use of technology in education; consequently, technological terminology was eliminated. Instead, the current study will focus on the instructional and pedagogical tasks associated with the organization of technology within the learning environment.

The scale was constructed around the six significant domains of the inner circle (i.e., the micro level) of the SQD-model (see Figure 1), a model based on the synthesis of qualitative evidence (Authors, 2012). Cronbach's Alpha (✔) and McDonald's Omega (ɣ) were used as estimates for scale reliability. The latter may provide a more precise measure than the more established coefficient of Cronbach's ✔ (Revelle & Zinbarg, 2009). Each item in the questionnaire was presented as a statement revolving around the six strategies. The labels and item wordings are presented in Table 1. Respondents were asked to rate each statement on a six-point Likert scale: (1) totally disagree (2) disagree (3) slightly disagree (4) slightly agree (5) agree (6) totally agree. Each of the six domains showed good reliabilities: *Using teacher educators as role models* (ROL, 4 items; ✔ = .91, ɣ = .90), *Reflecting on attitudes about the role of technology in education* (REF, 4 items; ✔ = .89, ɣ = .88), *Learning technology by design* (DES, 4 items; ✔ = .90, ɣ = .89), *Collaborating with peers* (COL, 4 items; ✔ = .87, ɣ = .85), *Scaffolding authentic technology experiences* (AUT, 4 items; ✔ = .83, ɣ = .83), and *Moving from traditional assessment to continuous feedback* (FEE, 4 items; ✔ = .93, ɣ = .92). The entire scale comprised 24 items and exhibited an excellent overall reliability, ✔ = .98, ɣ = .90.

3.3 Data analysis

In previous studies, several self-report instruments to measure the use and integration of ICT in educational contexts have been developed and verified by means of classical test

theory (CTT; Christensen & Knezek, 2008). In CTT, the analyses are performed on the test as a whole rather than on single items (Hambleton & Jones, 1993). Although this makes it possible to generate item statistics such as item-to-total correlations, these still only apply to *that* particular group of respondents on *that* particular collection of items (de Ayala, 2013). By contrast, item response theory (IRT) emphasizes the underlying trait affecting test performance irrespective of the respondents and the actual test (e.g., Wilson, 2005). IRT has therefore often been described as a successor to CTT in psychological assessment (Thomas, 2011). Some of the advantages of using IRT include an increase in measurement accuracy of the assessment instrument, the reduction of measurement error, and the estimation of item and person parameters independent of the item or person sample (Hambleton & Jones, 1993; Thomas, 2011). These benefits point to the transferability and the reuse of measurement instruments, together with the possibility "to design instruments with specific characteristics" (de Ayala, 2013, p.147). In this respect, given that the main goal of the present study is to develop a specific instrument to measure pre-service teachers' perceptions of the support and training to integrate ICT in classrooms, IRT and particularly Rasch modelling provide a promising psychometric approach (Boone et al., 2014; McCoach, Gable, & Madura, 2013).

The Rasch model (i.e., the one-parameter logistic model) was first described by Georg Rasch (1960), and refers to an item response model with the premise that all items in the test have the same discrimination, implying that only item difficulty can vary (Bond & Fox, 2007; Thomas, 2011). The Rasch model was originally developed as a unidimensional latent variable model for handling dichotomous data (i.e., responses are coded as 1 if correct and 0 if incorrect) (Bond & Fox, 2007; Rasch, 1960). Further extensions of the Rasch model for dealing with polytomous data were proposed, namely the Rating Scale Model (Andrich, 1978) and the Partial Credit Model (Wright & Masters, 1982).Given that pre-service teachers' responses comprise six response categories, polytomous Rasch models were applied in the

present study. The software *RUMM2030* was used to analyse the data with the help of marginal maximum likelihood estimation (Andrich, Sheridan, & Luo, 2012). Missing responses were coded as missing and accounted for by the full-information maximum likelihood procedure (Enders, 2010).

Assumptions of IRT models. There are three underlying assumptions common to unidimensional IRT models: (1) the observations on the manifest variables are a function of a continuous latent person variable, the trait (*unidimensionality*) (Linacre, 1998); (2) the responses to items are independent, meaning that the response to an item should not predict the correct answer to another item (*local independence*) (de Ayala, 2013; Smith, 2002); and (3) the probability of endorsing an item increases as the trait level increases (*monotonicity*) (de Ayala, 2013). For the Rasch model, an additional assumption is that all items discriminate equally (de Ayala, 2013).

The concept of local independence assumes that the success or failure of endorsing one item does not depend on the success or failure of endorsing another item. This implies that there are no dependencies between the items except those attributed to the latent trait (Marais & Andrich, 2008; Yen, 1993). Violations of local independence may be indicated by residual correlations and multidimensionality (Ip, 2010). For residual correlations, a value higher than .20 was considered to be indicative of local dependencies by De Ayala (2013). Moreover, if there is evidence on the multidimensionality of a scale, which can for instance be obtained from factor-analytic approaches (e.g., exploratory factor analysis; McCoach et al., 2013), the items reflect more than one trait, violating the unidimensionality assumption of Rasch models (Bond & Fox, 2007). Hence, for evaluating the quality of the SQD-scale by using Rasch models, testing for item dependencies and multidimensionality is considered to be essential.

Polytomous Rasch models. In a Likert scale with several categories, the respondents need to interpret the distance between the categories (e.g., between *totally disagree* and *disagree*, between *disagree* and *slightly disagree*, and so forth). If this distance between the item locations (in logits) of neighbouring response categories differs across the items, the polytomous Rasch model is referred to as the "unrestricted" Partial Credit Model (Wright & Masters, 1982). In contrast, if the distance between the categories is interpreted as equal, the more "restricted" Rating Scale Model applies (Andrich, 1978; Bond & Fox, 2007).

The location of the probability of responding in one of two adjacent response categories reaches 0.5, namely a *threshold* (Andrich, 1995), is used to study the functioning of the response categories in the test instrument. Although these thresholds are normally ordered in polytomous Rasch models that are based on rating scales, disordered item locations of the response categories may occur (Andrich, de Jong, & Sheridan, 1997). However, disordered categories do not necessarily indicate item misfit, as was shown by Adams, Wu, and Wilson (2012).

Model and item fit. De Ayala (2013) pointed out that the fit between the empirical model and the observed data is a necessary, yet insufficient condition for the validity of a test instrument. Evidence on the model-data fit can be obtained from test- and item-level statistics. The analysis of *test-level statistics* usually comprises the assessment of multidimensionality and likelihood-based information criteria such as Akaike's Information Criterion (*AIC*) and the Bayesian Information Criterion (*BIC*). Although the latter do not provide information on the actual fit between the model and the data, they can be used to compare models with different characteristics (e.g., Rating Scale versus Partial Credit Model). In this regard, a model with lower *AIC* and *BIC* values fits better than the competing model (Burnham & Anderson, 2004). It is noteworthy that the choice of a particular model should not only be based on the *AIC* but also on the *BIC*, which accounts for the sample size. Besides these

criteria, Fisher's likelihood-ratio test, which is based on the log-transformed likelihood values (*LogL*), can be performed to compare competing models (de Ayala, 2013). Moreover, in order to obtain information on the measurement error for respondents and the information the test provides, the test information curve and the standard error of estimation plot are evaluated (for details, please consult Wilson, 2005).

At the *item level*, item fit statistics, residual correlations, and the differential functioning across subgroups can be investigated. For instance, the *Fit Residuals* that are provided by RUMM2030 as item fit statistics represent a weighted difference between a person's response and the expected response under the model assumptions (Andrich et al., 2012). In order to evaluate these differences, Andrich and Kreiner (2010) suggested using cut-offs of ±2.5 for an adequate fit to the model. In this respect, fit residuals below –2.5 indicate an overfit and above +2.5 an underfit, respectively. Furthermore, $\square$ tests can be performed for each item to identify significant deviations of the model-based item characteristic curves from the observed item characteristic curves (Hagquist, Bruce, & Gustavsson, 2009).To account for multiple difference testing, the *p* values of the $\square$ statistics are Bonferroni adjusted (Bland & Altman, 1995).

Differential item functioning (DIF). The term "differential item functioning" (DIF) is used to describe items that function differently across different groups within the sample (e.g., gender, ethnic, or socioeconomic groups), given that equal levels of the trait are measured (Tennant & Pallant, 2007; Wilson, 2005). Two types of DIF can be distinguished: in *non-uniform DIF*, the magnitude of DIF varies across different levels of the latent trait, whereas in *uniform DIF* there is no variation (cf. Hagquist et al., 2009, p. 383). In the presence of non-uniform DIF, significant interaction effects between group membership and the latent trait exist, and may suggest that these items need to be discarded (see Andrich & Hagquist, 2001). However, items showing uniform DIF could be retained, for instance by

treating them as group-specific rather than common items. In the present study, we will check for DIF to test the generalizability and invariance of the SQD-scale at the item level.

Person reliability. The person separation index (PSI) is an index of internal consistency that is comparable to Cronbach's ✔ (Hagquist et al., 2009). This index is calculated using estimates from the Rasch model and represents the power of the test to discriminate amongst the respondents. Generally speaking, higher values of this index for the SQD-scale indicate higher discrimination between persons. The PSI is included next to the model and item fit to add a person perspective to our analyses.

4. Results

4.1 Descriptive statistics and exploratory factor analyses

Our analyses of the classical descriptive statistics did not reveal any floor or ceiling effects for the 24 items (see Table 2). In fact, the mean response tendencies met the actual mean categories of the rating scale. Moreover, the corrected item-to-total correlations (r_{it}) were above .50 and, together with the high reliability of the scale, provided evidence for scale homogeneity.

< Insert Table 2 about here >

In order to study the dimensionality of the 24-item scale, we first ran an exploratory factor analysis (EFA) on the total sample. Given that pre-service teachers' responses were categorical, we estimated the EFA on the basis of polychoric correlations in the R package *psych* (Revelle, 2014). The number of factors identified by the scree plot and eigenvalue criteria was reduced to a single factor, explaining 63% of variance. Factor loadings on this factor were high (see Table 3). Given that some research suggests that EFA results may depend on the sample and may therefore produce inconsistent findings (Osborne & Fitzpatrick, 2012), we randomly drew 20 samples comprising $n = 200$ pre-service teachers

each and re-ran the EFA. For each random sample, only one factor was identified, explaining

between 59% and 63% of variance. Moreover, the average factor loadings across the twenty

samples did not substantially deviate from those obtained from the total sample (Table 3).

Hence, we concluded that the newly developed scale was unidimensional.

< Insert Table 3 about here >

4.2 IRT analyses

In order to evaluate the scale properties and the psychometric quality of the newly

developed measure of the extent to which effective strategies are implemented in teacher

training ('SQD-scale'), polytomous item response theory (IRT) models that take into account

the categorical nature of teachers' responses were specified. In the following section, we will

describe these models with respect to their overall fit (see 'Model selection'), the properties of

the original SQD-scale (see 'Psychometric properties of the entire scale'), and a reduced scale

(see 'Psychometric properties of the reduced scale').

Model selection. As a first step, we used the entire scale of 24 items to test which

polytomous Rasch models represented the data. Under the assumption that each item has a

unique structure of the rating scale, the Partial Credit Model revealed the following fit

statistics: $LogL = -20\,207.7$, number of parameters $= 121$, $AIC = 40\,657.4$, $BIC = 41\,205.2$.

Under the assumption that all items have the same structure of the rating scale, the following

fit statistics were obtained from the Rating Scale Model: $LogL = -20\,265.2$, number of

parameters $= 29$, $AIC = 40\,588.3$, $BIC = 40\,719.7$. Fisher's Likelihood-Ratio Test revealed no

significant difference between the deviances of the two polytomous Rasch models,

χ^2 [92] $= 115.0$, $p = .053$. Moreover, the values of the information criteria were lower in the

Rating Scale Model, $\Delta AIC = 69.1$, $\Delta BIC = 485.5$. These findings corroborated our preference

for the Rating Scale Model, which assumed that the rating scale worked equally well across

the 24 items. Hence, we evaluated the psychometric properties of the SQD-scale on the basis of this model.

   Psychometric properties of the entire scale. Based on the Rating Scale Model, we first studied the item fit statistics in order to disentangle the specific fit between the observed and the expected category probability curves (de Ayala, 2013). As indicated by the fit residuals that ranged between –7.11 and 13.19 and exceeded the suggested cut-off criteria of ±2.5, several items showed a potential misfit to the model. Particularly for item *AUT2* (*During my pre-service training, I was able to learn to use ICT in the classroom through the internships*), an improbable fit of the polytomous Rasch model was indicated (*Fit Residual* = 13.19), meaning that the observed and expected response functions differed significantly. Nevertheless, besides the item fit statistics, the person separation index of 0.969 for the set of 24 items confirmed that the scale worked well with excellent power to discriminate between the persons.

   As a second step, we studied violations of the local item independence assumption by evaluating residual correlations among the 24 items. These correlations ranged between –.21 and .33, but were on average rather low, $M (r) = -.04$, $SD (r) = .08$. For three out of 276 item pairs, these correlations exceeded the suggested cut-off of .20. Interestingly, item *FEE2* correlated significantly with two items: *FEE3* ($r = .25$) and *FEE4* ($r = .33$). These correlations appear reasonable in light of the item formulations of *FEE2, FEE3*, and *FEE4*; all refer to aspects of developing and evaluating teachers' ICT competences, leading to related responses on these items (see Table 1). We therefore identified item *FEE2* as worrisome with respect to item dependencies.

   Against this background, we decided to exclude items *AUT2* and *FEE2* in order to reduce misfit and to decrease the number of residual correlations that may violate the assumption of local independence. Starting with item *AUT2*, we excluded one item at a time,

because the fit of the items depends on the frame of reference (Hagquist et al., 2009). For each step of excluding items, we evaluated the changes in model and item fit according to previously described guidelines. Excluding item *AUT2* did not significantly change the item fit statistics for the remaining items of the scale. Nevertheless, the model fit improved after the exclusion of *AUT2*, $LogL = -19\ 139.1$, number of parameters $= 28$, $AIC = 38\ 334.3$, $BIC = 38\ 461.1$; $\chi^2 [1] = 2\ 252.1$, $p < .001$, $\Delta AIC = 2\ 254.1$, $\Delta BIC = 2\ 258.6$. Although model fit had improved, the significant residual correlations within the item pairs *FEE2–FEE3* ($r = .24$) and *FEE2–FEE4* ($r = .31$) still remained. Item *FEE2* was subsequently excluded, generating a reduced scale of 22 items.

 Psychometric properties of the reduced scale. In comparison to the entire 24-item scale, the fit of the Rating Scale Model to the reduced 22-item scale indicated an improvement in the information criteria, although the number of parameters only changed marginally: $LogL = -18\ 326.8$, number of parameters $= 27$, $AIC = 36\ 707.7$, $BIC = 36\ 829.9$; $\Delta AIC = 3\ 880.7$, $\Delta BIC = 3\ 889.7$. The decrease in the $LogL$ value was statistically significant, $\chi^2 [2] = 3\ 876.7$, $p < .001$. Furthermore, the person separation index of 0.968 for the 22-item scale did not change substantially.

 Item fit residuals ranged between –5.64 and 7.00 and indicated an improvement in item fit. Nevertheless, only the residual of item *COL3* (*During my pre-service training, students helped each other to use ICT in an educational context*) was significant at the Bonferroni adjusted level (*Fit Residual* $= 7.00$, $p < .000454$). Inspecting the characteristic curve of this item showed that the observed values by and large fit the expected values well; however, there was some misfit at the lower and upper ends of the curve (see Figure 2). A further qualitative investigation of *COL3* revealed that the wording appeared imprecise: the item could be understood in different ways, because "students" may either refer to pre-service teachers themselves or the students they teach during their internship. We suggest

reformulating the item into '*During my pre-service training, student teachers helped each other to use ICT in an educational context'*, so that it is clear what students are referred to.

< Insert Figure 2 about here >

The category probability curves showed ordered thresholds for all items, particularly for item *COL3*. Following Bond and Fox (2007), we therefore decided to keep this item in the reduced SQD-scale.

< Insert Figure 3 about here >

The item location parameters and residual item fit statistics for the 22-item SQD-scale are shown in Table 4. It is noteworthy that the item fit statistics improved noticeably ($M = -0.44$, $SD = 3.01$), compared to the original 24-item scale ($M = -0.56$, $SD = 4.17$). Moreover, the mean location of the persons (0.18 logits), which is close to the centralized mean of the item locations (0.00 logits), implied that the SQD-trait of the respondents corresponded well with the difficulty level of the scale. Hence, the items are well-targeted for the respondents in the sample of pre-service teachers.

< Insert Table 4 about here >

The person-item map ("Wright map") of the reduced SQD-scale supported this match between item and person locations and further revealed that all item thresholds were ordered, suggesting that the response categories provide an ordered set of responses from "totally disagree" to "totally agree" (see Figure 4). The discrimination between the thresholds of response categories was satisfactory and the relative locations of the items in the Wright map showed a good coverage of the latent trait on the person side. Only few areas on the logit scale (e.g., for extremely low and high values of the latent trait) did not fully match items and persons. However, by and large, the correspondence between the two sides of the Wright map was acceptable and indicated a good representation of the construct map (Wilson, 2005).

The item difficulty in the Rating Scale Model reflects the degree to which pre-service teachers agreed with the statement presented in the item (see Figure 4). To illustrate: the items related to feedback (FEE4: "My competences in using ICT in the classroom were regularly evaluated") are consistently higher up in the Wright Map, which implies that this item is a more 'difficult' one to agree with, compared to the probability of a positive response to the easiest item (AUT3: "I was encouraged to gain experience in using ICT in a classroom setting"). Also the items related to the design of ICT-rich lessons appear to be more difficult. For instance, Item DES4 "I received a great deal of help developing ICT-rich lessons and projects to use for my internship" is located high in the Wright Map. As a consequence, items on strategies to prepare pre-service teachers can be ordered from "easy" to "hard".

< Insert Figure 4 about here >

After the two items were deleted, the range of the residual correlations decreased, $M (r) = -.05$, $SD (r) = .08$, $Min = -.21$, $Max = .26$. Only one out of 231 item pairs (*ROL1– ROL3*) had a correlation of .26, which can be explained by similar item wordings (see Table 1). Nevertheless, given that these two items did not correlate highly with further items, we decided to keep them in the final scale. Moreover, a principal component analysis of the item residuals revealed that the first principal component explained only 8.89% of variance (see Table 5), suggesting that there was no evidence on substantial item dependencies and multidimensionality (Smith, 2002). Hence, deleting items *FEE2* and *AUT2* improved the degree to which the Rating Scale Model fit the data.

< Insert Table 5 about here >

Moving from the *item* level to the *test* level, we evaluated the test information curve and we observed that most of the test information lay between –4.0 and +2.0 logits on the person scale (see Figure 5). Accordingly, the standard error of measurement plot revealed the

lowest standard errors and therefore the highest precision within this range of logits. Moreover, as the person parameters become lower or higher, less precision and test information can be obtained from the Rating Scale Model. It is noteworthy that the maximum amount of test information and the minimum amount of standard error were met in the negative part of the logit scale ($\theta \approx -1.0$ logits). An explanation for this is that the majority of item location parameters were also negative (see Table 4). Still, as only few thresholds were outside the range of maximal test information, test sensitivity corresponded well with the location of item response categories.

< Insert Figure 5 about here >

As a final step, we tested the comparability of item location parameters across gender groups by checking for uniform and non-uniform DIF. In these analyses, a significant main effect of the factor *Gender* implies uniform DIF and a significant interaction effect between *Gender* and *Class Interval* indicates non-uniform DIF (Andrich et al., 2012; Tennant & Pallant, 2007). The latter refers to the interaction between pre-service teachers' membership to a manifest group (i.e., gender) and the intervals of the trait (i.e., class interval). In our results, there were neither significant main effects of *Gender* nor interaction effects (see Table 6). Hence, the reduced SQD-scale did not show uniform and non-uniform DIF. We can therefore conclude that the probability of responding to the SQD-items in a specific category is independent of pre-service teachers' gender.

< Insert Table 6 about here >

5. Discussion and Conclusion

The aim of the current study was to provide an instrument to measure pre-service teachers' perceptions of the extent to which they experience the necessary support and training in order to integrate technology into classroom activities. Specifically, we focused on

the strategies included in the inner circle of the SQD-model as depicted in Figure 1: 1) using

teacher educators as role models, 2) reflecting on the role of technology in education, 3)

learning how to use technology by design, 4) collaboration with peers, 5) scaffolding

authentic technology experiences, 6) Providing continuous feedback.

5.1 The SQD-scale

The results of the current study revealed that the newly developed scale was

unidimensional and indicated good fit to a Rating Scale Model for 22 items. Following the

principles of basic item response theory implied the rejection of 2 out of 24 items in the SQD-

scale that measure pre-service teachers' perceptions of the support and training given by TTIs.

The resulting scale showed that six response categories work well for the 22 items. Moreover,

the reduced scale exhibited highly satisfying psychometric properties, and no DIF across

gender emerged, meaning that the scale operates equally for both male and female pre-service

teachers.

The results also indicate that the items on strategies to prepare pre-service teachers differ

in their degree of 'difficulty'. Specifically, the items can be ordered from "easy" to "hard".

On the one hand, the data show that the items related to feedback appear to be the most

difficult. This is in line with the results in the study of Pareja Roblin et al. (2012) suggesting

that assessment and feedback with respect to educational technology use is one of the main

problems TTIs experience (see also Haydn & Barton, 2007: Mouza et al., 2014).

Nevertheless, previous research indicates that on-going and process-oriented feedback is

beneficial to build pre-service teachers' abilities to use technology in the classroom (Authors,

2012; Boulton, 2014).

Furthermore, items related to the design of ICT-rich lessons seem to be challenging for

TTIs. Other studies argue that additional support for pre-service teachers is needed to prepare

and implement lessons incorporating technology (e.g., Angeli & Valanides, 2009; Jang, 2008;

Sadaf, Newby, & Ertmer, 2012). According to Lee and Lee (2014), TTIs and schools could form partnerships to collaboratively work toward identifying the best methods to support pre-service in their lesson planning and practice related to technology integration. According to these authors, lesson planning is a good starting activity to bridge the gap between theory and practice (cf. Agyei & Voogt, 2014). In a next step, field experiences allow pre-service teachers to apply their knowledge of educational technology in authentic settings (Kafyulilo, Fisser, & Voogt, 2014; Valtonen et al., 2015).

On the other hand, the results suggest that the probability of a positive response is higher for the easiest item "I was encouraged to gain experience in using ICT in a classroom setting". As a consequence, helping pre-service teachers to design ICT-rich lessons and providing adequate feedback can be considered more challenging for TTIs compared to stimulating pre-service teachers to use technology (cf. Mouza et al., 2014; Pareja et al., 2012).

5.2 Limitations of the study and recommendations for future research

The current study aimed to develop an instrument to measure strategies TTIs could use without taking any potential determinants (e.g., Infrastructure, leadership, ICT policy planning) into consideration. In a next step, the SQD-scale can be used as a tool to examine the impact of TTIs on the pre-service teachers' competencies to integrate educational technology and/or technology attainment targets. A transparent understanding in the effect of these strategies on specific competency frameworks can result in adequate policy decisions or guidelines of educational authorities (cf. Vanderlinde et al., 2008). This instrument also encourages TTIs to reflect on their efforts to support future teachers in the digital age.

Moreover, the SQD-scale may provide additional information for intervention studies. In particular, this instrument can be used to examine how effective strategies are adopted in (online) communities of practice or teacher design teams, for example. Besides, the SQD-scale can also help examine institutional characteristics of TTIs as a critical predictor. As

stated earlier, this study is restricted to the perceptions of pre-service teachers about strategies at the micro level (i.e., the inner circle of the SQD-model). Nevertheless, key themes such as "aligning theory and practice" or "access to resources" can also operate at the micro level. It is clear that we need a holistic and dynamic approach to developing ICT-competence among future teachers within the institutional boundaries of TTIs. Since institutional characteristics have been excluded from the instrument, future research could examine how these key themes relate to the SQD-scale.

Moreover, future studies are needed to explore how teacher educators perceive the strategies included in the SQD-scale. By doing so, researchers could triangulate the views of various stakeholders (e.g., student teachers, teacher educators, ICT coordinators, and heads of departments). To illustrate, Authors (2012) conducted a multiple case study to examine the ways in which the strategies included in the SQD-scale were promoted in three TTIs. The findings indicate that each TTI adopted different strategies to prepare pre-service teachers for technology integration. At the same time the findings suggest that these efforts remain insufficient in two of the three TTIs. The pre-service teachers in these two TTIs felt that their pre-service education did not give them sufficient opportunities to engage in authentic tasks wherein they could apply their knowledge about technology to the design of concrete activities that could be useful in their later practice (cf. Lee & Lee, 2014). Only in one TTI pre-service teachers were able to provide concrete examples of learning opportunities with technology during their pre-service education.

In another study (Authors, in press), the strategies of the SQD-scale were used to collect in-depth interviews with beginning teachers. The results revealed that all beginning teachers acknowledged the importance of the six strategies but not all of them where addressed during their pre-service learning experiences. It appeared that teacher educators modeling technology use was an important motivator for beginning teachers to use

technology in their own teaching, but field experiences seemed to be the most critical factor influencing their educational use of technology.

5.3 Implications of the study

The SQD-scale provides an instrument TTIs could use to measure pre-service teachers' perceptions of the extent to which they experience the support and training needed to integrate technology in their educational practice (cf. Chien et al., 2012; Goktas et al., 2008; Kaufman, 2015), and specifically in view of the six strategies included in the new scale: 1) using teacher educators as role models, 2) reflecting on the role of ICT in education, 3) learning how to use technology by design, 4) scaffolding authentic technology experiences, 5) collaborating with peers, and 6) providing continuous feedback. In implementing such strategies, the SQD-scale could be used to measure the extent to which pre-service teachers perceive these efforts. These results can provide a good stepping stone to better prepare pre-service teachers for technology integration in classroom practices.

A clear example is related to the strategy "using teacher educators as role models". Several studies noted that there are insufficient technology-using teacher educators who are able to inspire pre-service teachers to use technology (e.g., Authors, 2012; Tearle & Golder, 2008). Gokstas et al. (2009) for instance pointed out that teacher educators often lack the knowledge, skills, and self-efficacy to teach technology integration to pre-service teachers. As stated above, the involvement of teacher educators in teacher design teams may constitute an effective strategy to develop competencies necessary to integrate digital resources into teaching and to structure the professionalization of teacher educators. A teacher design team can be described as a group of two or more teachers or teacher educators who (re-)design curriculum materials together (Handelzalts, 2009). In practice, this means that teacher educators consider together how technology can support the content and didactical aspects of the teacher education in order to prepare pre-service teachers for technology integration.

Subsequently, they may design technology-rich materials, test them in educational settings and reflect on the results (cf. Agyei & Voogt, 2014; Kafyulilo, Fisser, & Voogt, 2014). The results from the current study demonstrate that the items of SQD-model provide measures that allow researchers to evaluate such innovative projects.

The discussion on the ways in which pre-service teachers (and teacher educators) can be better prepared to integrate technology into their teaching practices should also be seen as part of the development of the entire teacher education programme (Authors, 2013; Lim & Pannen, 2012). Technology integration needs to be infused as a systemic and systematic process throughout the entire programme (Polly et al., 2010). Therefore, it is necessary for policy development at the institutional level to develop a shared vision on how to prepare future teachers for technology integration (cf. Lim & Pannen, 2012).

To conclude, many teacher training institutions have recognized the challenges associated with developing pre-service teachers' competencies for educational technology use and have proposed original, innovative strategies to expand these competencies (cf. Banas & York, 2014; Lee & Lee, 2014; Mouza et al., 2015). In implementing such strategies, the SQD-scale could be used to measure effective strategies adopted to prepare future teachers for educational technology use and to better structure the professionalization of teacher educators. As a result, this instrument can lead to valid and reliable benchmarking of the support future teachers need for the use of technology in education.

References

Aesaert, K., Van Nijlen, D., Vanderlinde, R., & van Braak, J. (2014). Direct measures of digital information processing and communication skills in primary education: using item response theory for the development and validation of an ICT competence scale. *Computers & Education, 76*, 168-181.

Adams, R. J., Wu, M. L., & Wilson, M. (2012). The Rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement*, *72*, 547–573.

Agyei, D. D., & Voogt, J. M. (2011). Exploring the potential of the will, skill, tool model in Ghana: Predicting prospective and practicing teachers' use of technology. *Computers & Education*, *56*, 91–100.

Agyei, D. D., & Voogt, J. M. (In Press). Pre-service mathematics teachers' learning and teaching of activity-based lessons supported with spreadsheets. *Technology, Pedagogy and Education*, 1-21.

Andrich, D. (1995). Models for measurement, precision, and the nondichotomization of graded responses. *Psychometrika, 60,* 7-26.

Andrich, D. (1978). A rating scale formulation for ordered response categories. *Psychometrika, 43*, 561–573.

Andrich, D., de Jong, J. H. A. L., & Sheridan, B. E. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. In J. Rost & R. Langeheine (Eds.), *Applications of Latent Trait and Latent Class Models in the Social Sciences* (pp. 59–70). Münster:Waxmann.

Andrich, D., & Hagquist C. (2001). *Taking Account of Differential Item Functioning through Principals on Equating*. Research Report no. 12. Perth, Australia: Social Measurement Laboratory, Murdoch University.

Andrich, D., & Kreiner S. (2010). Quantifying response dependence between two dichotomous items using the Rasch model. *Applied Psychological Measurement, 34*(3), 181-192.

Andrich, D., Sheridan, B. S., & Luo, G. (2012). *RUMM2030: Rasch Unidimensional Measurement Models* [Computer software]. RUMM Laboratory Perth, Western Australia.

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*, 154–168.

Authors (2012).

Authors (2013).

Authors (2014).

Authors (In Press).

Banas, J. R., & York, C. S. (2014). Authentic learning exercises as a means to influence preservice teachers' technology integration self-efficacy and intentions to integrate technology. Australasian Journal of Educational Technology, 30(6).

Bate, F. (2010). A bridge too far? Explaining beginning teachers' use of ICT in Australian schools. *Australasian Journal of Educational Technology*, *26*, 1042–1061.

Bland, J. M., & Altmann, D. G. (1995). Multiple significance tests: The Bonferroni method. *British Medical Journal*, *310*:170.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Dordrecht: Springer.

Boulton, H. (2014). ePortfolios beyond pre-service teacher education: a new dawn?. *European Journal of Teacher Education, 37*(3), 374-389.

Brun, M., & Hinostroza, J. E. (2014). Learning to become a teacher in the 21st century: ICT integration in Initial Teacher Education in Chile. *Educational Technology & Society*, *17*, 222–238.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and

    BIC in Model Selection. *Sociological Methods & Research*, *33*, 261–304

Chien, Y., Chang, C., Yeh, T., & Chang, K. (2012). Engaging pre-service science teachers to

    act as active designers of technology integration: A MAGDAIRE framework.

    Teaching and Teacher Education, *28*, 578–588.

Christensen, R., & Knezek, G. (2008). Self-report measures and findings for information

    technology attitudes and competencies. In J. Voogt, & G. Knezek (Eds.), *International*

    Handbook of Information Technology in Primary and Secondary Education (pp. 349–

    365). New York, NY: Springer.

de Ayala, R. J. (2013). The IRT tradition and its applications. In T. D. Little (Ed.), *The Oxford*

    Handbook of Quantitative Methods (vol. 1, pp. 144–169). New York, NY: Oxford

    University Press.

Drent, M., & Meelissen, M. (2008). Which factors obstruct or stimulate teacher educators to

    use ICT innovatively? *Computers & Education*, *51*, 187–199.

Enders, C. K. (2010). *Applied Missing Data Analysis*. New York, NY: Guilford.

Gao, P., Wong, A. F., Choy, D., & Wu, J. (2011). Beginning teachers' understanding

    performances of technology integration. *Asia Pacific Journal of Education*, *31*, 211–

    223.

Goktas, Y., Yildirim, S., & Yildirim, Z. (2009). Main barriers and possible enablers of ICTs

    integration into pre-service teacher education programs. *Educational Technology &*

    Society, *12*, 193–204.

Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing

    research: An introduction and illustrative example. *International Journal of Nursing*

    Studies, *46*, 380–393.

Haydn, T.A., & Barton, R. (2007). Common needs and different agendas: How trainee

    teachers make progress in their ability to use ICT in subject teaching. Some lessons

    from the UK. *Computers & Education, 49*, 1018-1036.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item

    response theory and their applications to test development. *Educational Measurement:*

    Issues and Practice, 12, 535–556.

Handelzalts, A. (2009). *Collaborative curriculum development in teacher design teams*.

    University of Twente.

Hyrkäs, K., Appelqvist-Schmidlechner, K., & Oksa, L. (2003). Validating an instrument for

    clinical supervision using an expert panel. International Journal of Nursing Studies,

    40(6), 619-625.

International Society for Technology in Education (ISTE) (2012). *ISTE Standards Teachers*.

    Retrieved from http://www.iste.org/docs/pdfs/20-14_ISTE_Standards-T_PDF.pdf

Ip, E. H. (2010). Empirically indistinguishable multidimensional IRT and locally dependent

    unidimensional item response models. *British Journal of Mathematical and Statistical*

    Psychology, *63*, 395–416.

Jang, S. J. (2008). The effects of integrating technology, observation and writing into a

    teacher education method course. *Computers & Education*, *50*, 853–865.

Kafyulilo, A., Fisser, P., & Voogt, J. (2014). Teacher design in teams as a professional

    development arrangement for developing technology integration knowledge and skills

    of science teachers in Tanzania. *Education and Information Technologies*, 1-18.

Kaufman, K. (2015). Information Communication Technology: Challenges & Some Prospects

    from Preservice Education to the Classroom. *Mid-Atlantic Education Review*, *2*, 1–11.

Kay, R. H. (2006). Evaluating strategies used to incorporate technology into preservice education: A review of the literature. *Journal of Research on Technology in Education*, *38*, 383–408.

Kennisnet (2012). *ICT proficiency of teachers (ICT-bekwaamheid van leraren)*. The Netherlands: Kennisnet.

Lee, Y., & Lee, J. (2014). Enhancing pre-service teachers' self-efficacy beliefs for technology integration through lesson planning practice. *Computers & Education, 73*, 121-128.

Lim, C. P., & Pannen, P. (2012). Building the capacity of Indonesian education universities for ICT in pre-service teacher education: A case study of a strategic planning exercise. *Australasian Journal of Educational Technology, 28*(6), 1061-1067.

Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, *2*, 266–283.

Marais, I., & Andrich, D. (2008). Formalising dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement*, *9*, 1–16.

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument Development in the Affective Domain* (3rd ed.). New York, NY: Springer.

Niess, M. L. (2005). Preparing teachers to teach science and mathematics with technology: Developing a technology pedagogical content knowledge. *Teaching and Teacher Education*, *21*, 509–523.

Mouza, C., Karchmer-Klein, R., Nandakumar, R., Ozden, S. Y., & Hu, L. (2014). Investigating the impact of an integrated approach to the development of preservice teachers' technological pedagogical content knowledge (TPACK). Computers & Education, 71, 206-221.

OECD (2010). *Inspired by technology, driven by pedagogy: A systemic approach to technology-based school innovations*. Paris: OECD Publishing.

Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research & Evaluation*, *17*(15).

Ottenbreit-Leftwich, A., Glazewski, K., Newby, T., & Ertmer, P. (2010). Teacher value beliefs associated with using technology: Addressing professional and student needs. *Computers & Education*, *55*, 1321–1335.

Pareja Roblin, N., Tondeur, J., van Braak, J., Voogt, J., & Fisser, P. (2012). *The journey from pre-service to practice: Exploring the connections between prior learning experiences and beginning teachers' uses of technology*. Paper presented at the European Conference on Education Research (ECER): Cadiz, Spain. September 2012.

Polly, D., Mims, C., Shepherd, C. E., & Inan, F. (2010). Evidence of impact: transforming teacher education with preparing tomorrow's teachers to teach with technology (PT3)grants. *Teaching and Teacher Education*, *26*, 863–870.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Achievement Tests*. Copenhagen: Danish Institute for Educational Research.

Revelle, W. (2014). *psych: Procedures for Psychological, Psychometric, and Personality Research*. R package version 1.5.1.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*, 145–154.

Sadaf, A., Newby, T. J., & Ertmer, P. A. (2012). Exploring pre-service teachers' beliefs about using Web 2.0 technologies in K-12 classroom. *Computers & Education*, *59*, 937–945.

Sang, G., Tondeur, J., Chai, C. S., & Dong, Y. (2015). Validation and profile of Chinese pre-service teachers' technological pedagogical content knowledge scale. *Asia-Pacific Journal of Teacher Education*, 1-17.

Sang, G., Valcke, M., van Braak, J., & Tondeur, J. (2010). Student teachers' thinking processes and ICT integration: predictors of prospective teaching behaviors with educational technology. *Computers & Education*, *54*, 103–112.

Smith, E. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, *3*, 205–231.

Tearle, P., & Golder, G. (2008). The use of ICT in the teaching and learning of physical education in compulsory education: how do we prepare the workforce of the future? *European Journal of Teacher Education*, *31*, 55–72.

Tennant, A., & Pallant, J. F. (2007). DIF matters: A practical approach to test if differential item functioning makes a difference. *Rasch Measurement Transactions*, *20*, 1082–1084.

Thomas, M. L. (2011). The value of item response theory in clinical assessment: A review. *Assessment*, *18*, 291–307.

Tømte, C., Enochsson, A. B., Buskqvist, U., & Kårstein, A. (2015). Educating online student teachers to master professional digital competence: The TPACK-framework goes online. *Computers & Education*, *84*, 26–35.

Thorpe, G. L., & Favia, A. (2012). *Data Analysis Using Item Response Theory Methodology: An Introduction to Selected Programs and Applications*. Retrieved from http://digitalcommons.library.umaine.edu/psy_facpub/20/

UNESCO (2008). *UNESCO's ICT Competency Standards for Teachers*. Retrieved from http://cst.unesco-ci.org/sites/projects/cst/default.aspx

Valtonen, T., Kukkonen, J., Kontkanen, S., Sormunen, K., Dillon, P., & Sointu, E. (2015). The impact of authentic learning experiences with ICT on pre-service teachers' intentions to use ICT for teaching and learning. *Computers & Education, 81*, 49-58.

Vanderlinde, R., Van Braak, J., & Tondeur, J. (2010). Using an online tool to support school-based ICT policy planning in primary education. *Journal of Computer Assisted Learning, 26*(5), 434-447.

Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Erlbaum.

Wright, B., & Masters, G. (1982). *Rating Scale Analysis: Rasch Measurement*. Chicago, IL: MESA Press.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependencies. *Journal of Educational Measurement*, *30*, 187–213.