Assessing Teacher Digital Competence: the Construction of an Instrument for Measuring the Knowledge of Pre-Service Teachers

José Luis Lázaro Cantabrana1, Mireia Usart Rodríguez2, Mercè Gisbert Cervera3
1Department de Pedagogia, University Rovira i Virgili, Spain [joseluis.lazaro@urv.cat]
2Department de Pedagogia, University Rovira i Virgili, Spain [mireia.usart@urv.cat]
3Department de Pedagogia, University Rovira i Virgili, Spain [merce.gisbert@urv.cat]

Received on 26 October 2018; revised on 13 November 2018; accepted on 28 November 2018; published on 15 January 2019
DOI: 10.7821/naer.2019.1.370

ABSTRACT
Assessing competences always poses a challenge and can be even more complicated when tackling a multidimensional competence like teacher digital competence (TDC). TDC is understood to consist of different dimensions linked to its components. This complexity gives rise to the need to organize and systematize both TDC training and its evaluation through a standard based on validated benchmark indicators. Designing and developing an instrument for TDC assessment has been a two-phase process. The COMDID-A self-assessment tool was developed in the first phase and COMDID-C, an instrument for assessing knowledge related to TDC, in the second. In this article we present the process of constructing the COMDID-C instrument. For this first stage, we worked with two samples, an expert validation and a pilot test sample. Due to the complexity of the test, we conducted a preliminary evaluation of the validity of its content, construction and reliability. Our results indicate that the test is well designed and consistent with its intended purpose. The next step will be administering the test to a larger sample that will allow the instrument to be externally validated.

KEYWORDS: TEACHER ASSESSMENT, TEACHER DIGITAL COMPETENCE, DIGITAL TECHNOLOGY, COMPETENCE ASSESSMENT, TEACHER TRAINING.

1 INTRODUCTION
The importance of training teachers in the digital environment and its close relationship with the quality of education in the 21st century has been made clear in the content of different reports published by international institutions. The European Commission (2007 and 2018) defines DC as one of the nine key competences citizens need in order to participate in today’s society. Specifically, the European Commission (2018, p. 5) specifies that “Digital competence involves the confident, critical and responsible use of, and engagement with, digital technologies for learning, at work, and for participation in society.” On this basis, we believe that today’s teachers should be able to train citizens to use digital technologies (DT) as a natural part of their daily lives.

In accordance with the objectives of the Education and Training 2020 Strategic Framework (European Union, 2009), teachers must link digital-age skills or competences to their professional practice. Various international and national frameworks of reference have been developed in this regard which define the digital competence of teachers, which we will refer to as teacher digital competence (TDC), and its components. Table 1 contains a summary of the main frameworks of reference that have been used most recently, which we will discuss below.

From the perspective of general frameworks of reference, we would like to highlight some aspects of those mentioned above:

- At the international level, the European Commission (Recker & Punie, 2017, p. 7) has published DigCompEdu, which specifies the digital competences that a teacher must possess in today’s society in order to effectively practice his or her profession. The proposal is designed to encourage reflection so that governments can develop their own frameworks of reference based on a common language and a shared point of departure.

- In Spain, there are two institutional frameworks of note that, in addition to contributing their own definitions of TDC, provide an evaluation rubric based on dimensions or areas, indicators and levels of skill development. The INTEF (2017) defines TDC as the set of competences that 21st century teachers must develop to improve the efficacy of their educational practice and for their own ongoing professional development.

- For the Government of Catalonia (2016, p. 2), TDC is “the ability of teachers to apply and transfer all their knowledge, strategies, skills and attitudes about learning and knowledge technologies (LKT) into real and concrete situations of their professional praxis in order to: a) facilitate student learning and the acquisition of digital competence; b) implement processes of improvement and innovation in teaching in accordance with the needs of the digital age; and c) contribute to their professional development in accordance with the processes of change that occur in society and in educational centers.”

For us, in keeping with that described above, TDC is made up of a set of capacities, abilities and attitudes that the teacher must develop in order to incorporate digital technologies into his or her professional practice and development.

*To whom correspondence should be addressed:
Facultat de Ciències de l’Educació i Psicologia
Campus Sescelades - Universitat Rovira i Virgili
Carretera de Valls, s/n
43007 - Tarragona, Spain
A thorough analysis of the content of these frameworks in terms of their evaluation indicators shows that there is also a close relationship between their indicators:

a) Correspondence between the teacher digital competence (COMDID) proposal and the Government of Catalonia proposal. Of the COMDID’s 22 descriptors, there are two that are not included in the Catalan government’s proposal. These refer to information handling and the creation of knowledge (from a didactic perspective) and to the use of a personal learning environment. In the Government of Catalonia’s proposal, there are two descriptors, both in the professional development dimension, that do not appear in the COMDID. These are related to reflective practice and participation in research. This comparison shows that the degree of coincidence between the two proposals is very high, as COMDID is a broad proposal that is virtually entirely included within the Catalan government’s work.

b) Correspondence between COMDID and DigCompEdu. Of the COMDID’s 22 descriptors, five of them were not identified in the European Commission’s document. These refer to the methodological approach of the institution, to learning environments, to the management of spaces with digital technologies at the school, to the participation in projects that involve digital technologies and to personal learning environments. On the other hand, a descriptor referring to reflective praxis has been included in DigCompEdu that does not appear in the COMDID proposal. So, COMDID has proven to be a broader proposal than that of the European Commission.

Based on this content analysis, we believe that the evaluation rubric created by Lázaro and Gisbert (2015) is an adequate
instrument for assessing TDC, in line with two governmental proposals: one international (DigCompEdu) and one national (Government of Catalonia). Some TDC assessments have used international frameworks for reference, such as DESECO or DigCompEdu, either exactly as they are laid out in the proposal (Sancho & Padilla, 2016) or in adapted versions (Gutiérrez & Serrano, 2016).

COMDID has been used as an evaluation rubric for TDC in other research and innovative experiments conducted by the ARGGET research group (Lázaro, Esteve, Gisbert, & Sanromán, 2016; Silva, Miranda, Gisbert, Morales, & Onetto, 2016) and a version has recently been adapted for the Latin American context (Lázaro, Gisbert, & Silva, 2018). Therefore, we consider that this rubric, in addition to having been validated, has a history that allows us to use it as a point of departure and a reference for the creation of the test.

3 DESIGN OF THE ASSESSMENT PROCESS

We started from the concept of assessment as a component of the teaching-learning process that must have a formative function, not only for the teacher but also for the student. We see evaluation as a process that should guide students in their learning and develop their capacity to self-regulate that learning (Carless, 2007).

At the time of the assessment, it is important to remember that teachers carry out their work in different areas: the classroom, the educational institution, the community and in the context of their own personal and professional development. The areas we cite should serve as scenarios in which teaching tasks and competencies have to be put into action and are therefore settings in which they can be evaluated (Lázaro & Gisbert, 2015, p. 34).

Thus, the assessment of students’ TDC was designed based on the COMDID rubric to be used at two times:

a) Final TDC assessment. Evaluation by means of a tool, specifically a CRT that measures students’ knowledge based on the components of TDC. The CRT in this case is used to assess the absolute status of the subject (student) with respect to mastery of a well-defined concept (teacher digital competence) and is useful for classifying students into one of the possible mutually exclusive categories as competent/not competent, in relation to a cut-off point established based on the judgment of experts in TDC. It is important to point out that, unlike the previous assessment, this test is not based on the students’ self-perception of their competence, but rather it objectively measures their capacities in certain situations inherent to the teaching profession.

b) After completing the training process, which lasted an entire academic year (12 ECTS credits) and required the students to participate in various activities oriented towards the development of TDC, the assessment process was undertaken, and the results obtained using the evaluation tool presented in this article. The tool was used to collect the data corresponding to the final evaluation of the process, which was a summative assessment.

3.1 Objectives

The review of the literature underscores the need for an assessment instrument that allows us to reliably and validly measure knowledge related to competences, specifically, to measure knowledge in samples of undergraduate students in primary education. Therefore, we set out the following specific objectives:

- Objective 1. To construct an instrument for the objective assessment of TDC knowledge in pre-service teachers.
- Objective 2. To establish a cut-off point and conduct a preliminary pilot study to lay the foundations for the subsequent external validation of the COMDID-C instrument.

4 METHOD

4.1 Sample

After the literature review and the COMDID-C question design processes were complete, the study was carried out in two phases: first the group of experts (sample 1) was contacted, and then a group of students (sample 2) participated in the pilot phase. The experts who participated in this study are researchers and teachers with experience in at least one of two areas: initial teacher training or TDC training. The minimum period of teaching experience among the expert sample was two years and the maximum was 30. Experience in TDC ranged from one to seven years. In total, six women and five men (hereinafter called the experts) from Rovira i Virgili University participated in the study. The youngest expert was 23 years old and the oldest was 56 (M = 39.4; SD = 12.34). They were all part of the COMDID-A assessment. Sample 2 in the pilot phase consisted of 25 students in the second year of the double degree in early child-
hood and primary education with a specialization in English at Rovira i Virgili University. The average age in this sample was 22.41 years (SD=2.71). There were, in total, three men (12%) and 22 women (88%) who participated in the validation of the COMDID-C CRT for the evaluation of digital competence.

**Design and tool**

A paper version of test was administered during the final session of the subject called Organization of the School Space, Materials and Teacher Skills. The test consisted of 88 questions divided between two parallel forms of the test that we sought to validate. These tests were divided into the four dimensions of TDC: D1. Didactic, curricular and methodological aspects; D2. Planning, organization and management of digital technological resources and spaces; D3. Relational aspects, ethics and security; D4. Personal and professional aspects. Half of the questions (44) correspond to “Test A” form (Table 2) of true or false (score 0 or 1) and the other 44 questions were included in “Test B” form (Table 3) of answers which range from the incorrect answer (to which we assign 0 points) to the completely correct one (1 point), all the questions are answered on a scale of five graduated options (0, 0.25, 0.5, 0.75 and 1). These two tests were designed in accordance with the self-assessment questionnaire COMDID-A (REFS), which has been validated and whose factorial structure is divided into the same four dimensions that this instrument evaluates (see Table 2).

After administering the two parallel forms of the test to a sample of students, they were asked about it and their proposals for improvement in terms of item formulation were taken into account, in accordance with the criteria of different experts who analyzed these comments in two work sessions. For example, in these sessions it was decided to eliminate the names of specific ICT tools to increase the validity of the content and the reliability of the COMDID over time. For instance, in the question linked to personal learning environments (PLE), the mention of the tool Symbaloo was deleted in an attempt to avoid making specific references, in this case, to a platform for compiling resources that allow one to configure a PLE.

Before we could analyze the reliability of the COMDID-C instrument, and because it is a CRT that measures the level of competence of the students, it was necessary to establish the cut-off point, in other words, the score above which the CRT could be considered passed and the subject competent in TDC. This point was determined by experts (judges), in the matter being assessed; however, the problem that arose was how to reconcile the different criteria used by each judge in finding a cut-off point. In this case, we used the Angoff method to address this issue.

The Angoff method (1971) is still the most used approach in practice, and it is particularly prevalent in education, and it has been adapted for use with different objectives (Cizek & Bunch, 2007). A group of assessors formulate their judgments based on a hypothetical group of students with minimal TDC. To every question on the test, the judges indicate the hypothetical students’ responses, and this yields a final value or minimum score, which is the cut-off point for the group to whom the test will ultimately be applied (Muñiz, 2010).

**4.2 Data analysis**

Measuring the cut-off point: In keeping with the Angoff method described above, the group of experts (class lecturers, team researchers and experts in TDC) determined the cut-off to be 70 percent for the specific group that comprises our sample. If the student of this group did not achieve the indicated score, they

<table>
<thead>
<tr>
<th>Table 3. Example of Test A question type</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dimension 4. Personal and professional</strong></td>
</tr>
<tr>
<td><strong>Descriptor 4.1. Personal Learning Environment.</strong></td>
</tr>
<tr>
<td><strong>Type</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4. Example of Test B question type</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dimension 4. Personal and professional</strong></td>
</tr>
<tr>
<td><strong>Descriptor 4.1. Personal Learning Environment.</strong></td>
</tr>
<tr>
<td><strong>Type</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
were classified as not competent. It is important to mention that this method can be applied both to the multiple-choice test (Test B), and to the dichotomous test (Test A). Therefore, we consider this result a good departure point for measuring the reliability of the two parallel COMDID-C forms.

The validity of the content of the COMDID-C items was addressed throughout the development process, as detailed throughout section 3.

For the reliability analysis, due to the design of the subject and in order to be able to measure TDC, the session had to be held at the end of the course, which made it impossible to administer the test twice to this sample. This is why we administered the parallel forms of the COMDID-C, as mentioned in the previous section. With this data we measured Cohen’s kappa (Cohen, 1960), the most commonly used coefficient of internal consistency for this type of test (Landis & Koch, 1977; Shoukri, Asyali, & Donner, 2004). At the same time, with the objective of complementing those results, we also analyzed the construct validity or internal reliability with the Livingston coefficient (Livingston, 1972) for both of the parallel forms (Test A and Test B). We will discuss the results in the next section.

Stability: The stability of this test over time (test-retest stability or reliability) will be studied with future student groups (Livingston & Lewis, 1995).

5 RESULTS

The quantitative analysis of the data was performed with SPSS software version 21.0 for Windows.

To calculate both Cohen’s kappa and the Livingston coefficient with a single application, we estimated the competent and not competent students in the sample.

Table 5. Contingency table for the student sample (N=25)

<table>
<thead>
<tr>
<th>Cut-off point = 70</th>
<th>Test B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competents</td>
<td>Not competent</td>
</tr>
<tr>
<td>Test A</td>
<td>Competents</td>
</tr>
<tr>
<td>Not Competents</td>
<td>0</td>
</tr>
<tr>
<td>Test B</td>
<td>19</td>
</tr>
</tbody>
</table>

The calculation of Cohen’s kappa yielded the result $k = -0.603$ ($p \leq .005$, $n = 25$ categorizations). In general, kappa coefficient values ranging between 0.6 and 0.8 are considered acceptable, while those over 0.8 are interpreted as very good (Landis & Koch, 1977). Therefore, our parallel versions of the test can be considered reliable.

The calculation of the Livingston coefficient with the data shown in Table 5 yielded the following results.

Table 6. Statistical data (mean, standard deviation and alpha) of tests A and B

<table>
<thead>
<tr>
<th>Cut-off point = 70</th>
<th>X</th>
<th>Sx</th>
<th>$r_{xx}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test A</td>
<td>7.6</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>Test B</td>
<td>7.4</td>
<td>0.7</td>
<td>0.32</td>
</tr>
</tbody>
</table>

For Test A, we obtained $k_2 = 0.472$ ($p \leq .005$, $n = 25$). For Test B, the Livingston coefficient is $k_2 = 0.723$ ($p \leq .005$, $n = 25$). Conceptually, $k_2$ can be interpreted as the percentage of cases that would be classified in the same category if the same test were taken again. As with traditional reliability coefficients, values equal to or exceeding 0.70 can be considered acceptable. On these grounds, we can confirm that both forms of the test are reliable according to the data, but that form A will require further study before its validity as a version of COMDID can be confirmed (see the following section for discussion).

6 DISCUSSION AND CONCLUSIONS

The objectives of this study were, first of all, to construct an instrument for the objective assessment of TDC knowledge in pre-service teachers. This instrument is called COMDID-C, because, as explained in the introduction, it is based on the COMDID rubric. The construction of this new tool has been approached as part of a comprehensive, mixed method (both qualitative and quantitative) process that considers all of the stakeholders for the study of TDC.

Competence involves putting into action conceptual knowledge, procedural knowledge and attitudes to be able to resolve a particular situation (OECD, 2011; Perrenoud, 2005). Thus, we are aware that competency assessment is a complex process and that it must be approached from a broad perspective and through the use of different assessment techniques and instruments (Palliser, Fullana, Planas, & Valle, 2010). With this in mind, we have addressed the limitations of using a single assessment tool to measure capacities related to a complex competence like TDC. Before we can continue to advance the reliability process for this instrument, we believe a study with a larger sample must be conducted in order to collect evidence of the general behavior of the tool and to continue with the process of its development, studying in particular its external validity.

With regard to the second objective, establishing the cut-off point and carrying out a preliminary pilot study to establish the bases for the subsequent external validation of the COMDID-C instrument, the kappa value indicates that the two tests are parallel, in other words, that we can administer either one indistinguishably. In addition, the response time is long for those 88 questions, so we believe that we should apply 44 from Test A or B. Furthermore, we need a larger sample in order for the Livingston coefficient to yield better results and ensure validity, with a minimum of 44 students to increase the value of Cohen’s kappa and report greater reliability for the COMDID.

Specifically, and from the results of the psychometric analysis, it has become clear that, although reliable, version A of the COMDID (with dichotomous responses) should continue to be studied in larger samples. Within the constraints of this first quantitative study, and in keeping with Clark and Watson (1995), we can affirm that a multiple-choice tests offers better statistical and reliability results.

Similarly, we stress the need to design the assessment ensuring coherency between what is being evaluated and the procedure used for it (Villa & Poblete, 2011, p. 150). In this regard, we believe that the test we have used, administered at the end of the formative process for a summative purpose, offers reliable data for measuring the knowledge acquired by the students and the development of their TDC.

In the near future we suggest the test should be administered by means of a technological solution that allows the student to obtain immediate feedback and recommendations on how they can improve their skills. Thus, the evaluation would be more oriented towards learning, fostering the student’s capacity for self-regulation (Gil-Flores, 2012, p. 135).
FUNDING

Research funded by AGAU. ADEEDIMP project: Evaluación i certificació de la competència digital docent en la formació inicial de mestres: una proposta de model per al sistema universitari català (Ref. 2017ARMIF00031).

REFERENCES


OCDE. (2011). Informe habilidades y competencias del siglo XXI para los aprendices del nuevo milenio en los países de la OCDE.


