*Article*

# Design and Validation of a Novel Tool to Assess Citizens' Netiquette and Information and Data Literacy Using Interactive Simulations

**Juan Bartolomé [1,*]** and **Pablo Garaizar [2]**

1    TECNALIA, Basque Research and Technology Alliance (BRTA), 48160 Derio Bizkaia, Spain
2    Engineering Faculty, University of Deusto, 48007 Bilbao, Spain; garaizar@deusto.es
*    Correspondence: juan.bartolome@tecnalia.com; Tel.: +34-664101155

**Abstract:** Until recently, most of the digital literacy frameworks have been based on assessment frameworks used by commercial entities. The release of the DigComp framework has allowed the development of tailored implementations for the evaluation of digital competence. However, the majority of these digital literacy frameworks are based on self-assessments, measuring only low-order cognitive skills. This paper reports on a study to develop and validate an assessment instrument, including interactive simulations to assess citizens' digital competence. These formats are particularly important for the evaluation of complex cognitive constructs such as digital competence. Additionally, we selected two different approaches for designing the tests based on their scope, at the competence or competence area level. Their overall and dimensional validity and reliability were analysed. We summarise the issues addressed in each phase and key points to consider in new implementations. For both approaches, items present satisfactory difficulty and discrimination indicators. Validity was ensured through expert validation, and the Rasch analysis revealed good EAP/PV reliabilities. Therefore, the tests have sound psychometric properties that make them reliable and valid instruments for measuring digital competence. This paper contributes to an increasing number of tools designed to evaluate digital competence and highlights the necessity of measuring higher-order cognitive skills.

**Keywords:** digital competence; computer-based assessment; netiquette; information and data literacy; simulations

## 1. Introduction

In a rapidly evolving world, digital representation of information and its communication through digital technologies have transformed our daily life with severe consequences in terms of sustainability in society. Citizens must face demands of different natures of the digital world [1]. The current society presents a new scenario that demands new perspectives for cyber connection and user empowerment. The United Nations Sustainable Development Goals (SDGs) have identified main challenges and their respective 17 goals divided in 169 targets. Moreover, none of the goals and targets are disconnected from the potential and effects of digital technology. Ensuring access to technology is not sufficient. For achieving the SDGs, it is essential to empower people with the right capabilities to use technology meaningfully to participate in today's society [2]. According to the literature review and the consultation with experts and policy officers at European and international levels carried out by Ala-Mutka [3], the acquisition of digital competence (DC) is considered to be as relevant as the other key competences towards a sustainable society, and many of them have tried to define which DC each citizen should have, as it provides important benefits in today's society. Matching the skills of citizens to the requirements of the demand for employment has been identified as a key factor in sustainable development for the future workforce [4]. It is crucial to reduce the digital divide, which is closely related with

the economic, social and cultural conditions of citizens and impedes sustainable development [5,6]. Furthermore, over the past years, various definitions have been provided about what DC is. This variety may be due to the fact that DC is a context-dependent definition [7,8]. In a government policies context, Ferrari [7] defined DC as "*the set of knowledge, skills, attitudes, strategies and awareness which are required when ICT and digital media are used to perform tasks, resolve problems, communicate, manage information, collaborate, create and share content, and build knowledge in an effective, efficient, and adequate way in a critical, creative, autonomous, flexible, ethical, and a sensible form for work, entertainment, participation, learning, socialization, consumption, and empowerment*". Consequently, DC is critical for empowering citizens to live in a society where they are consumers and creators of digital technology in critical, creative, autonomous and ethical ways, which are also essential for sustainable development [9]. The European Commission launched the Digital Competence Framework for citizens (DigComp) [10], and its updates in 2016 and 2017, with the aim of boosting the development of DC in Europe.

In this context, assessing DC has become a topic of growing interest in recent years and relevant studies have examined the principal advances and limitations [8,11–14]. However, despite employing different approaches, some issues still require further study; most of the assessment systems consist of self-assessments, do not cover the three components of DC (knowledge, skills and attitudes) and measure mainly low-order cognitive skills (according to the Bloom's taxonomy, the lower-order cognitive skills include remembering, understanding and applying, while the higher-order cognitive skills include analysing, evaluating and creating). In addition to these shortcomings, until recently, a majority of the digital literacy frameworks were based on assessment frameworks from commercial enterprises [8]. Consequently, the selection of DCs taught and assessed was influenced by the framework chosen, based mainly on commercial applications such as Microsoft's Office Suite and operating system. The launch of DigComp in 2013 facilitated the development of tailored implementations, providing a reference framework to work on DC [12]. However, most of the implementations related to competence assessment are self-reports compounded by multiple-choice items and Likert scales, only measuring low-order cognitive skills (e.g., IKANOS, probably the best known self-diagnostic tool at a European level available on http://test.ikanos.eus/ (accessed on 3 March 2022)). Furthermore, the skill component of the DC is barely evaluated, probably because the development of simulations or task-based assessments is complicated and time consuming.

Technology enhanced assessment (TEA) provides innovative and authentic item formats, such as interactive simulations with the true-to-life settings necessary for assessing skills such as communication or collaboration [15–18], and opportunities to carry out the evaluation in a safe setting such as in simulated environments. In recent years, a growing number of studies have examined and found a direct correlation between users' performance and engagement and how the evaluation items are designed [19–21]. Therefore, the design of the evaluation items is critical so that they require displaying the expected knowledge and skills. This fact is specifically challenging in the assessment of complex cognitive constructs. Test designers tend to use different dynamic formats such as interactive simulations. According to Heer [22], the cognitive domain is described as the combination of the cognitive process dimension (which has six categories: remembering, understanding, applying, analysing, evaluating and creating) and the knowledge dimension (which has four categories: factual, conceptual, procedural and metacognitive). Each dimension has its own categories and cognitive processes, and obviously, the same item format may not be suitable to gather evidence from different processes and levels of proficiency. So, assessments commonly include different types of questions to allow such inferences. However, the great effort required for developing specific types of questions, such as dynamic formats, limits the extent of inferences.

The assessment frameworks identified for accreditation purposes are used in workforce contexts to ensure that employees have the DCs required to perform in the workplace and can be categorised as commercial enterprises (mainly the European computer driving

license (ECDL), also known as ICDL, and Certiport's IC3) [8] or custom implementations based on DigComp [12]. Regarding the assessment instruments identified by Law et al. [8], despite the inclusion of different item formats such as interactive simulations, the assessments prioritise the technology itself, instead of the use of different applications to solve certain tasks. In addition, the tests have been designed based on licensed private software, i.e., the items have been designed showing tasks based on licensed private software, not representing many organisations that use other software, e.g., Google Workspace. Even more, the situations presented in their assessments to be solved by the examinees mostly represent computer-based tasks. However, this fact does not correspond to reality, where in 2019 more than 90% of young people used mobile devices to access the Internet and 52% used a portable computer [23]. Therefore, mobile devices should be included in the assessment framework or could lead to a relevant limitation. Finally, assessments for accreditation purposes tend to be reliable to a certain extent, and its construct and internal validity are normally weaker than the research-oriented ones.

This study is a custom implementation based on the latest version, DigComp 2.1, and is closely related to BAIT, the DCs certification service of the Basque Government [24], which is also a custom implementation based on DigComp. Credentialing-focused assessments are required to be conceived for use in a scalable and continuous manner in safe settings. So, this type of assessment should be designed with focus on general and practical technical abilities (e.g., use of Office software), to facilitate their use in different areas and as long as possible.

On this basis, we developed an assessment tool to assess citizens' DC selecting two case studies: information and data literacy (IDL) and netiquette. These case studies represent two different approaches that are currently being adopted by some relevant initiatives identified as successful cases [12], tests based on competence area or tests based on DCs. For the selection of the competence area, we chose one of the three main competence areas according to the DigComp framework, IDL, but other competence areas such as communication and collaboration could have been chosen due to their relevance. To select a DC, we were in the same situation and chose a competence that is not usually assessed in depth, where only low cognitive skills are usually assessed. In addition, we selected netiquette because the results of including dynamic item formats would be very noticeable. In other DCs, the improvement could be at different levels. We applied a design-based research (DBR) methodology that was based on the analysis of different sources of information to carry out the development of the evaluation tool and its later validation. In summary, we sought to achieve the following objectives:

- To design a tool for the assessment of DC that supports dynamic formats such as interactive simulations, which are particularly relevant when measuring complex cognitive constructs such as DC in safe settings.
- To describe the design principles applied during the different steps of the development of the tests for evaluating the DCs selected, with the aim that they can be extended to the rest of the DCs included in the reference framework.

Additionally, we address the following research questions:

- Is it possible to assess IDL through a DBR-designed test using simulations?
- Is it possible to assess netiquette through a DBR-designed test using simulations?

This manuscript is structured as follows. In Section 1, first we present the reference framework for the evaluation of DC. Second, we review recent evidence related to the selected case studies: IDL and netiquette. Third, we introduce item response theory (IRT), which has been extensively used in the field of education in test construction, basically as a measure of latent traits, and provides ways of assessing the properties of a measurement instrument in terms of reliability and validity. Section 2 explains the methodology applied, whereas Section 3 presents the results. At the end, we conclude our manuscript and introduce the directions for future work.

### 1.1. Reference Framework for the Evaluation of DC

In 2013, the Institute for Prospective Technological Studies (IPTS) from the European Commission's Joint Research Centre launched the Digital Competence Framework (Dig-Comp), integrating existing conceptualisations of DC [9]. This framework arranged the dimensions of DC in five competence areas: information and data literacy, communication and collaboration, digital content creation, safety and problem solving. In total, 21 DCs are distributed in the five competence areas. In 2016, DigComp 2.0 was published [25] and updated the terminology, concepts and descriptors of DCs. In 2017, DigComp 2.1 was released [26] and applied significant changes, such as increasing the initial three proficiency levels to eight, making use of Bloom's taxonomy in the definition of the descriptors of DCs.

DigComp is a reference framework structured in five dimensions: (1) competence areas involving the different DCs, (2) descriptors for each DC, (3) proficiency levels at the DC level, (4) knowledge, skills and attitudes expected in each DC and (5) different purposes of applicability. We used the DigComp framework as the reference framework due to its remarkable strengths: (1) it was designed after a deep analysis of the available DC frameworks, (2) it followed a meticulous process of consultation and development by experts in the area of DC and (3) as a result, it provides a comprehensive view based on DCs and competence areas. For similar reasons, the United Nations Educational, Scientific and Cultural Organization (UNESCO) also selected DigComp as the reference DC framework for the development of the Digital Literacy Global Framework (DLGF) [8,27]. Even more, the World Bank also identified the DigComp framework, in a recent report, as one of the most comprehensive and widely used frameworks for general DC [28].

DigComp describes DC regardless of the technologies and devices employed. Nevertheless, common software tools tend to provide similar functions despite the fact that the interface design may vary [29]. Moreover, findings from recent studies also questioned that DC is independent of the task context and the technology used, since in some specific fields, concrete digital technologies or handling specific digital technologies may be a relevant DC [8].

Based on the data collection approach, three major categories were identified in the custom implementations based on DigComp [26]: (1) performance-based assessment, where examinees have to solve tasks that are usually expected to be faced in a real-life context by using simulations or typical applications such as Office suites, (2) knowledge-based assessment, where the declarative and procedural knowledge of examinees is measured and (3) self-assessment, which is mainly based on Likert scales and examinees self-evaluate their level of knowledge and skills. Other authors such as Sparks et al. [30] illustrated different designs of assessment instruments according to their purposes: research purposes, accreditation, institutional quality assurance, self-assessment to support professional development, etc. Therefore, considering that the context of the present study is accreditation, we chose a performance-based assessment approach for the design of our instrument, and aspects of reliability and validity of the instrument were considered from the beginning. Additionally, considering that the target group of citizenship can be very diverse, we considered usability aspects during the design of the tool.

Regarding the types of items selected in the design of evaluation instruments, test designers tend to use constrained response item formats. Their implementation is simple and facilitates the automatic correction. However, these formats are not the most suitable for assessing higher-order skills. To assess higher-order skills according to the intermedium and advanced levels of DigComp, more sophisticated formats are necessary, such as purpose-built games or interactive simulations, to ensure an effective evaluation of DC [31]. Furthermore, despite the study carried out by Heer [22] to select different item formats to meet the assessment purposes, empirical evidence choosing the most suitable item types according to the assessment objectives is scarce.

Finally, the multidimensionality of the DC construct has been identified in several studies. For example, DC has been theoretically structured in five competence areas in the DigComp framework [25]. However, theoretical and empirical studies have reported

contradicting results. For example, Reichert et al. analysed the most commonly used digital literacy frameworks and found that empirical evidence on the use of digital applications allows for distinguishing between a general digital literacy component and four application-specific components (the web-based information retrieval factor, the knowledge-based information retrieval factor, the word processing factor and the digital presentation factor) [32]. Jin et al. found in their custom implementation based on DigComp that DC can be considered as a general one-dimensional construct [33]. In the systematic literature review conducted by Siddiq et al. [11], in most studies where dimensionality was checked, it was concluded that DC is a unidimensional construct, i.e., the construct has a unique underlying dimension that can be measured using a single measure of the test. Although further studies have continued in the same line, e.g., [34–36], the need for further research was also suggested. In addition to frameworks of DC, various national and international assessment studies were conceived based on a multidimensional framework, e.g., the International Computer and Information Literacy Study (ICILS). However, the empirical results presented differences in terms of quantity of dimensions and categories identified in the dimensions.

### 1.2. Information and Data Literacy

According to DigComp, IDL is one of the competence areas composed of three DCs (see Table 1) [26].

**Table 1.** IDL area and DCs as defined in DigComp [26].

| Digital Competence | Description |
| --- | --- |
| Browsing, searching and filtering data, information and digital content | To articulate information needs, to search for data, information and content in digital environments, to access and navigate between them. To create and update personal search strategies. |
| Evaluating data, information and digital content | To analyse, compare and critically evaluate the credibility and reliability of sources of data, information and digital content. To analyse, interpret and critically evaluate the data, information and digital content. |
| Managing data, information and digital content | To organise, store and retrieve data, information and content in digital environments. To organise and process them in a structured environment. |

This area is also known as information literacy or digital information literacy, and it is constantly changing due to the recurrent changes in how citizens access and manage information through different types of devices. Citizens, and especially youth, are replacing traditional media with social networks, which are currently one of the means most used, but at the same time supposes an ungoverned source of information that tends to create confusion, generate controversies and distrust [37–39], enable users to be active content creators [40] and influence young people in their choice of role models [41]. Moreover, the ease and speed of propagation that social networks facilitate for disinformation has become one of the most dangerous threats [42–45], in conjunction with the emergence of discourses based on emotional appeal to influence choice by making use of different ways such as click baiting, algorithms based on artificial intelligence, creation of filtered bubbles, personalisation of information, etc. [44–46]. The 2019 Eurobarometer already showed an increase in concern over issues such as the rapid growth of fake news (74%) and towards social media (65%) [47]. IDL has been identified as a key literacy to identify fake news [48]. So, in this context, it is necessary to examine and assess how citizens perceive and evaluate the media in terms of fake news.

There are many self-reports where individuals must self-assess their level, and most of them are tools composed of multiple-choice questions measuring low-order cognitive skills, e.g., [11,12,49–51]. In addition, assessments cannot be carried out using simple self-assessment tests. They offer a solution that is easy to implement but tends to obtain unrealistic results from examinees, caused by their overconfidence, especially examinees with very low ability [52]. The existence of the Dunning–Kruger Effect has also been proven to exist in the IDL area [53]. There are also some exceptions, e.g., using open-ended tasks with scoring rubrics [54,55], but these alternatives would be very complicated to integrate in a certification context requiring safe settings.

From the point of view of operationalising the construct of IDL for assessment purposes, Sparks et al. [30] indicated that test designers appear to take two possible approaches: (1) selecting a particular framework aligned with the construct defined in their implementation and then designing items according to the descriptors of the framework (this option is suitable for assessing a specific set of skills) or (2) operationalising the construct at a conceptual level, thereby developing authentic tasks that evaluate ECDL in a broader way. This option is suitable for defining the construct more holistically and examining whether examinees can put their knowledge into action in a real context. Consequently, the intended learning objectives and the type of assessment foreseen should be clarified from the beginning. Even more, beyond a specific construct definition, in the development of the assessment other issues should be considered, e.g., the contexts where information is going to be accessed, evaluated and used or whether a specific technology is an assessment target in itself or constitutes a way for achieving an objective.

Regarding the implementation of the assessment tools, Sparks et al. [30] categorised different types of assessment as: (1) consisting of constructed response questions focused on IDL, such as the International Computer and Information Literacy Study (ICILS), (2) consisting of constructed response questions focused on technology literacy, such as ECDL and (3) consisting of performance-based tasks focused on IDL, such as the Interactive Skills Assessment Tool (iSkills, Mount Maunganui, New Zealand).

IDL assessment in higher education is a key issue too [30,56], and interest in developing instruments to assess IDL has been growing in recent years. However, most of the tests are developed from two perspectives, librarian and academic, and are often domain specific [57,58].

With regard to the validation of the quality of the assessment instruments, the classical test theory was applied in most of the tests identified, and the most commonly performed analyses were content and discriminant validity and internal consistency reliability [59]. Therefore, experts argued the need to have free available assessment instruments for measuring IDL, performing a more effective assessment, validated and independent of the domain and the context [57].

*1.3. Netiquette*

According to DigComp [10], netiquette is one of the six DCs defined in the communication and collaboration competence area, defined as: "*To be aware of behavioural norms and know-how while using digital technologies and interacting in digital environments. To adapt communication strategies to the specific audience and to be aware of cultural and generational diversity in digital environments*".

In our present society, where ICTs are present into most areas and social networks and the extensive use of mobile devices have radically modified the way of interacting among people, netiquette is becoming a crucial DC [60]. Thus, a new scenery emerges for understanding human relations, from how interpersonal skills are exercised online to how social behaviours are exhibited in groups and online communities [61]. Cabezas-González et al. found that individuals who communicate online frequently and make use of social networks very frequently tend to show lower levels of DC, contrary to expectations [62]. So, it is of great importance to investigate the current education of individuals in communication and collaboration and in netiquette too [63]. Nevertheless, netiquette has been

barely defined and still does not seem to have attracted the required attention [64]. Only a few studies have analysed the guidelines related to the correct use of electronic mail, e.g., [65,66], or presented general guidelines for the Internet, e.g., [67]. No studies have attempted to define which DCs a citizen should have to communicate efficiently through everyday tools such as instant messaging applications, social networks or email. So, it is necessary to review the theoretical background and analyse the experimental proposals.

Regarding the implementation of the assessment tools, the empirical articles identified included the development of tailored tests, e.g., [68,69], whose validity and reliability evidence are insufficient; netiquette DC has not yet been assessed in depth and most of them include only a few general questions [8,12]. From a broader point of view, experts have identified the lack of instruments for evaluating individuals' DC in the communication and collaboration competence area [11]. Only BAIT, closely related to this study, provides a test exclusively dedicated to the assessment of netiquette [24].

*1.4. Item Response Theory (IRT)*

IRT, also referred to as item characteristic curve theory, attempts to give a probabilistic foundation to the problem of measuring unobservable traits and constructs, or latent traits [70,71]. IRT is widely used to calibrate and evaluate items in assessment instruments and provide ways of assessing the properties of a measurement instrument in terms of reliability and validity [34,70,72–75]. The main features and advantages that characterise IRT are: (1) the existence of latent traits that can explain an examinee's behaviour in a test, (2) the relationship between performance and the set of traits assessed, (3) the dimensionality specification, (4) the position of the item in the trait's value set, (5) assessment instruments with properties that do not depend on the specific group of respondents or the specific set of items showed, as both items and examinees receive a score on the same scale at the same time, (6) in contrast to classical test theory (CTT), basic units of analysis are based on the items and not on the assessment instrument and (7) the reliability of an assessment instrument is dependent on the action between the examinee and the assessment instrument.

Considering the strengths of the theory, we used the Rasch measurement model to investigate the reliability and validity of the tests developed in our study. The Rasch measurement model is the simplest model available within an IRT context and facilitates interpretation assuming that the response of the examinees to an item only depends on their proficiency and the difficulty of the item [74]. Furthermore, in IRT, the internal validity of a test is evaluated in terms of the fit of the items to the model. Marginal maximum likelihood (ML) is the most commonly used method in the estimation of the models in IRT and presumes that the parameters of an individual are aleatory variables with a certain distribution [75].

Applying the most suitable IRT model firstly relies on the characteristics of the items used, dichotomous or polytomous. For dichotomous items, such as the ones used in our tests, the most used models are the logistic models with one, two or three parameters. The parameters to characterise the items include [76]: their difficulty (situating the item on the ability scale, which states the probability of being answered correctly), discrimination (representing the degree of variation in the success rate of individuals as a function of their ability) and a pseudo guessing parameter (representing the lower asymptote where even less-capable individuals will score by guessing). IRT is based on the principle that it is possible to measure latent traits, i.e., traits which are not directly perceptible. Some items can comprise a specific trait (e.g., competences for evaluating the information) [71].

In addition, measures constructed using the Rasch measurement model are unidimensional and have expected structures of item calibrations that cover the difficulty range of difficulty within a domain in an assessment domain. The results are valid only to the extent that the dimensions are different and clear, i.e., there are no items assessing different variables at the same time; therefore, the unidimensionality assumption is realistic. Hence, other models such as the multidimensional IRT models appeared, which consider a con-

struct consisting of various factors. The multidimensional random coefficient multinomial logit (MRCML) model was presented as an alternative to confirmatory factor analysis (CFA) [77]. CFA and multidimensional IRT are methods applied to validate a possible organisation of the information. The multidimensional Rasch model is the simplest of MIRT models, which assumes that all item loadings are set to unity with the Rasch model [78]. In our analysis, we used the software ConQuest to analyse the difficulty of the items and the covariance across dimensions, as the software package is based on the MRCML model [79].

Finally, relatively small sample sizes could be sufficient for Rach analysis, and about 200 examinees suffice for obtaining accurate parameter estimates [80].

## 2. Materials and Methods

We applied a DBR methodology to develop and validate the assessment tool. DBR is a widely used methodology in the learning sciences to analyse the development of solutions [81,82]. We combined different methods during the iterative design process of the assessment instrument for two implementations [83]. According to Reeves, each cycle consists of different phases [84] (see Figure 1).
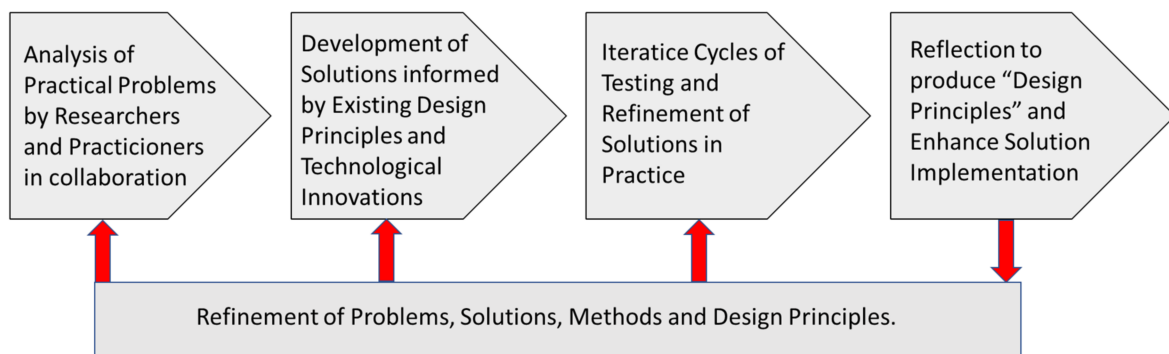


**Figure 1.** DBR approach in technology research [84].

The specific approach that we followed was guided by the different elements, guidelines and considerations suggested [82] for each phase of DBR proposed by Reeves [84] (see Table 2).

**Table 2.** Phases of the DBR methodology adapted to our research proposal (table based on [82]).

| Phase | Element |
|---|---|
| PHASE 1: analysis of the problem by researchers and practitioners in collaboration | Statement of problem<br>Consultation with researchers and practitioners<br>Research questions<br>Literature review |
| PHASE 2: development of theoretical framework solutions based on existing design principles and technological innovations | Theoretical framework<br>Development of draft principles to guide the design of the solution<br>Description of proposed solution |
| PHASE 3: iterative cycles of testing and refinement of the solution in practice | Implementation of intervention (first iteration with digital competence centre facilitators and second iteration with citizens)<br>Participants<br>Data collection<br>Data analysis |
| PHASE 4: reflection to produce "design principles" and enhance solution implementation | Design principles<br>Designed artefact |

In this paper, we describe our specific approach and results in each phase, including two iterative cycles in phase 3. We present the findings of the two interactive cycles in the

Section 3. We carried out the different phases in order, even though some of them could be managed simultaneously.

### 2.1. Phase 1: Analysis of the Problem by Researchers and Practitioners in Collaboration

At the start of phase 1, we stated the problem. To comprehend the problem, we examined the available solutions by carrying out a literature review, as has been shown in the introduction of this manuscript. We identified a lack of suitable instruments for measuring individuals' DC, measuring not only low-order cognitive skills. It is necessary to include innovative and authentic item formats in the tests to assess higher-order skills according to the medium and advance levels defined in DigComp framework. Otherwise, we are mainly assessing the knowledge component of DC. The purpose of this study is guided by this problem identified, and the aim is the development and validation of a potential solution. Furthermore, as stated in the introduction of the manuscript, various studies mentioned were based on consultation with researchers and practitioners. After the review of current knowledge and practice, we defined our objectives based on the design principles identified in several key studies. In parallel, the authors of the manuscript have been participating in The DigComp Community of Practice (DigComp CoP), which was launched in late 2019 by All Digital to promote the adoption and support the development of DigComp framework [85]. We have been participating in working groups, exchanging material and experience, accessing good practices, learning from peers and being informed about the latest developments concerning DigComp.

### 2.2. Phase 2: Development of Theoretical Framework Solutions Based on Existing Design Principles and Technological Innovations

The design of an intervention based on a detailed understanding of the problem is guided by design principles that are prescriptive theoretical arguments. [83]. Therefore, we defined an initial solution based on key studies identified [24,86,87] and from other key studies identified in the literature review, such as [8,9,11,32]. Once the new design proposition was established, it was necessary to examine and improve it after the testing and analysis [83].

For the design of the assessment instrument, we took DigComp 2.1 as the reference framework. DigComp offers a clear view of the different components of DC (knowledge, skills and attitudes) when using digital devices and services, which are needed to achieve a full participation in our society and can be adapted to many areas of life. Specifically, we focused on "*enhancing employability*" as the application scenario, since this study is closely related to BAIT [24]. Initially, we selected 4 DCs as case studies and the first six levels (foundation, intermediate and advanced) to be assessed. It can be considered that these are the most commonly demanded DC levels citizens for their employment. We also considered each DC as an independent construct and developed one test for each of them in order to be measured independently. From the literature review, we identified a series of sub-competences to be included for each test (see Table 3). The descriptors defined for each DC, sub-competences and corresponding levels can be examined in Table S1 in the Supplementary Material. Based on these descriptors, we developed the assessment items.

The items were distributed in each sub-competence non-uniformly, i.e., we considered that some sub-competences required more items to be measured correctly. While the DCs described in DigComp may appear stable in the short term, in the current context where technology is continually causing profound habit changes, the construction of DC requires a constant revision [11,34]. For example, in the selection of sub-competences for the netiquette DC, we did not include anything related to its application in video conferencing. Months later, due to the pandemic, these issues became relevant to this area due to an unprecedented rise in the use of this type of tool.

**Table 3.** Case studies of DCs based on DigComp [26] and sub-competences selected.

| Competence Area | Digital Competence | Sub-Competence |
|---|---|---|
| Communication and collaboration | Netiquette | Sub-competence1: apply basic netiquette guidelines when using email (e.g., use of blind carbon copy (BCC), forward an email/content, etc.). Sub-competence2: apply simple online writing rules (no capital letters, respect the spelling, referring to others by their aliases or nicknames, etc.) and use emoticons appropriately when communicating online. Sub-competence3: recognise appropriate behaviours on social networks, such as receiving permission from others before publishing (especially when children are involved); avoiding SPAM (e.g., sending invitations or other messages to everyone); using words or a non-clear language that may be misunderstood. Sub-competence4: Recognise inappropriate online behaviour, such as stalking, trolling or cyber bullying. Able to deal with negative behaviours such as flagging disrespectful publications or notifying the police. |
| Information and data literacy | Browsing, searching and filtering data, information and digital content | Sub-competence5: analyse information needs, search for data and information in digital environments, filter and locate. Sub-competence6: define the search strategy required at any given moment. |
| Information and data literacy | Evaluating data, information and digital content | Sub-competence7: examine and evaluate the credibility and reliability of sources of data and information. Sub-competence8: examine and evaluate digital content, data and information. |
| Information and data literacy | Managing data, information and digital content | Sub-competence9: organise, store and process data, information and content in digital environments. |

With respect to the three components of DC, we opted to exclude the evaluation of the attitudinal component from the scope of our study. In fact, the attitudinal component is complex and there is a lack of consensus on how to evaluate it, and even more, is not going to be directly assessed in BAIT [24]. Then, according to the analysis performed during the literature review, we selected implementing a performance-based approach, where individuals are monitored in a computer-based assessment (CBA) setting. We designed the items to see whether examinees were able to understand any digital environment in an effective way instead of evaluating their knowledge about specific applications. They have to put their knowledge into action and higher-order skills can be triggered and measured. This way of measuring obtains the most realistic situation of the individuals' proficiency levels of DC.

To achieve the goal, we designed an online web assessment tool following the same architecture as BAIT [24] to make easier the applicability of the results in BAIT. Other aspects that we had to consider during the design were: the test delivery mode (it would be under controlled conditions), the amount and type of content/questions (a significant number of knowledge and skills questions would be needed in order to evaluate the sub-competences selected for each DC) and the time needed for taking each test. We also decided to include different dynamic formats such as purpose-built interactive simulations,

as long as they could be monitored and assessed in a certification environment requiring safe settings. The number of items for each test was 41, 40 and 30 items, respectively, for each of the DCs of the IDL test and 44 items for the netiquette test. To design the items, we followed the design criteria outlined below:

- The shorter and simpler the better.
- Related to practical situations and common situations, especially in real-world scenarios.
- Neutral with respect to commercial brands and specific technological solutions. If this is not possible, use the most commonly used solutions as a basis. In the simulations, provide "*alt*" messages (alternative text to images) when hovering over the different options to help users who do not normally work with this tool.
- Address the selected competence elements (knowledge and skills) and refer to the three macro proficiency levels (foundation, intermediate and advanced).
- Balance the number of knowledge and skill questions (k/s) in each test: 22/22 for the netiquette test and 25/35 for the IDL test.
- All the items were dichotomous for all the formats (correct 1 and incorrect 0). The complexity of the tasks or any partial responses during the resolution of a task were not considered. We made this decision to simplify their understanding. On the other hand, the assessment criteria for each item would have been more complicated.

We used different item formats: multiple-choice questions, interactive simulations, image/simulation-based questions and open tasks. They must be displayed on a single screen to be responded to in a unique step without scrolling.

Interactive simulations, in which real-life situations were represented, and participants had to solve the tasks demanded by carrying out the required actions, such as sharing a document stored in the cloud or locating the nearest open pharmacy from the mobile phone. In the design of the simulations, we selected scenarios that can be commonly encountered in the context of the selected DCs. We sought to measure cognitive skills in a situation where digital technology must be applied. We were not interested in measuring technology use per se. This approach is better adapted to the fast technological change. We developed the simulations using a commercial solution called Articulate Storyline (ASL) [88]. ASL is a powerful solution for designing interactive simulations based on branching scenarios, which are a great way of providing authentic assessment. Simulations offer individuals a chance to put theory into practice by facing realistic situations that they might encounter in real life while interacting with different devices (e.g., mobile devices, laptops, workstations, etc.). Branching scenarios, where different choices take an individual down different paths, offer an opportunity for test designers to trace the performance of the examinees by truly assessing their aptitude. The simulations can be designed allowing for real behaviours that are usually performed in a real context (clicks, double clicks, enter text, right click, shortcut keys, etc.). So, for solving the tasks, we considered different paths and determined a limit of wrong clicks allowed, considering the difficulty of the item. With this approach, participants could explore programs and situations to a certain extent using judgment and decision making, rather than determining by memory the location of all functionalities. The different tasks shown in the simulations were abstracted from real applications widely used. The behaviour of the simulations was similar to the real applications (e.g., right-clicking shows you the context menu if applicable, clicking on a link underlined in blue color redirects you to the linked page, etc.). We designed the tasks in order to be delivered in a controlled environment. For each task, we collected the individual response time and the result (task solved or not). Furthermore, ASL allows the creation of scripts embedded in the simulations that allow the creation of variables to gather more information about the performance of examinees while solving the tasks. So, we additionally registered the number of clicks (correct or not) in each step of a scenario and the last step where an examinee finally failed. An example of the design of a simulation based on a mobile device in ASL can be seen in Figure 2.
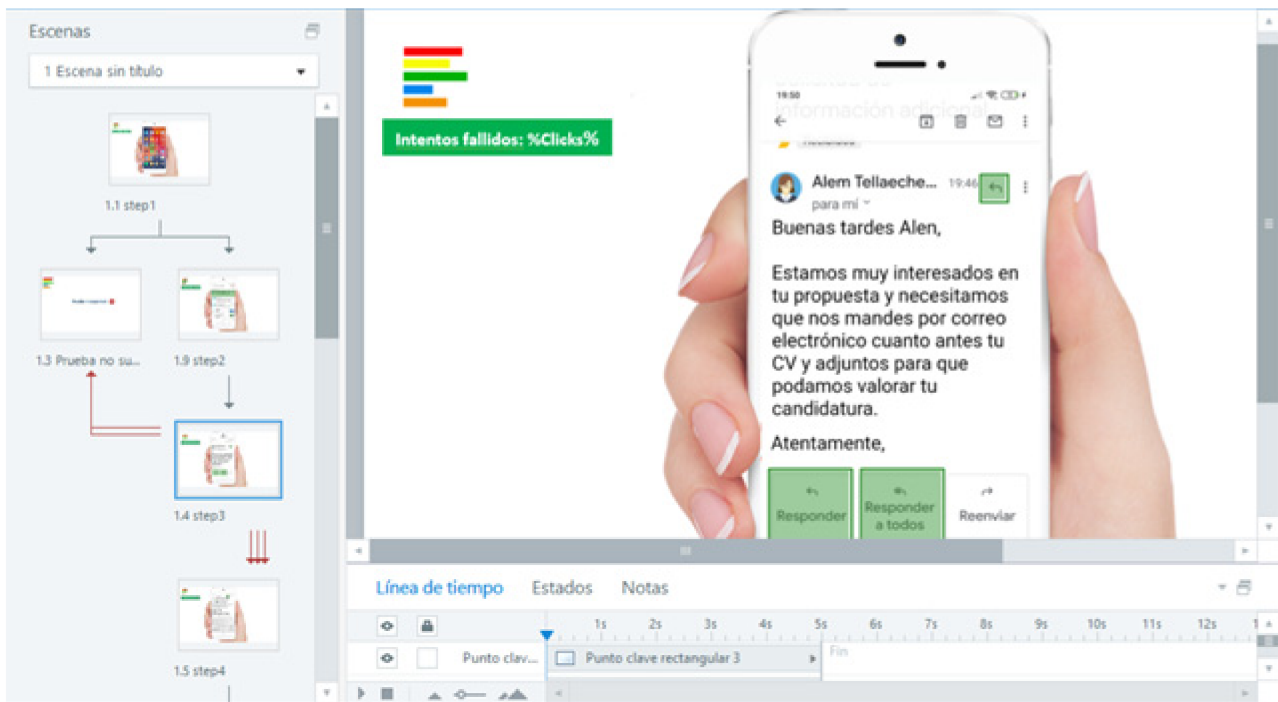
**Figure 2.** Example of the design of a simulation based on a mobile device in ASL.

Image/simulation-based questions, where a situation was presented to the participants and they had to critically evaluate the situations or put their knowledge into action by carrying out the required tasks, were used. This format can be appropriate for triggering and measuring higher-order skills related to intermediate and advanced levels of DigComp framework [87]. We applied the same design principles as for the interactive simulations, except that this format does not have a limit of wrong clicks, i.e., participants could examine the different areas freely.

Open tasks, where participants had to carry out the actions required to solve the tasks demanded by interacting with the computer and its applications, e.g., opening a spreadsheet and applying the necessary filters to locate a specific piece of information or accessing a simulated job vacancies portal "Lanbila" to solve specific tasks, were also used. We implemented this type of item by creating custom developments and integrating them into the assessment platform to evaluate the responses automatically (see Figures 3 and 4).



**Figure 3.** Design of an open task.

**Figure 4.** "*Lanbila*" web site.

Finally, all the different item formats selected were integrated in the CBA. An example of the interface can be seen in Figure 5. The platform was enabled to register the responses to the questions, the results obtained, the response times and additionally, for the simulations, the number of wrong clicks and the last step achieved (to know which path followed in their solutions).
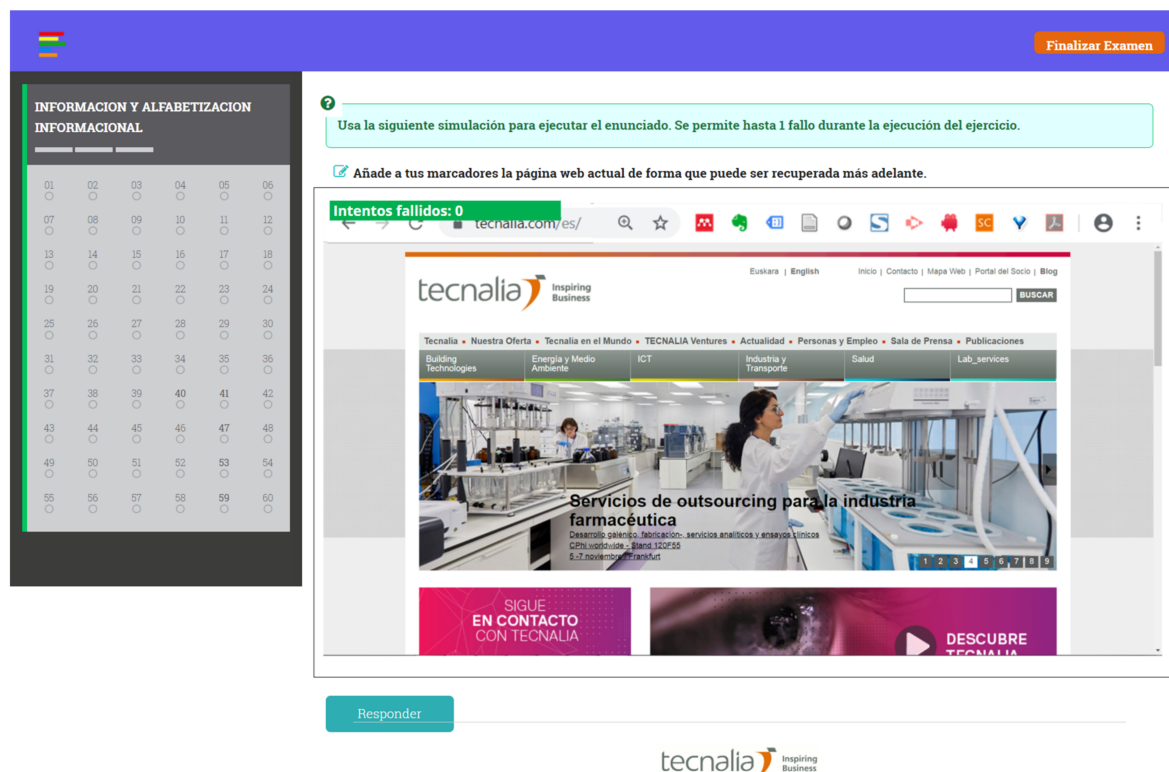


**Figure 5.** An example of the interface of the tests showing a simulation-based item.

After designing our assessment solution, we moved to the test phase to analyse the designed solution. Finally, based on the results of the test phases, we reviewed and refined the design elements and principles along with the implementation processes.

### 2.3. Phase 3: Iterative Cycles of Testing and Refinement of the Solution in Practice

We put into practice and evaluated the proposed solution in practice. In this section, we describe the results for both iterative cycles, detailing the methodology followed, given that it mainly represents the data collection and analysis phases of the study.

Finally, we would like to remark that due to the pandemic, we had to implement significant modifications in the data collection process of the second iteration. We could not gather the results in a controlled environment under a supervised session, as we initially planned. We had to organise an open call and most of participants carried out the tests from their homes in different conditions. Indeed, several comments and suggestions received were related to this fact. So, we had to examine all the feedback received in depth and discard reviews that will not be useful considering that the final tests will be run in a controlled and supervised environment.

#### 2.3.1. First Iteration with DC Centre Facilitators

During March of 2020, DC centre facilitators from the KZgunea telecentre network (KZgunea) [89] individually completed the tests and sent us their feedback. The four tests available at http://www.evaluatucompetenciadigital.com (accessed on 3 March 2022) were shown in Spanish, and the items were loaded into the tests in the same order. However, participants could navigate through the questions and change the order of their responses. The actions performed during the tests and the order of responses by the facilitators were gathered within the platform, and results were automatically generated. Summarising, in the cycle one testing phase we sought to investigate:

1. Content and wording of the items.
2. Facilitators' suggestions to improve some items identified as "to be improved".
3. Facilitators' suggestions to improve the questions/tests.

#### 2.3.2. Second Iteration with Citizens

During 22–28 March of 2021, the All Digital Week was held [90], offering various online activities with the aim of promoting the acquisition of digital skills. We decided to support the action by organising an online activity inviting citizens to assess their DC by completing the tests available. The campaign was mainly aimed at citizens familiar with IT Txartela [91] or people interested in improving their DC. To access our target group, we used several strategies: (1) we put a banner on the IT Txartela website, (2) publishing the event on social media and (3) accessing personal contacts such as friends, family and colleagues to reach a larger number of participants. To make participation in the study more attractive, we decided to give away some gadgets among the participants.

For the final version of the tests, we decided to create: (1) one test based on a competence area, with a selection of items from the three DCs of the IDL area and (2) s second test based on a DC, with a selection of items from the netiquette DC. Note that we discarded the cross relationships between competences identified in DigComp and measured each DC individually. This decision favoured the external validity of the tests. In addition, we had to explain in depth all the steps and decisions taken in each phase in order to be adopted by a wider audience [92]. The main factors that were considered for the development of the final tests were:

- Time needed for completing each test should be less than 30 min, to decrease the probability of users dropping out too early. So, the test for IDL competence area included 60 items and the test for netiquette DC included 44 items.
- The distribution of items in the IDL test was similar for each DC. In the netiquette test, the distribution was carried out ensuring that all sub-competences were present. The distribution of sub-competences was realised according to the literature review

carried out at the beginning of the study and the feedback received by the facilitators (see Table 4).

- According to the macro proficiency levels, we considered the following proportion for each test: 25% foundation level, 50% intermediate level and 25% advanced level. The proficiency levels were assigned to the items following a pragmatic approach mapping the verbs of the statements with the Bloom's taxonomy [93].

**Table 4.** Distribution of number of items for each sub-competence.

| Test | N° of Items | Sub-Competence |
|------|-------------|----------------|
| Netiquette | 10 | Sub-competence1 |
| | 11 | Sub-competence2 |
| | 16 | Sub-competence3 |
| | 7 | Sub-competence4 |
| Information and data literacy | 10 | Sub-competence5 |
| | 10 | Sub-competence6 |
| | 10 | Sub-competence7 |
| | 10 | Sub-competence8 |
| | 20 | Sub-competence9 |

The final versions of both tests are available for use in following studies at http://evaluatucompetenciadigital.com (accessed on 3 March 2022). Before starting the tests with citizens, examinees were provided with guided interactive help to be familiarised with the test environment. The items were loaded into the tests in the same order. However, participants could navigate through the questions and change the order of their responses. The actions performed during the tests and the order of responses by the participants were gathered within the platform, and results were automatically generated. Summarising, in the cycle two testing phase we sought to:

- Evaluate the tests with end users.
- Analyse the data gathered from the participants using different item response theory models and investigate the appropriateness of the models by examining different indicators of model fit.
- Examine the reliability and validity of the tests, which are concepts commonly used to evaluate the quality of the assessment instruments [31,94].

### 2.4. Phase 4: Reflection to Produce "Design Principles" and Enhance Solution Implementation

We followed a DBR because this methodology is suitable for describing the iterative process of the design and development of the main outputs of our study, the tests for evaluating the DC selected and to specify the main aspects considered and decisions made. The design principles described in the study contain procedural knowledge of the procedures, results and context followed during the different steps. We implemented our solution taking DigComp as a reference, which was created to be used as a reference for the development of tailored initiatives, providing a common terminology adaptable to our requisites. DigComp is not technology dependent and describes the competences in general terms. Stakeholders interested in developing their own implementation should identify which knowledge and skills are relevant and whether some specific applications or digital devices are key elements according to their peculiarities. So, we specified which knowledge and skills were of interest for our target group and designed an artefact to assess citizens' DC. For implementing the most suitable items, we used different formats such as interactive simulations and other dynamic formats. Note that for other DCs, other formats might be more appropriate. The readers will be able to decide which aspects might be of interest for their own implementations according to their specific settings.

## 3. Results

*3.1. Phase 3: Iterative Cycles of Testing and Refinement of the Solution in Practice*

### 3.1.1. First Iteration with DC Centre Facilitators

The participants were facilitators (*n* = 93) from 75 different centres of KZgunea. We did not follow any additional selection. The services provided by KZgunea include training in DC and support to the IT Txartela certification service [89]. Their expertise is of great value, as they support citizens' needs daily.

We sent the invitation with the details to participate in the study to the coordinator of KZgunea by email, including a template for collecting the information. The participation was voluntary. After the tests were completed, the coordinator sent us the filled templates. Then, we started with the analysis of the information. First, we identified which items obtained more comments and suggestions. We examined the items that obtained at least three mentions related to their comprehension difficulty or presented technical difficulties. In those cases, we analysed their comments and suggestions in detail. In addition, all the suggestions for improvement were analysed. Information related to the degree of difficulty of the items was duly annotated. If the comments on their level were too distant from the level initially selected, the question was reviewed in depth. Specifically, for the simulations, we increased or decreased the limit of wrong clicks allowed to adjust the difficulty level of the item. As a result, several items were reviewed and some of them modified (details can be found in the Table S2 in the Supplementary Material).

Next, we examined all the suggestions received related to the incorporation of new items. Suggestions that were oriented to specific applications were excluded since our approach was to see if examinees were able to understand any digital environment in an effective way instead of evaluating their knowledge about specific applications. As a result, several new items were developed (Table S3 in the Supplementary Material).

Finally, we also received some general comments and suggestions related to the test environment and new possibilities to be incorporated to the tests. The comments about the problems they had due to the use of different browsers and resolutions were dismissed because the tests were not taken in a controlled and optimised environment, using a specific and secured browser as we initially planned. We also were suggested to include the possibility to check the correct answers to the failed questions at the end of the test. Despite being out of the scope of this study, this suggestion was duly annotated for future implementations. As a result of this iteration, the tests were modified and prepared for the next iteration.

### 3.1.2. Second Iteration with Citizens

The participants were citizens who anonymously completed the IDL and netiquette tests, 329 and 214 participants, respectively. We did not follow any additional selection criteria, but we asked them to additionally fill in their sex and age range for demographics purposes. The overall completion time for the netiquette test was M = 916 s (SD = 585 s), i.e., more than 15 min, and M = 2007 s (SD = 1253 s), i.e., almost 34 min for the IDL test.

We removed 113 respondents from the IDL test and 16 respondents from the netiquette test due to their attempts having more than five skipped items, their attempts lasting only 5 min or less and it being judged unlikely that a participant could realistically read the items and answer in such a short time. The distribution of participants' records and their demographics are summarised in Table 5. There was a demographic weighting in favour of users in the age range of 25–54 and slightly more male users, especially in the IDL test. Table 6 shows the scores achieved in the tests.

**Table 5.** Respondent's record distribution and demographics.

| Test | Gender | Age Range |
|---|---|---|
| Netiquette (*n* = 201) entry 2 | Male 54.6% Female 46.4% | (16–24) 15.7% (25–54) 76.9% (55–74) 7.4% |
| IDL (*n* = 209) | Male 64.8% Female 35.2% | (16–24) 22.6% (25–54) 68.3% (55–74) 9.0% |

**Table 6.** Summary of descriptive data for the tests.

| Test | Mean | Standard Deviation |
|---|---|---|
| Netiquette (*n* = 201) | 24.70 | 8.70 |
| IDL (*n* = 209) | 42.69 | 11.25 |

In the development of a quantitative instrument for assessment purposes, it is crucial to measure its quality [95,96], which mainly consists of measuring its validity (if the assessment instrument assesses what should be measured and refers to how test scores are interpreted and used [94]) and reliability (if the assessment instrument produces similar results under equivalent conditions) [97]. To obtain evidence of quality, different methods are available, and studies involving the development of an assessment instrument should include enough evidence [95].

To obtain evidence of validity, both content and construct validity were considered. With regards to content validity, in the first iteration we already carried out a validation process based on expert judgment, specifically based on digital competence centre facilitators. We sought to confirm that the content of the tests represented their intended construct and were appropriate for accomplishing the testing purposes. Furthermore, some of the items included in the tests were previously examined by our previous work [85]. We analysed the item response processes of different types of items included in the tests, obtaining useful insights to understand the performance of the examinees and investigate if the assessment criteria for each item were correctly established. For the rest of the items included in the tests, we applied in the design of the items the main findings of our previous work [85]. Furthermore, this type of solution tends to present weaker internal validity (i.e., the evidence that the design reflects what is measured). Therefore, it was necessary to balance the internal and external validity through methodological decisions and the design of the tests. External validity is the extent to which the results of our study can be generalised to other contexts. Therefore, the relations between the measures obtained and potential confounding variables such as the participants socioeconomic status, gender or age could be investigated [31]. Hereafter, we outline the steps taken to obtain basic descriptive evidence of validity in the construction and validation of the tests:

- Examine the difficulty parameter (*p*-value) of the items and the discrimination indices as a starting indicator to justify the choice of the model.
- Analyse the dimensional validity and reliability.

First, we conducted a classical item analysis to examine the difficulty parameter (*p*-value) of the items. Items whose *p*-value is close to 0.00 (very difficult) or close to 1.00 (very easy) should be removed. In addition, it is necessary to investigate whether the items have similar discrimination indices [76] as a starting indicator to justify the choice of the model. The one parameter logistic model (1 PLM) has only one free parameter (the difficulty parameter) and expects that all items have similar discrimination indices of all items. Otherwise, the 1 PLM should not be applied. Therefore, we calculated the distribution of the point-biserial correlations, which is the Pearson correlation between each item and the total test score for each examinee. Items with a point-biserial value

smaller than 0.15 should be removed [98]. We only had to remove Item5 in the netiquette test because its correlation was <0.15 (see Table 7).

**Table 7.** Item characteristics: *p*-value and point-biserial correlations. Item5, in bold, was eliminated.

| Information and Data Literacy Test | | | Netiquette Test | | |
|---|---|---|---|---|---|
| **Item** | ***p*-Value** | **Point-Biserial Correlations** | **Item** | ***p*-Value** | **Point-Biserial Correlations** |
| Item1 | 0.87 | 0.567 | Item2 | 0.72 | 0.541 |
| Item2 | 0.83 | 0.435 | Item3 | 0.55 | 0.257 |
| Item3 | 0.53 | 0.527 | Item4 | 0.37 | 0.458 |
| Item4 | 0.82 | 0.468 | **Item5** | **0.21** | **0.113** |
| Item5 | 0.25 | 0.200 | Item6 | 0.61 | 0.311 |
| Item6 | 0.68 | 0.553 | Item7 | 0.56 | 0.431 |
| Item7 | 0.82 | 0.627 | Item8 | 0.31 | 0.398 |
| Item8 | 0.46 | 0.386 | Item9 | 0.67 | 0.474 |
| Item9 | 0.41 | 0.269 | Item10 | 0.59 | 0.422 |
| Item10 | 0.66 | 0.444 | Item11 | 0.29 | 0.256 |
| Item11 | 0.39 | 0.467 | Item13 | 0.67 | 0.413 |
| Item12 | 0.62 | 0.450 | Item14 | 0.75 | 0.396 |
| Item13 | 0.47 | 0.439 | Item15 | 0.67 | 0.615 |
| Item14 | 0.51 | 0.450 | Item16 | 0.64 | 0.478 |
| Item15 | 0.63 | 0.342 | Item17 | 0.44 | 0.183 |
| Item16 | 0.73 | 0.500 | Item19 | 0.64 | 0.568 |
| Item17 | 0.76 | 0.549 | Item20 | 0.64 | 0.466 |
| Item18 | 0.71 | 0.431 | Item22 | 0.60 | 0.441 |
| Item19 | 0.45 | 0.299 | Item23 | 0.43 | 0.233 |
| Item20 | 0.67 | 0.383 | Item24 | 0.41 | 0.377 |
| Item21 | 0.82 | 0.317 | Item25 | 0.69 | 0.604 |
| Item22 | 0.89 | 0.560 | Item28 | 0.70 | 0.455 |
| Item23 | 0.78 | 0.504 | Item29 | 0.66 | 0.659 |
| Item25 | 0.87 | 0.541 | Item30 | 0.66 | 0.374 |
| Item26 | 0.84 | 0.405 | Item31 | 0.29 | 0.164 |
| Item27 | 0.84 | 0.453 | Item32 | 0.60 | 0.329 |
| Item28 | 0.60 | 0.261 | Item33 | 0.62 | 0.571 |
| Item29 | 0.71 | 0.414 | Item34 | 0.66 | 0.534 |
| Item30 | 0.67 | 0.291 | Item35 | 0.50 | 0.554 |
| Item31 | 0.81 | 0.626 | Item36 | 0.60 | 0.694 |
| Item32 | 0.80 | 0.445 | Item37 | 0.41 | 0.243 |
| Item33 | 0.88 | 0.506 | Item38 | 0.67 | 0.534 |
| Item34 | 0.59 | 0.282 | Item39 | 0.76 | 0.501 |
| Item35 | 0.76 | 0.575 | Item40 | 0.37 | 0.173 |
| Item36 | 0.80 | 0.517 | Item41 | 0.81 | 0.622 |
| Item37 | 0.87 | 0.460 | Item42 | 0.56 | 0.373 |
| Item38 | 0.65 | 0.404 | Item43 | 0.55 | 0.537 |
| Item39 | 0.70 | 0.350 | Item44 | 0.34 | 0.229 |
| Item40 | 0.76 | 0.529 | Item45 | 0.61 | 0.545 |
| Item41 | 0.78 | 0.415 | Item46 | 0.74 | 0.568 |
| Item42 | 0.89 | 0.580 | Item47 | 0.60 | 0.433 |
| Item43 | 0.69 | 0.344 | Item48 | 0.51 | 0.251 |
| Item44 | 0.80 | 0.433 | Item49 | 0.50 | 0.431 |
| Item45 | 0.71 | 0.552 | Item50 | 0.79 | 0.191 |
| Item46 | 0.52 | 0.358 | | | |
| Item47 | 0.83 | 0.556 | | | |
| Item48 | 0.90 | 0.443 | | | |
| Item49 | 0.59 | 0.374 | | | |
| Item50 | 0.74 | 0.410 | | | |
| Item51 | 0.94 | 0.585 | | | |
| Item52 | 0.91 | 0.528 | | | |
| Item53 | 0.78 | 0.552 | | | |
| Item54 | 0.46 | 0.219 | | | |
| Item55 | 0.93 | 0.501 | | | |
| Item56 | 0.74 | 0.435 | | | |
| Item57 | 0.86 | 0.543 | | | |
| Item58 | 0.66 | 0.512 | | | |
| Item59 | 0.68 | 0.660 | | | |
| Item60 | 0.57 | 0.342 | | | |

Evaluating the internal construct validity and dimensionality of a new measure is a relevant element of evidence, to examine whether the effects observed in our study are caused by the manipulation of the independent variable and not by other factors [31].

As a preliminary analysis, we performed an exploratory factor analysis (EFA), which is one of the most frequently applied techniques in test development and validation studies to explore the set of latent variables or common factors that explain responses to test items. We applied the principal components factor analysis with a varimax rotation to examine the factor loadings and dimensionality of both tests. Before carrying out the EFA, we calculated Bartlett's test of sphericity to examine factorability of the data and Kaiser–Meyer–Olkin's (KMO) test to evaluate sampling appropriateness. Results confirmed a significant test statistic for Bartlett's test of sphericity. For the IDL test, a chi square of 339,326, $p < 0.001$ and a KMO value of 0.717 were obtained, and for the netiquette test a chi square of 344,640, $p < 0.001$ and a KMO value of 0.818 were obtained, which means that the data had adequate structure detection. The exploratory factor analysis of the data, using the principal component extraction method and a varimax rotation of all the items, revealed one strong factor explaining 80% of the total variance for the IDL test and 70% for the netiquette test. These outcomes provided support for concluding that there is one strong general factor that all items in both tests relate to and can be interpreted to be the participants' general DC.

Then, we investigated which model fits the data significantly better, the less restricted model (multidimensional Rasch model) or the simpler model (unidimensional Rasch model). We took two different approaches for both tests. For the IDL test, we considered the three DCs of the competence area as independent dimensions. For the netiquette test, we considered the four sub-competences selected as independent dimensions. We calculated the difference in deviances in the estimation of the two different models, which is expected to follow a chi-square distribution, and the degrees of freedom, which is the difference in the number of parameters. Thus, we can statistically calculate which model fits the data significantly better.

For the IDL test, the difference between the deviances of these two models follows a chi-square distribution with five degrees of freedom and an estimated difference of 75.9 in the deviance. Therefore, the three-dimensional model fit the data better than the unidimensional one (see Table 8).

**Table 8.** Main model indicators for IDL test.

| Model | Deviance | Number of Parameters |
|---|---|---|
| 1-dim | 11,900.3 | 61 |
| 3-dim | 11,824.4 | 66 |

We calculated the weighted mean-square fit statistic for all the items to check the alignment of the items with the multidimensional Rasch model. This statistic shows the amount of inaccuracy of the measurement system [98], which should be near the unit. However, values falling within the range of 0.75 to 1.33 are widely accepted [99]. We only found Item50 not acceptable, whose fit was 0.74 (see Table 9).

**Table 9.** Item analysis results (multidimensional model).

| | |
|---|---|
| Sample size | 209 |
| Number of items in calibration | 60 |
| Weighted fit MNSQ (0.75, 1.33) T sig. | 1 (0.74) |
| Reliability estimates: EAP/PV reliability | |
| DC 1.1 | 0.880 |
| DC 1.2 | 0.840 |
| DC 1.3 | 0.875 |

We also examined the EAP/PV estimate of reliability for the test that is provided by the Conquest software [100], which is similar to other reliability estimates such as Cronbach's alpha [101]. The estimated latent correlations between the three dimensions were high (see Table 10), implying that they may be evaluating the same trait, i.e., there is one strong factor that that underlies all items, which can be interpreted as general DC. We found similar findings in the netiquette test, as we show below.

**Table 10.** Correlations among the three dimensions (based on DCs).

| Dimensions | DC 1.1 | DC 1.2 | DC 1.3 |
|------------|--------|--------|--------|
| DC 1.1 | | | |
| DC 1.2 | 0.785 | | |
| DC 1.3 | 0.929 | 0.847 | |

Figure 6 shows the locations of examinees and items on the same scale using a Wright map, which is a powerful yet simple graphical tool. The "X" illustrates the location of examinees on each dimension. On the right side of the graph, the items are shown. The items are represented on the right, increasing in difficulty from bottom to top. If the examinee and the item are aligned, then the probability of responding to that item correctly is practically 50%. If the examinees' location is higher, the probability of responding to that item correctly increases and vice versa.

For the netiquette test, we note that the difference between the deviances of these two models follows a chi-square distribution with nine degrees of freedom and an estimated difference of 26.2 in the deviance. Therefore, the four-dimensional model fit the data better than the unidimensional one (see Table 11).

**Table 11.** Main model indicators for the netiquette test.

| Model | Deviance | Number of Parameters |
|-------|----------|---------------------|
| 1-dim | 10,100.0 | 44 |
| 4-dim | 10,073.8 | 53 |

We calculated the weighted mean-square fit statistic for all the items to check the alignment of the items with the multidimensional Rasch model and did not find any item that fell outside the acceptable range (see Table 12).

**Table 12.** Item analysis results (multidimensional model).

| | |
|---|---|
| Sample size | 201 |
| Number of items in calibration | 43 |
| Weighted fit MNSQ (0.75, 1.33) T sig. | none |
| Reliability estimates: EAP/PV reliability | |
| Sub-competence1 | 0.822 |
| Sub-competence2 | 0.795 |
| Sub-competence3 | 0.859 |
| Sub-competence4 | 0.774 |

```
         Dim. 1.1          | Dim. 1.2  |  Dim. 1.3  |
                           |           |            |
                           |         X |            |
                           |         X |            |
    4                      |         X |          X |
                           |         X |          X |
                           |        XX |         XX |
                         X |        XX |        XXX |
                           |     XXXXX |      XXXXX |
                           |     XXXXX |   XXXXXXXX |
    3                    X |      XXXX |     XXXXXX |
                         X |    XXXXXX |    XXXXXXX |
                        XX | XXXXXXXXXXX |   XXXXXXXX |
                       XXX |  XXXXXXXXX | XXXXXXXXXX |
          XXXXXXXXXXXXXXXXX |           |   XXXXXXXX |
    2              XXXXX | XXXXXXXXXXX | XXXXXXXXX |5
               XXXXXXX |  XXXXXXXXX | XXXXXXXXXX |53
             XXXXXXXXX |  XXXXXXXXX |  XXXXXXXXX |
              XXXXXXX |   XXXXXXXX | XXXXXXXXXXX |45
           XXXXXXXXXXX |   XXXXXXX |   XXXXXXX |59
            XXXXXXXXXX |   XXXXXXX | XXXXXXXXXX |9 11 27 33 48
    1       XXXXXXXXXX |   XXXXXX |    XXXXXX |19
             XXXXXXX |   XXXXXXXX |     XXXX |8 13 37 57
            XXXXXXXXX |     XXXXX |    XXXXX |14 29 38 42 58
          XXXXXXXXXXXX |      XXXX |   XXXXXX |3 28 44
             XXXXXXX |      XXXX |     XXXX |49 55
           XXXXXXXXX |       XXX |      XXX |12 23 34 39 40 52
    0        XXXXXXXX |       XXX |      XXX |15 31 35
           XXXXXXXXX |       XXX |       XX |6 10 20 21 30 43
             XXXXXX |       XXX |          |18 25 26 46
             XXXXX |         X |        X |16 56 60
               XX |       XXX |        X |17 24 36
   -1          XXXX |        XX |        X |32 41
               XX |         X |        X |22
                X |         X |        X |2 4 7 47 51
              XXX |         X |          |
                X |           |        X |1 50 54
                X |           |       XX |
   -2           X |           |        X |
                  |           |            |
                X |           |            |
                  |           |            |
               XX |           |            |
                  |           |            |
   -3             |           |            |
                  |           |            |
                  |           |            |
                  |           |            |
                  |           |            |
   -4             |           |            |
=================================================================
Each 'X' represents 1.2 cases
```
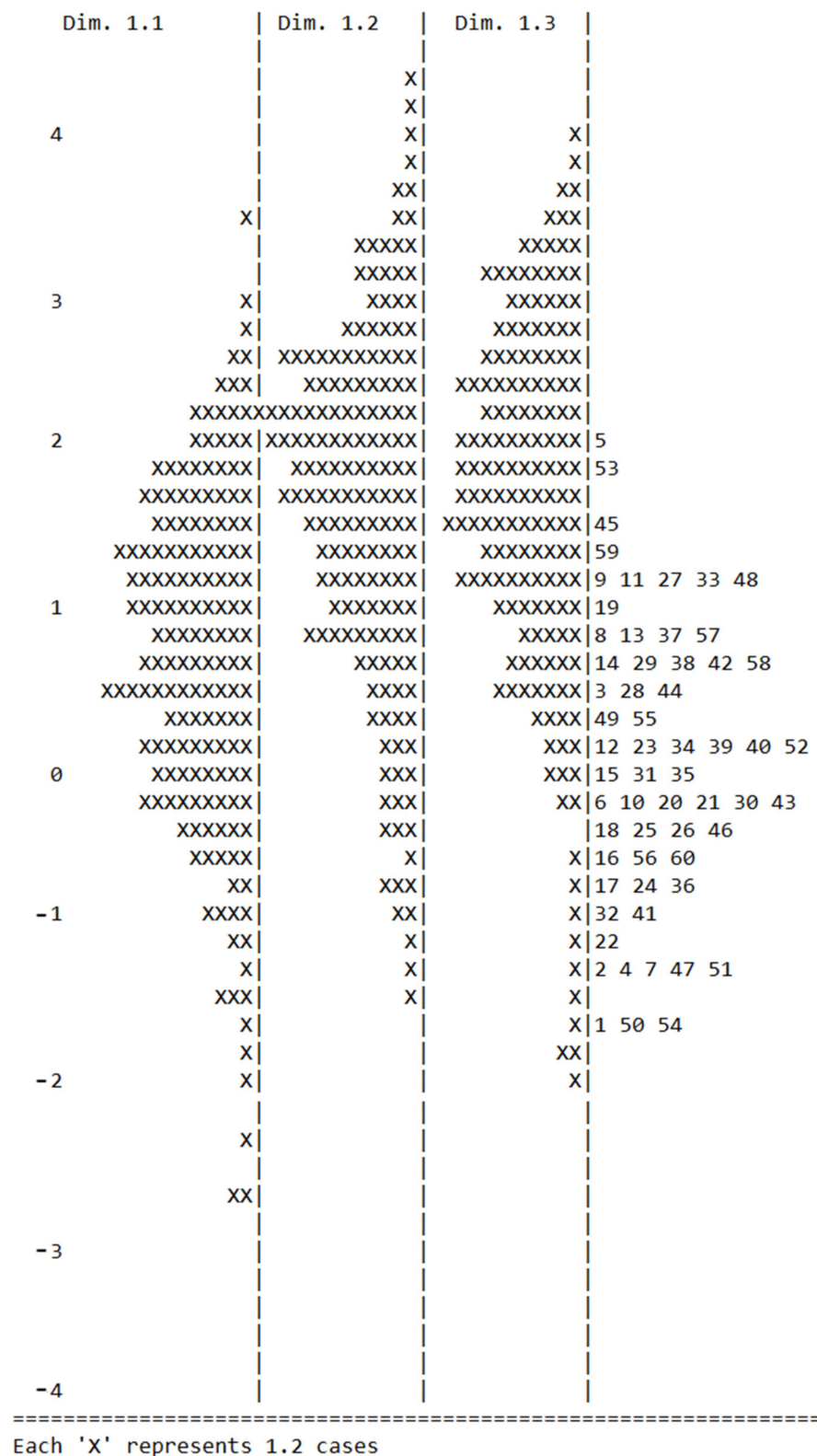
**Figure 6.** Representation of the dimensions of the IDL test.

The estimated latent correlations between the four dimensions were high for the netiquette test too (see Table 13).

**Table 13.** Correlations among the four dimensions corresponding to the four sub-competences (SC).

| Dimensions | SC1 | SC2 | SC3 | SC4 |
|---|---|---|---|---|
| SC1 | | | | |
| SC2 | 0.854 | | | |
| SC3 | 0.913 | 0.845 | | |
| SC4 | 0.827 | 0.806 | 0.862 | |

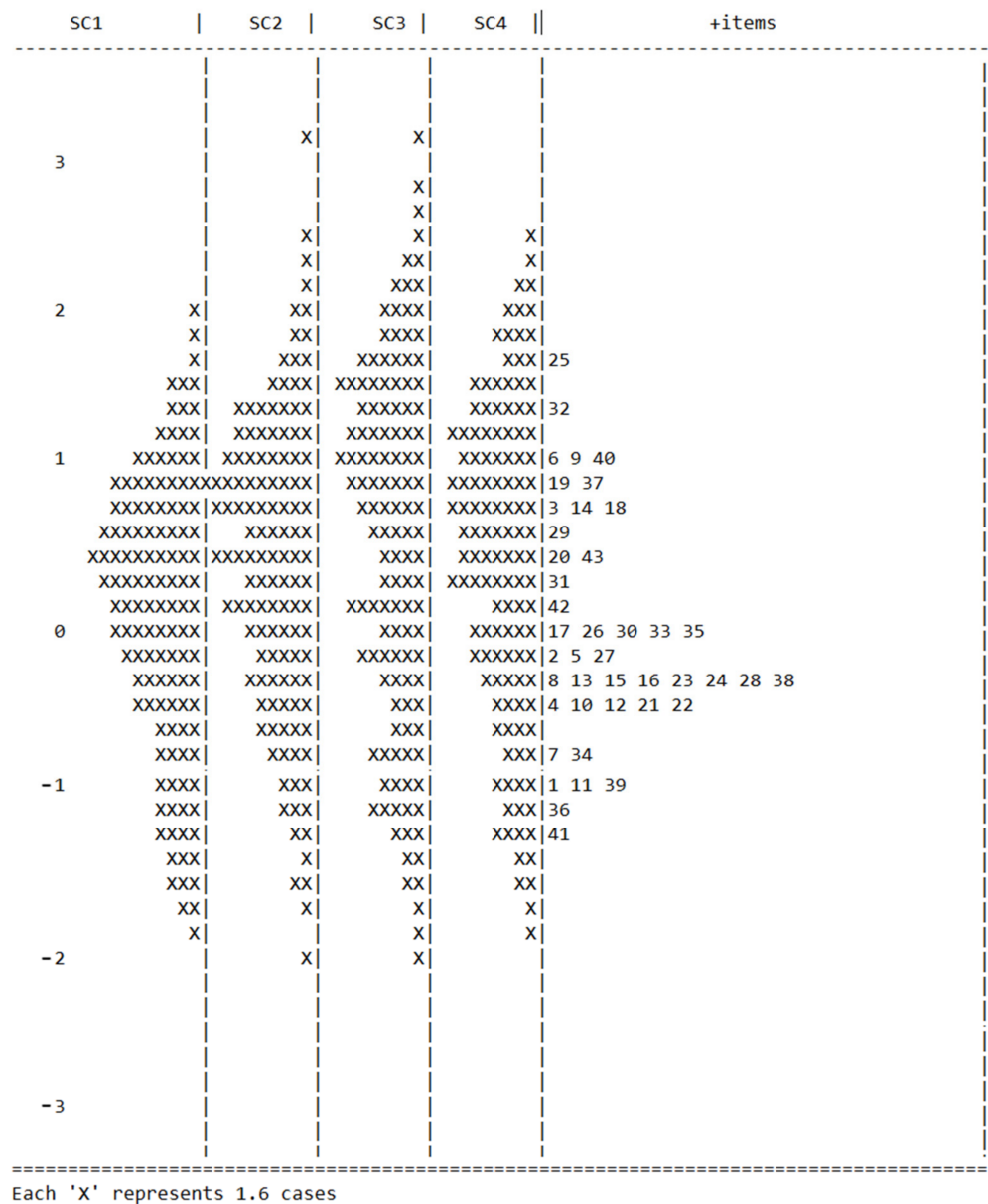Figure 7 shows the location of examinees and items on the same scale base using a Wright map.



**Figure 7.** Representation of the dimensions of the netiquette test.

Finally, we calculated the EAP/PV estimate to investigate the internal consistency of all the dimensions, with values between 0.78 and 0.88 (see Tables 9 and 12), with all coefficients higher than 0.70 indicating good internal consistency. Cronbach's alpha for the overall tests was 0.93 (IDL) and 0.89 (netiquette).

## 4. Discussion and Conclusions

For reaching the SDGs, it is vital to empower citizenship with the right capabilities to use technology meaningfully to participate in today's society, where all areas are affected by the effects of digital technology to a greater or lesser extent. In this context, the recognition of DC is one of the main lines of action of the European Commission in recent years along with the promotion of a common framework of reference for DC. However, most accreditation systems have generally been insufficient [8,9,11]. Thanks to the release of DigComp and the possibilities of developing authentic item formats such as interactive simulations in TEAs, the development of custom implementations of instruments for the assessment of DC, measuring not only low-order cognitive skills, has been facilitated. In addition, DC requires constant reviewing due to the constant changes in how citizens access and manage information through different types of devices.

This paper reports on a DBR study on DC assessment in the context of citizenship interested in accrediting their DC. We sought to design an assessment instrument incorporating different item formats to assess higher-order skills, which are normally scarce in today's landscape. So, we decided to design assessment items according to the descriptors of the intermediate and advanced levels of DigComp. Therefore, it was crucial to choose the most adequate item formats, such as purpose-built interactive simulations, and correctly design them to trigger the intended behaviours and measure higher-order cognitive skills when necessary. Note that this fact is particularly important when assessing complex cognitive constructs and requires examinees to put their knowledge into action, providing the most precise picture of their level of DC.

One of the strengths of our study was the selection of case studies. To the best of our knowledge, netiquette DC has not yet been assessed in depth as we carried out in our study. Regarding the IDL test, other authors have designed assessment tests with similar aims, but most of them have been developed from two perspectives, librarian and academic, and they are often domain-specific, as many of them are self-reporting tools or are tools based on multiple-choice questions focused on low-order cognitive skills [22,57,58]. In addition, we had to review the DCs in depth to select the content and sub-competences to be assessed in each test, due to their nature of constant change. During the first iteration with the facilitators, part of the efforts was dedicated to validating the proposal of contents and sub-competences initially identified from the literature review. Thus, we selected two use cases following different approaches: (1) a test based on the IDL competence area and (2) a test based on netiquette DC. Both approaches have their singularities but are valid and have been selected by different stakeholders, e.g., based on competence area [102] and based on DC [24]. So, the main objective of our study was to investigate the peculiarities of both approaches and describe the design principles applied during the different steps of the development of the tests. Apart from that, the tests have sound psychometric properties that make them reliable and valid instruments for measuring DC, even though the two approaches differed in the degree of depth in assessing the competencies covered by their respective tests. That is, in the case of the netiquette test, the number of questions per DC was much higher than in the case of the three DCs included in the IDL test.

Another strength of our study was the methodology followed. The development was carried out in an iterative process, which included two cycles of testing and refinement, validating the content of the construct to be assessed and the design of the items, ensuring that the expected knowledge and skills are covered, the item formats selected are suitable for that aim, the usability of the items/tests is correct, e.g., the simulations include all the actions and different possible paths, and finally, the items are well written and easy to understand, which is a key factor in the development of an item bank of DC [86].

In the development of a quantitative instrument for assessment purposes, it is crucial to measure its quality [95–97]. However, recent reviews of tools for the assessment of DC concluded that the evidence provided is not enough [11,13,14]. In view of this, we designed our study by planning several studies throughout the different phases to obtain enough evidence to ensure the quality of the instruments. Although the multidimensionality

of the construct of DC has been identified in several studies, theoretical and empirical studies have reported contradicting results [32]. More specifically, Vuorikari et al. [25] theoretically described the construct of DC in DigComp framework 2.0, but the dimensions identified have not been empirically confirmed or require further research. Our findings showed that the three-dimensional model for the IDL test and the four-dimensional model for the netiquette test fitted the data better than the unidimensional model. However, considering the high correlations obtained, it seems that all the items relate to one strong factor, which can be interpreted as general DC. Recent studies have pointed in the same direction, e.g., [103]. The results obtained in the netiquette test are also interesting, where female participants obtained higher average scores than male participants. Even more, participants in the 55–74 age range obtained better results. The number of participants of this age range was very low (7,4%), and it would be interesting to confirm these results by piloting the test with a bigger number of participants, including more people of this age range.

Moreover, some limitations should be considered when interpreting the findings of this study. The sample of participants was relatively small considering the target group of interest, especially in some age ranges. Although the main age range of interest of our study was 25–54, it would be of interest to see if the results would be replicated with a larger sample and more varied in terms of age. Another point worth mentioning is that, although the tool has been designed for citizenship in general, in the validation process with KZgunea experts, the validation criteria that were applied might not have been applied in other regions, i.e., perhaps the facilitators of other telecentre networks might consider other sub-competences of more interest, etc. So, there may be some variability between different regions that may require the application of adaptations.

However, some weaknesses should be mentioned. Developing interactive simulations and integrating them in the assessments in order to be automatically scored require a big effort. In addition, it is necessary to design the interactive simulations based on applications that are as neutral as possible, so that examinees who have not used them will be able to carry out the tasks in a logical manner. Unfortunately, most of the applications are constantly being updated, many of them introducing major design changes. This fact makes it advisable to constantly review the simulations, examining whether any changes are important enough to consider redesigning them. In a similar way, the assessment criteria should also be checked constantly, as certain learning objectives may become obsolete and new ones may emerge. The reviews and updates will have to be regular and constant.

The objective of our study was to describe the different phases of the design and validation of the tests for evaluating the DC selected, specifying the main aspects considered and decisions made. In addition, we presented validity and reliability evidence to ensure the quality of the tests. That is the reason why we followed a DBR, as this methodology is useful to describe the design principles applied during the different steps of the development of the tests for evaluating the DCs selected, with the aim that they can be extended to the rest of the remaining DCs included in the reference framework. Stakeholders interested in developing their own implementations might find this study of interest to decide which insights might be of interest for their own implementations according to their specific settings. The DigComp framework covers twenty-one different DCs, each of them with their own peculiarities. Depending on the descriptors to be assessed for a particular DC, it may be appropriate to use one item format versus another. Further studies may consider incorporating new innovative items, which could trigger more complex response processes, such as developing and integrating purpose-built games. We should also underline that we are in a rapidly digitalising world with constant changes in people's behaviour and habits caused by the continuous emergence of new technologies and applications. DC is a complex and constantly evolving construct, and it should be reviewed periodically.

## References

1. List, A.; Brante, E.W.; Klee, H.L. A framework of pre-service teachers' conceptions about digital literacy: Comparing the United States and Sweden. *Comput. Educ.* **2020**, *148*, 103788. [CrossRef]
2. O'Sullivan, K.; Clark, S.; Marshall, K.; MacLachlan, M. A Just Digital framework to ensure equitable achievement of the Sustainable Development Goals. *Nat. Commun.* **2021**, *12*, 6345. [CrossRef]
3. Ala-Mutka, K. *Mapping Digital Competence: Towards a Conceptual Understanding*; Institute for Prospective Technological Studies: Sevilla, Spain, 2011; pp. 7–60.
4. Abidoye, R.; Lim, B.T.H.; Lin, Y.C.; Ma, J. Equipping Property Graduates for the Digital Age. *Sustainability* **2022**, *14*, 640. [CrossRef]
5. Portillo, J.; Garay, U.; Tejada, E.; Bilbao, N. Self-perception of the digital competence of educators during the COVID-19 pandemic: A cross-analysis of different educational stages. *Sustainability* **2020**, *12*, 10128. [CrossRef]
6. Sá, M.J.; Santos, A.I.; Serpa, S.; Miguel Ferreira, C. Digitainability—Digital Competences Post-COVID-19 for a Sustainable Society. *Sustainability* **2021**, *13*, 9564. [CrossRef]
7. Ferrari, A. *Digital Competence in Practice: An Analysis of Frameworks*; JRC IPTS: Seville, Spain, 2012. [CrossRef]
8. Law, N.W.Y.; Woo, D.J.; de la Torre, J.; Wong, K.W.G. *A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2*; UNESCO: Paris, France, 2018; p. 146.
9. Santos, A.I.; Serpa, S. The importance of promoting digital literacy in higher education. *Int. J. Soc. Sci. Stud.* **2017**, *5*, 90. [CrossRef]
10. Ferrari, A. *DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe*; Publications Office of the European Union: Brussels, Belgium, 2013. [CrossRef]
11. Siddiq, F.; Hatlevik, O.E.; Olsen, R.V.; Throndsen, I.; Scherer, R. Taking a future perspective by learning from the past—A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educ. Res. Rev.* **2016**, *19*, 58–84. [CrossRef]
12. Kluzer, S.; Priego, L.P. *Digcomp into Action: Get Inspired, Make it Happen. A User Guide to the European Digital Competence Framework*; Joint Research Centre: Seville, Spain, 2018.
13. Zhao, Y.; Llorente, A.M.P.; Gómez, M.C.S. Digital competence in higher education research: A systematic literature review. *Comput. Educ.* **2021**, *168*, 104212. [CrossRef]
14. Saltos-Rivas, R.; Novoa-Hernández, P.; Rodríguez, R.S. On the quality of quantitative instruments to measure digital competence in higher education: A systematic mapping study. *PLoS ONE* **2021**, *16*, e0257344. [CrossRef] [PubMed]
15. Greiff, S.; Wüstenberg, S.; Avvisati, F. Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving. *Comput. Educ.* **2015**, *91*, 92–105. [CrossRef]
16. Osborne, R.; Dunne, E.; Farrand, P. Integrating technologies into "authentic" assessment design: An affordances approach. *Res. Learn. Technol.* **2013**, *21*, 21986. [CrossRef]
17. Timmis, S.; Broadfoot, P.; Sutherland, R.; Oldfield, A. Rethinking assessment in a digital age: Opportunities, challenges and risks. *Br. Educ. Res. J.* **2016**, *42*, 454–476. [CrossRef]
18. Binkley, M.; Erstad, O.; Herman, J.; Raizen, S.; Ripley, M.; Miller-Ricci, M.; Rumble, M. Defining twenty-first century skills. In *Assessment and Teaching of 21st Century Skills*; Springer: Dordrecht, The Netherlands, 2012; pp. 17–66. [CrossRef]

19. Nguyen, Q.; Rienties, B.; Toetenel, L.; Ferguson, R.; Whitelock, D. Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates. *Comput. Hum. Behav.* **2017**, *76*, 703–714. [CrossRef]

20. Rienties, B.; Toetenel, L. The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. *Comput. Hum. Behav.* **2016**, *60*, 333–341. [CrossRef]

21. Papamitsiou, Z.; Economides, A.A. Learning analytics for smart learning environments: A meta-analysis of empirical research results from 2009 to 2015. In *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 1–23. [CrossRef]

22. Heer, R. *A Model of Learning Objectives–Based on a Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*; Center for Excellence in Learning and Teaching, Iowa State University: Ames, IA, USA, 2012; Available online: www.celt.iastate.edu/wp-content/uploads/2015/09/RevisedBloomsHandout-1.pdf (accessed on 19 January 2022).

23. Eurostat. Being Young in Europe Today–Digital World. 2017. Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Being_young_in_Europe_today (accessed on 19 January 2022).

24. BAIT—Evaluation and Certification System of Digital Competences. Available online: http://www.bait.eus (accessed on 19 January 2022).

25. Vuorikari, R.; Punie, Y.; Carretero Gomez, S.; Van Den Brande, G. *DigComp 2.0: The Digital Competence Framework for Citizens*; EUR 27948 EN; Publications Office of the European Union: Luxembourg, 2016. [CrossRef]

26. Carretero, S.; Vuorikari, R.; Punie, Y. *DigComp 2.1: The Digital Competence Framework for Citizens with Eight Proficiency Levels and Examples of Use*; EUR 28558 EN; Publications Office of the European Union: Luxembourg, 2017. [CrossRef]

27. Laanpere, M. Recommendations on Assessment Tools for Monitoring Digital Literacy within UNESCO's Digital Literacy Global Framework. Information Paper No, 56. 2019. Available online: https://unesdoc.unesco.org/ark:/48223/pf0000366740 (accessed on 3 March 2022).

28. Bashir, S.; Miyamoto, K. *Digital Skills: Frameworks and Programs*; World Bank: Washington, DC, USA, 2020; Available online: https://openknowledge.worldbank.org/handle/10986/35080 (accessed on 3 March 2022).

29. Fraillon, J. International large-scale computer-based studies on information technology literacy in education. In *Second Handbook of Information Technology in Primary and Secondary Education*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 1161–1179.

30. Sparks, J.R.; Katz, I.R.; Beile, P.M. Assessing digital information literacy in higher education: A review of existing frameworks and assessments with recommendations for next-generation assessment. *ETS Res. Rep. Ser.* **2016**, *2016*, 1–33. [CrossRef]

31. Messick, S. Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *Am. Psychol.* **1995**, *50*, 741–749. [CrossRef]

32. Reichert, F.; Zhang, D.J.; Law, N.W.; Wong, G.K.; de la Torre, J. Exploring the structure of digital literacy competence assessed using authentic software applications. *Educ. Technol. Res. Dev.* **2020**, *68*, 2991–3013. [CrossRef]

33. Jin, K.Y.; Reichert, F.; Cagasan, L.P., Jr.; de la Torre, J.; Law, N. Measuring digital literacy across three age cohorts: Exploring test dimensionality and performance differences. *Comput. Educ.* **2020**, *157*, 103968. [CrossRef]

34. Aesaert, K.; Van Nijlen, D.; Vanderlinde, R.; van Braak, J. 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. *Comput. Educ.* **2014**, *76*, 168–181. [CrossRef]

35. Goldhammer, F.; Naumann, J.; Keßel, Y. Assessing individual differences in basic computer skills. *Eur. J. Psychol. Assess.* **2013**, *29*, 263–275. [CrossRef]

36. Huggins, A.C.; Ritzhaupt, A.D.; Dawson, K. Measuring information and communication technology literacy using a performance assessment: Validation of the student tool for technology literacy (ST2L). *Comput. Educ.* **2014**, *77*, 1–12. [CrossRef]

37. Pérez-Escoda, A.; Esteban, L.M.P. Retos del periodismo frente a las redes sociales, las fake news y la desconfianza de la generación Z. *Rev. Lat. Comun. Soc.* **2021**, *79*, 67–85. [CrossRef]

38. Dessart, L. Social media engagement: A model of antecedents and relational outcomes. *J. Mark. Manag.* **2017**, *33*, 375–399. [CrossRef]

39. Pérez-Escoda, A.; Pedrero-Esteban, L.M.; Rubio-Romero, J.; Jiménez-Narros, C. Fake News Reaching Young People on Social Networks: Distrust Challenging Media Literacy. *Publications* **2021**, *9*, 24. [CrossRef]

40. Larrondo-Ureta, A.; Peña-Fernández, S.; Agirreazkuenaga-Onaindia, I. Hacia una mayor participación de la audiencia: Experiencias transmedia para jóvenes. *Estud. Sobre Mensaje Periodístico* **2020**, *26*, 1445–1454. [CrossRef]

41. Castillo-Abdul, B.; Romero-Rodríguez, L.M.; Larrea-Ayala, A. Kid influencers in Spain: Understanding the themes they address and preteens' engagement with their YouTube channels. *Heliyon* **2020**, *6*, e05056. [CrossRef] [PubMed]

42. Vraga, E.K.; Bode, L. Defining misinformation and understanding its bounded nature: Using expertise and evidence for describing misinformation. *Political Commun.* **2020**, *37*, 136–144. [CrossRef]

43. Masip, P.; Suau, J.; Ruiz-Caballero, C. Perceptions on media and disinformation: Ideology and polarization in the Spanish media system. *Prof. Inf.* **2020**, *29*, 1–13. [CrossRef]

44. Viner, K. How Technology Disrupted the Truth. *The Guardian*. 12 July 2016. Available online: https://www.theguardian.com/media/2016/jul/12/how-technology-disrupted-the-truth (accessed on 19 January 2022).

45. Orso, D.; Federici, N.; Copetti, R.; Vetrugno, L.; Bove, T. Infodemic and the spread of fake news in the COVID-19-era. *Eur. J. Emerg. Med.* **2020**, *27*, 327–328. [CrossRef]

46. Kopecky, K.; Szotkowski, R.; Aznar-Díaz, I.; Romero-Rodríguez, J.M. The phenomenon of sharenting and its risks in the online environment. Experiences from Czech Republic and Spain. *Child. Youth Serv. Rev.* **2020**, *110*, 104812. [CrossRef]

47. European Commission. Standard Eurobarometer 93. Summer 2020. Report. Public Opinion in the European Union. 2020. Available online: https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/ResultDoc/download/DocumentKy/91061 (accessed on 3 March 2022).

48. Jones-Jang, S.M.; Mortensen, T.; Liu, J. Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *Am. Behav. Sci.* **2021**, *65*, 371–388. [CrossRef]

49. Walsh, A. Information literacy assessment: Where do we start? *J. Libr. Inf. Sci.* **2009**, *41*, 19–28. [CrossRef]

50. Catalano, A. The effect of a situated learning environment in a distance education information literacy course. *J. Acad. Libr.* **2015**, *41*, 653–659. [CrossRef]

51. Foo, S.; Majid, S.; Chang, Y.K. Assessing information literacy skills among young information age students in Singapore. *Aslib J. Inf. Manag.* **2017**, *69*, 335–353. [CrossRef]

52. Kruger, J.; Dunning, D. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* **1999**, *77*, 1121. [CrossRef] [PubMed]

53. Mahmood, K. Do people overestimate their information literacy skills? A systematic review of empirical evidence on the Dunning-Kruger effect. *Commun. Inf. Lit.* **2016**, *10*, 3. [CrossRef]

54. Leichner, N.; Peter, J.; Mayer, A.-K.; Krampen, G. Assessing information literacy among German psychology students. *Ref. Serv. Rev.* **2013**, *41*, 660–674. [CrossRef]

55. Markowski, B.; McCartin, L.; Evers, S. Meeting students where they are: Using rubric-based assessment to modify an information literacy curriculum. *Commun. Inf. Lit.* **2018**, *12*, 5. [CrossRef]

56. Association of College & Research Libraries [ACRL]. *Framework for Information Literacy for Higher Education*; American Library Association: Chicago, IL, USA, 2016; Available online: http://www.ala.org/acrl/standards/ilframework (accessed on 19 January 2022).

57. Hollis, H. Information literacy as a measurable construct: A need for more freely available, validated and wide-ranging instruments. *J. Inf. Lit.* **2018**, *12*, 76–88.

58. Catalano, A.J. *Streamlining LIS Research: A Compendium of Tried and True Tests, Measurements, and Other Instruments: A Compendium of Tried and True Tests, Measurements, and Other Instruments*; ABC-CLIO: Santa Barbara, CA, USA, 2016.

59. Mahmood, K. A systematic review of evidence on psychometric properties of information literacy tests. *Libr. Rev.* **2017**, *66*, 442–455. [CrossRef]

60. Vaterlaus, J.M.; Aylward, A.; Tarabochia, D.; Martin, J.D. "A smartphone made my life easier": An exploratory study on age of adolescent Smartphone acquisition and well-being. *Comput. Hum. Behav.* **2021**, *114*, 106563. [CrossRef]

61. Galera, M.D.C.G.; Muñoz, C.F.; Pedrosa, L.P. Youth empowerment through social networks. Creating participative digital citizenship. *Commun. Soc.* **2017**, *30*, 129–140. [CrossRef]

62. Cabezas-González, M.; Casillas-Martín, S.; Muñoz-Repiso, A.G.V. Basic Education Students' Digital Competence in the Area of Communication: The Influence of Online Communication and the Use of Social Networks. *Sustainability* **2021**, *13*, 4442. [CrossRef]

63. Kozík, T.; Slivová, J. Netiquette in electronic communication. *Int. J. Eng. Pedagog.* **2014**, *4*, 67–70. [CrossRef]

64. Soler-Costa, R.; Lafarga-Ostáriz, P.; Mauri-Medrano, M.; Moreno-Guerrero, A.J. Netiquette: Ethic, education, and behavior on internet—a systematic literatura review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1212. [CrossRef]

65. Brusco, J.M. Know your netiquette. *AORN J.* **2011**, *94*, 279–286. [CrossRef] [PubMed]

66. Hammond, L.; Moseley, K. Reeling in proper "netiquette". *Nurs. Made Incred. Easy* **2018**, *16*, 50–53. [CrossRef]

67. McMurdo, G. Netiquettes for networkers. *J. Inf. Sci.* **1995**, *21*, 305–318. [CrossRef]

68. Linek, S.B.; Ostermaier-Grabow, A. Netiquette between students and their lecturers on Facebook: Injunctive and descriptive social norms. *Soc. Media + Soc.* **2018**, *4*, 2056305118789629. [CrossRef]

69. Arouri, Y.M.; Hamaidi, D.A. Undergraduate Students' Perspectives of the Extent of Practicing Netiquettes in a Jordanian Southern University. *Int. J. Emerg. Technol. Learn.* **2017**, *12*, 84. [CrossRef]

70. Muñiz Fernández, J. *Introducción a la Teoría de Respuesta a los Ítems*; Repositorio Institucional de la Universidad de Oviedo: Pirámide, Mexico, 1997.

71. Baker, F.B.; Kim, S.H. (Eds.) *Item Response Theory: Parameter Estimation Techniques*; CRC Press: Boca Raton, FL, USA, 2004.

72. Hambleton, R.K.; Jones, R.W. Comparison of classical test theory and item response theory and their applications to test development. *Educ. Meas. Issues Pract.* **1993**, *12*, 535–556.

73. Wilson, M. *Constructing Measures: An Item Response Modeling Approach*; Routledge: London, UK, 2004. [CrossRef]

74. Rasch, G. *Probabilistic Models for Some Intelligence and Achievement Tests*; Danish Institute for Educational Research: Copenhagen, Denmark; MESA Press: Chicago, IL, USA, 1983.

75. Thissen, D. Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika* **1982**, *47*, 175–186. [CrossRef]

76. Hambleton, R.K.; Swaminathan, H.; Rogers, H.J. *Fundamentals of Item Response Theory*; Sage: Newcastle upon Tyne District, UK, 1991; Volume 2.

77. Reckase, M.D. Multidimensional Item Response Theory Models. In *Multidimensional Item Response Theory*; Springer: New York, NY, USA, 2009. [CrossRef]

78.   Adams, R.J.; Wilson, M.; Wang, W. The multidimensional random coefficients multinomial logit model. *Appl. Psychol. Meas.* **1997**, *21*, e23. [CrossRef]

79.   Adams, R.J.; Wu, M.L.; Wilson, M.R. *ACER ConQuest 3.0. 1*; Computer Software; Australian Council for Educational Research: Melbourne, Australia, 2012.

80.   Wright, B.D.; Stone, M.H. *Best Test Design*; Australian Council for Educational Research: Melbourne, Australia, 1979.

81.   Sandoval, W. Conjecture mapping: An approach to systematic educational design research. *J. Learn. Sci.* **2014**, *23*, 18–36. [CrossRef]

82.   Herrington, J.; McKenney, S.; Reeves, T.; Oliver, R. Design-based research and doctoral students: Guidelines for preparing a dissertation proposal. In Proceedings of the ED-MEDIA 2007—World Conference on Educational Multimedia, Hypermedia & Telecommunications 2007, Vancouver, BC, Canada, 25–29 June 2007; Montgomerie, C., Seale, J., Eds.; Association for the Advancement of Computing in Education (AACE): Vancouver, BC, Canada, 2007; pp. 4089–4097. Available online: https://www.learntechlib.org/primary/p/25967/ (accessed on 19 January 2022).

83.   McKenney, S.; Reeves, T.C. *Conducting Educational Design Research*; Routledge: London, UK, 2018. [CrossRef]

84.   Reeves, T. Design research from a technology perspective. In *Educational Design Research*; Routledge: London, UK, 2006; pp. 64–78.

85.   All Digital. Available online: https://all-digital.org/ (accessed on 19 January 2022).

86.   Bartolomé, J.; Garaizar, P.; Larrucea, X. A Pragmatic Approach for Evaluating and Accrediting Digital Competence of Digital Profiles: A Case Study of Entrepreneurs and Remote Workers. *Technol. Knowl. Learn.* **2021**, 1–36. [CrossRef]

87.   Bartolomé, J.; Garaizar, P.; Bastida, L. Validating item response processes in digital competence assessment through eye-tracking techniques. In Proceedings of the Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality 2020, Salamanca, Spain, 21–23 October 2020; pp. 738–746. [CrossRef]

88.   Articulate Storyline 360. Available online: https://articulate.com/360/storyline (accessed on 19 January 2022).

89.   Kzgunea. Available online: https://www.kzgunea.eus/es/inicio (accessed on 19 January 2022).

90.   All Digital Week. Available online: https://alldigitalweek.org/ (accessed on 19 January 2022).

91.   IT Txartela, Sistema de Certificación de Competencias Básicas en Tecnologías de la Información. Available online: http://www.it-txartela.net (accessed on 19 January 2022).

92.   Van Deursen, A.J.; Helsper, E.J.; Eynon, R. Development and validation of the Internet Skills Scale (ISS). *Inf. Commun. Soc.* **2016**, *19*, 804–823. [CrossRef]

93.   Krathwohl, D.R. A revision of Bloom's taxonomy: An overview. *Theory Pract.* **2002**, *41*, 212–218. [CrossRef]

94.   American Educational Research Association; American Psychological Association y National Council on Measurement in Education. *Standards for Educational and Psychological Testing*; American Educational Research Association: Washington, DC, USA, 2014.

95.   Mueller, R.O.; Knapp, T.R. Reliability and validity. In *The Reviewer's Guide to Quantitative Methods in the Social Sciences*; Routledge: London, UK, 2018; pp. 397–401.

96.   Bandalos, D.L. *Measurement Theory and Applications for the Social Sciences*; Guilford Publications: New York, NY, USA, 2018.

97.   Scholtes, V.A.; Terwee, C.B.; Poolman, R.W. What makes a measurement instrument valid and reliable? *Injury* **2011**, *42*, 236–240. [CrossRef]

98.   Varma, S.; Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinform.* **2006**, *7*, 91. [CrossRef]

99.   Wu, M.; Adams, R.J. Properties of Rasch residual fit statistics. *J. Appl. Meas.* **2013**, *14*, 339–355. [PubMed]

100.   Adams, R.J.; Khoo, S.T. *Quest*; ACER: Melbourne, Australia, 1996.

101.   Adams, R.J. Reliability as a measurement design effect. *Stud. Educ. Eval.* **2005**, *31*, 162–172. [CrossRef]

102.   Iglesias-Rodríguez, A.; Hernández-Martín, A.; Martín-González, Y.; Herráez-Corredera, P. Design, Validation and Implementation of a Questionnaire to Assess Teenagers' Digital Competence in the Area of Communication in Digital Environments. *Sustainability* **2021**, *13*, 6733. [CrossRef]

103.   Clifford, I.; Kluzer, S.; Troia, S.; Jakobsone, M.; Zandbergs, U. *DigCompSat. A Self-Reflection Tool for the European Digital Framework for Citizens (No. JRC123226)*; Joint Research Centre: Seville, Spain, 2020.