This is an Accepted Manuscript version of the following published article:

Garcia, C., Argelagós, E., & Privado, J. (2021). Assessment of higher education students' information problem-solving skills in educational sciences. Information Development, 37 (3), 359-375. https://doi.org/10.1177/0266666920976189.

Assessment of higher education students' information problem-solving skills in educational sciences

Many of the current tests that evaluate information problem-solving skills suffer from ecological validity weakness and from library-bias. The Procedural Information Problem-Solving Knowledge Evaluation in Education test (PIKE-E) aims to assess information problem-solving skills of college students in relation to an academic literature review task in educational sciences. It entails a confirmatory analysis of the PIKE-P test in which it is based. The PIKE-E was completed by 700 students from three different Spanish-speaking countries. In our research, we do not assume that information problem solving skills at the international and cultural level are equal, but Internet access is practically the same in Higher Education. Results show the existence of five first-order factors, Defining the research question, Search strategies planning, Searching and locating sources, Selecting and processing information and Organizing and presenting information, and a general single second-order factor, Information problem-solving, which coincide with contemporary theoretical models on information literacy. The PIKE-E can be used to tackle in which specific areas -concerning information skills development-, students entering education degrees need to improve to succeed in their studies.

Keywords: information problem-solving; skills; assessment; higher education, educational sciences

Introduction

Although many students regularly use the Internet and do not have problems with their technical skills, on many occasions they lack of critical thinking that helps to differentiate reliable and relevant information from other that is not. In the university context, the ability to correctly find and use specific information is important especially at master's and doctoral levels, where students need to use of a lot of bibliography and rigorous documentary sources to write their tasks following an academic standard. However, students' prior experience and practice of this ability appears to be limited when entering their university studies. Additionally, previous experience and practice of this skill differs remarkably between freshmen. This combination of deficiencies and weaknesses requires teachers to invest resources and time on information skills instruction (Lanning & Mallek, 2017; Varlejs & Stec, 2014). Albeit information specialists and librarians have been traditionally responsible for training in these skills, students still show

difficulties when evaluating information suitably, especially in the light of the specificity of knowledge and the necessary use of specialised databases. This has led professionals from other academic fields (i.e, medicine, psychology, or engineering) to develop instructional approaches in order to convey this skill in a much more specific way (Taylor, 2012). Our aim is to provide a test to assess substantial students' information problem-solving skills in educational sciences, in order to focus their subsequent training in a much-tailored way.

Information problem-solving skills assessment

Throughout the years, researchers have analysed this ability, breaking it down into different components or phases. In 2000, the Association of College and Research Libraries, ACRL, described five standards for information literacy assessment comprising: (a) the ability to determine the nature and extent of the information needed, (b) to access the information effectively and efficiently, (c) to evaluate sources and information critically and incorporate it, (d) to use the information for a certain purpose, and (e) to use the information according to economic, legal, and social issues. These standards were revised and updated in 2012 and 2016 to include ethical use and dissemination of information. Moreover, in their report for UNESCO concerning information literacy indicators, Catts and Lau (2008) outlined six of them: (a) definition and articulation of information needs, (b) location and access of information, (c) assessment of information, (d) organization of information, (e) use of information and communication, and (f) an ethical approach to information. No systematic and useful standards in Spanish-speaking countries to carry out the process of design and validation of information problem-solving assessment instruments have been published (Martínez et al., 2014).

From a theoretical perspective, most researchers agree that this ability involves handling information in general (Machin-Mastromatteo, 2012) and requires from students "to identify information needs, locate information sources, extract and organize information from each source, and synthesize information from a variety of sources" (Brand-Gruwel & Gerjets, 2008, p.3), with the aim of achieving a specific purpose (Garcia & Badia, 2017). It is an ability that recognises when an information is needed and acts subsequently to solve the problem (Leichner, Peter, Mayer & Krampen, 2013). Besides, this ability is not acquired neither by browsing the Web nor by using social networks (Badia & Becerril, 2015; Bannert & Reimann, 2011; Puustinen, Volckaert-Legrier, Coquin, & Bernicot, 2009). It requires instruction to find the proper information and a critical attitude to evaluate it (Argelagós & Pifarré, 2012;

Frerejean, Velthorst, van Strien, Kirschner, & Brand-Gruwel, 2019; Salmeron, Naumann, García, & Fajardo, 2015; Van Deursen & Van Diepen, 2013).

In order to evaluate these information literacy skills, instruments have been devised worldwide. Many of them were developed following the ACRL standards (Boh Podgornik, Dolnicar, Šorgo & Bartol, 2016; Emmet & Emde, 2007; Mittermeyer; 2005; O'Connor, Radcliff, & Gedeon, 2002; Salisbury & Karasmanis, 2011) and some focused on library-related issues (Mery, Newby, & Peng, 2011). Others adapted ACRL (2000) standards to specific languages or cultures (e.g., Al Aufi & Al Azri, 2013). In Figure 1, we present a conceptual map with a classification of the instruments collected in our bibliographic review. The instruments can be classified according to whether they assess the execution of the informational task (performance assessment) or individuals' given answers without putting into practice any informational task (reported assessment). The first type is a direct assessment since it looks at how individuals solve an information problem with a computer or other device. The process carried out is observed or tracked, and/or the product obtained is collected. The second type analyses responses to interviews and questionnaires as to what they would do or how they see themselves concerning an informational task.

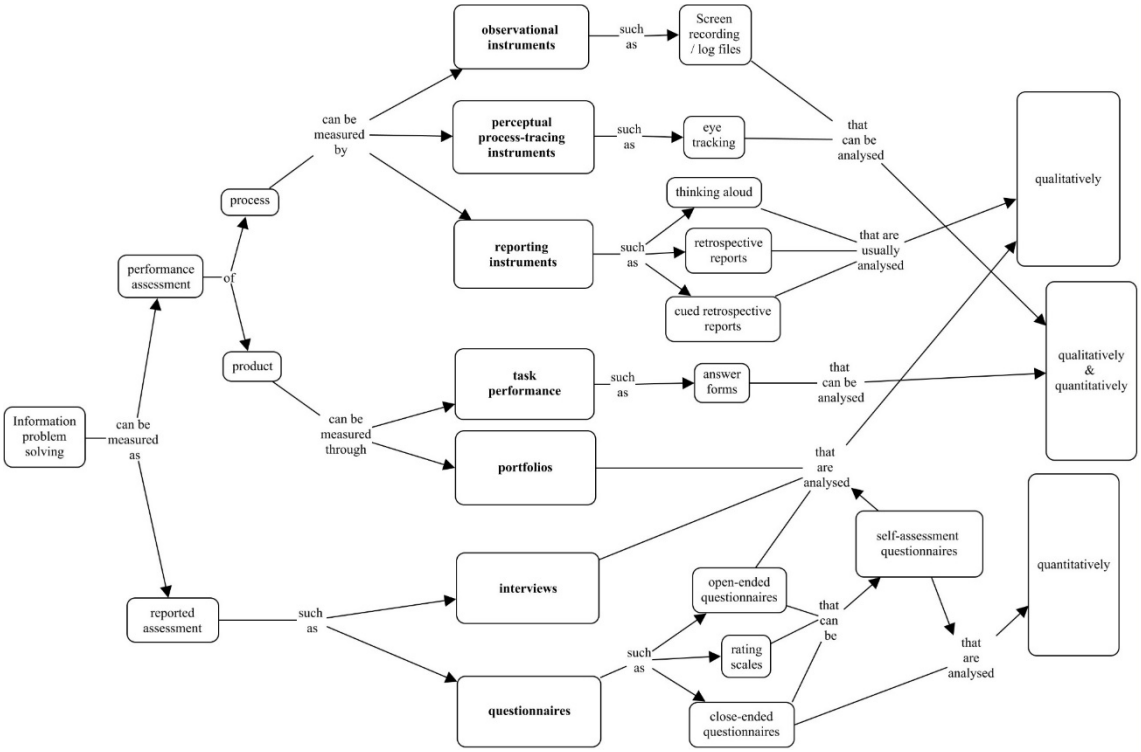

Figure 1. Types of instruments to assess information problem-solving

Concerning process assessment during performance, three main instruments have been used to evaluate IPS: observational, perceptual process-tracking and reporting instruments. In the observational instruments, an external observer captures the participants' behaviour and environment by means of field notes or video recordings of a classroom (e.g., Bregman, 2012) or using a screen recording software (i.e., Camstudio, Camtasia) which allows for a fine-grained analysis of actions taken during the process of solving an information problem (Argelagós, Brand-Gruwel, Jarodzka & Pifarré, 2018). This recording allows to analyse the time spent and the frequency of each action: search queries introduced by individuals in a search engine, results selected in a search engine results page (SERP), webpages visited, links selected from a webpage. Log files can unobtrusively record many participants at the same time. This instrument has already been used in many studies on IPS skills (e.g., Argelagós & Pifarré, 2012; Frerejean et al., 2019; Mason, Junyent, & Tornatora, 2014; van Deursen et al., 2014; Whitelock-Wainwright et al., 2020).

In perceptual process-tracing instruments, the eye movements' behaviour made by users are registered by an eye-tracking apparatus and provide information on what is visually attended to, in what order, and for how long (Holmqvist, Nyström, Andersson, Dewhurst, Jarodzka, & van de Weijer, 2011). Eye tracking is useful to analyse the visual exploration of a SERP during an IPS task (Şendurur & Yildirim, 2015; van Strien, Kammerer, Brand-Gruwel, & Boshuizen, 2016). As a limitation, the acquisition of an eye-tracking apparatus is expensive and it is time consuming, both to record and analyse.

Finally, the reporting instruments allow for valid inferences about the cognitive skills underlying task performance (Argelagós et al., 2018; Van Gog, Paas, Van Merriënboer, & Witte, 2005) as participants indicate what they have done after the task. Most widely used instruments are (1) think-aloud protocols, that consists of registering the utterances that the user is verbalising while solving a task (Ericsson and Simon, 1993), (2) retrospective reporting, in which participants report the thoughts while they are working on a task immediately after task performance (Van Gog, Paas, Van Merriënboer, & Witte, 2005), and (3) cued-retrospective reporting, in which participants are prompted with recordings of their task performance including recordings of their actions and eye movements (Hansen, 1991), and therefore the task performance is not influenced by the fact of verbalising their thoughts at the same time (Schmeck, Opfermann, Van Gog, Paas,& Leutner, 2015; Schwonke, Berthold & Renkl, 2009)

Concerning product assessment, the task performance can be analysed through: (1) answer forms, that are the answers given by the users while solving the task, which are collected or registered in order to apply a qualitative or quantitative analysis (e.g., Pifarré & Argelagós, 2020; Somerville et al., 2008; Walhout, Oomen, Jarodzka & Brand-Gruwel, 2017). And (2) portfolios, that are able to collect, store, and create working products in different formats as a result of the learning process of acquiring information literacy skills (e.g. Prastiwi, Kartowagiran, & Susantini, 2020; Scharf, Elliot, Huey, Briller & Joshi, 2007) and they have been used to assess literacy skills from a qualitative approach (e.g., Sharma, 2007; Sonley, Turner, Myer, & Cotton, 2007).

Reported assessment gathers interviews and questionnaires where individuals inform of what they would do or feel concerning an information task. Semi-structured interviews have been used to gain a deeper and more nuanced understanding of participants' feelings about information literacy experience (e. g., Detmering, McClellan, & Willenborg, 2019), and also to qualitatively complement other instruments to assess digital literacy (e.g., Hsu, Wenting, & Hughes, 2019).

Questionnaires collect information about participant's knowledge, beliefs, attitudes, and behaviour in an objective manner (Sapsford, 1999). They can be (1) open-ended (e.g. Thornton, 2006), (2) rating scales (i.e., Likert) (e.g. Kurbanoglu et al., 2006; Pinto, 2011; Carr et al., 2011; Booker et al, 2012), (e.g.) and (3) closed-ended questionnaires (e.g. Hsieh et al., 2013; Leichner et al., 2013; McKinney et al., 2011; Reed et al., 2007; O'Connor, Radcliff & Gedeon, 2001; Ondrusek et al., 2005; Staley et al., 2010). A special type of questionnaire, widely used to evaluate information literacy is the self-assessment or self-reporting (Pinto, 2010; Puteh & Ibrahim, 2010; Small & Snyder, 2010; Timmers & Glas, 2010). Finally, rubrics are considered a possible alternative to classical tests and questionnaires (Marzal, Solano, Vázquez, Muñoz, & Ros, 2011; Oakleaf, 2009).

In 2016, Rosman, Mayer and Krampen developed and validated a test to measure psychology students' information-seeking skills. Their theoretical framework combined the ACRL (2000) standards with the IPS skill decomposition approach of Brand-Gruwel et al. (2005). They used a situational judgment test format (Motowidlo, Hooper & Jackson, 2006) to avoid self-report measures due to possible bias in the scores. They believed that putting students before a certain situation that they must judge and answer based on how they would behave would be a way of controlling the problems of non-ecological tests or those that are too time-consuming, such as simulations or information tasks that involve searches and results. The Procedural Information-seeking Knowledge Evaluation – Psychology version test (PIKE-

P) was designed to measure psychology students' information-seeking skills in relation to declarative and procedural concepts. Results showed satisfactory content validity, acceptable reliability and high convergent validity. The test was also able to explain variance in information search task scores. However, they did not analyse the factorial structure of the test and assumed that all the items formed a single factor. Authors identified as main limitations their sole focus on information-seeking skills (hence omitting information evaluation), the possible deviation of the sample and the partial measurement of general psychology knowledge.

A large part of the instruments found in Spanish to evaluate the informational competencies of university students are self-assessment questionnaires (Carlos & Ramírez, 2017; Gisbert, Espuny & González, 2011; Velázquez, 2010), as the IL-HUMASS survey whose aim is to find out how students conceptualize their experiences and behaviours related to information (Pinto-Molina & Puertas-Valdeiglesias, 2012). Other instruments found are a semi-structured interview (Barberá et al., 2016), a combination of an objective test, ordinal scale and performance pills (Martínez Abad et al., 2014), a record of the actions carried out during the search task (González et la., 2013), a 3D tool, a simulator on digital competence (Mon & Cervera, 2013) and the application of untranslated English tests (iSkills, SAILS) to Spanish-speaking university students (Lau et al., 2016).

Our literature review has shown that multiple information problem-solving skills tests and surveys suffer a library-bias as many target the specific sub-skill of information search, as stated by Boh Podgornik et al. (2016). Library bias refers to the emphasis of various tests to evaluate information search-and-find competencies through catalogues, libraries, databases, etc. that underestimate the importance of other competencies necessary for solving information problems, such as the definition of the search object or problem, the evaluation of the search results, the processing of the information found, and their organization and presentation. What is generally common in these tests is that they focus their attention on criteria such as search strategies and keyword choice, generation and combination of search terms to satisfy the requirements of a particular research task, understanding and application of Boolean operators, use of scholarly databases and library catalogues or browsing of one or more resources to locate pertinent information. They are focused, in general, in items related to how the student uses search to find the information effectively and efficiently. Albeit, in their comparative study between experts and novices on the IPS process, Brand-Gruwel, Wopereis & Vermetten (2005) found that experts spent more time on the skills "defining the problem", "processing the information" and "presenting the information"

than to specific search skills. In this sense, the bias of many tests towards this search-and-find skill would make it difficult to assess the ability of the student to carry out the complete IPS process (or, in other words, the bibliographic review following each IPS skill) necessary to write an academic text in educational sciences. In addition, many tests may have become obsolete very quickly as technologies and tools evolve constantly. Besides, self-assessment type tests may over- or under-estimate the actual respective skills (Van Deursen & Van Dijk, 2010). In their research, Rosman et al (2015) showed the inconsistency of self-assessment tests as compared to achievement tests when evaluating information literacy.

Although using performance assessment instruments to evaluate information problem solving are recommended (Sparks, Katz & Beile, 2016) as they require students to perform real-life applications of knowledge and skills (Oakleaf, 2008), its creation, application and scoring cost is too high to be applied to hundreds or thousands of students. On the other hand, Spanish-speaking instruments are dismissed either because they are self-assessed questionnaires or expensive to apply in time or price. Of course, English ones have the language barrier to Spanish-speaking students. So, we decide to develop a quantitative instrument in Spanish, on-line, inexpensive, easy to apply, adjusted to the students' academic context, that includes a complex behaviour measurement and with a comprehensive perspective based on the IPS model of Brand-Gruwel, Wopereis and Vermetten (2005), by means of an adaptation and a confirmatory analysis of the PIKE -P test. In our research, we assume that although the context of technological development is not the same in Germany and in Spanish-speaking countries, Internet penetration is close to 70% in Latin-America and 91% in Spain and its use in Higher Education is close to 100%. For this reason, Spanish students have the same possibilities of searching for references in online databases as their German counterparts. The final aim of this research is to have a validated and adequate Spanish test to assess the ability of a student to carry out the review of the literature needed to write an academic text in educational sciences at the University.

Method

Item generation

Items were generated based on the PIKE-P test developed by Rosman et al. (2015). The translation process used in this study was defined according to the standard steps that anthropological, sociological and psychological literature recommends (e.g., Brislin, 1970; Van de Vijver & Leung, 1997). The first

step consisted of the initial translation from German into Spanish. The second step involved back-translation — that is, from Spanish back into German — which was carried out in such a way not to produce simple word-for-word equivalence and also to improve the cultural equivalence of the measure (Ardila, 2000; Sousa, & Rojjanasrirat, 2011). The third step involved adapting and making needed adjustments to the Spanish version, which included technical aspects and content. The students faced a test in which they had to choose between the decisions they would have to make when developing a bibliographic review task in educational sciences. In each item, three possible approaches of different instrumental suitability to face the corresponding decision were described, but only one was the correct one.

The instrument was applied to three independent samples from different Spanish-speaking countries (Spain, Colombia and Ecuador) in order to increase the generalizability of the results. The evidence of the test's internal validity and its factorial invariance was studied according to the nationality and gender of the participants.

Test adaptations

- Six questions were added, to cover the five IPS skills. The initial version contained 22 questions and covered three skills: Defining the research question, Search strategies planning, and Searching and locating sources. The six questions added in the Spanish version covered the fourth and the fifth skill: Selecting and processing information, and Organizing and presenting information. Altogether, the adapted version of the scale in the Spanish context had 28 items. The questions added were formulated according to the original test, which is, following a situational judgment test format.
- The number of response options for each question was reduced, from four in the original test to three. The response options were limited in the Spanish version in order to avoid excessively extending the test and, consequently, its duration at the time of its sitting. Our aim was to avoid having students grow weary and drop out of the test. The reduction in answer options did not affect the way students took the test.
- Each item corresponded to a different situation in the new version of the test: adequate situation (2 points), inadequate situation (0 points) and neutral situation (1 point), that is, neither very adequate nor very inadequate.

- Educational questions. The questions in the German version that related to psychology were reformulated to an educational context.
- Spanish databases accessible to participants. The PSYINDEX database (German) was replaced by the DIALNET database (in Spanish). On the other hand, the PSYCINFO database (paid) was replaced by the Academic Search Premier database (equivalent to PSYCINFO and with free access for participants). Similarly, the "thesaurus terms" (specific to the PSYCINFO database) were replaced by "subject terms" (specific to Academic Search Premier).
- On-site and virtual universities. The fact that universities can be face-to-face or virtual was taken into account. Brick-and-mortar universities have access to printed books, electronic journals and databases; their virtual equivalents have access to e-books and electronic articles, and database searches. Because students could use either physical or virtual libraries to conduct their research, in some questions and response items we specified whether the search engine related to a physical or virtual university.

Example of how PIKE-E evaluates IPS are seen in table 1. In order to guarantee that the questions could be well understood by the participants, a team of experts familiar with this field evaluated and revised each item, reaching a consensus whenever necessary.

Table 1. Examples of PIKE-E items

IPS skill	Example of item
Define the research question	*You are preparing a short presentation for a seminar. The title of the presentation is: «The efficacy of the therapeutic technique of flooding (flooding) or implosive therapy in arachnophobia»* *Which is more appropriate for the bibliographic research?* *A) Efficacy* *B) Arachnophobia* *C) Implosive*
Search strategies planning	*You want to make a presentation on the influence of leadership in group relationships. You have the following article:* *Barrow, J. C. (1975). An empirical framework of leadership effectiveness and investigation of leader subordinate-task causality relationship. Dissertation Abstract International, 35, 3631.* *What would you do to find more articles?* *A) I would extract the keywords from the article and plan to use them as a starting point for future searches (for example, in a specialized database).* *B) I would look for more articles by this author, as authors often publish multiple articles on the same topic.* *C) I would search for the bibliography of the article.*

Searching and locating sources	*You are looking for longitudinal studies on the efficacy of cognitive behaviour therapy in a specialized database. What do you do to miss as few studies as possible?* *A) I perform two searches for the keywords (thesaurus terms or "Subject Terms") "cognitive behaviour therapy" and "longitudinal study" and I link both searches using AND.* *B) I type «longitudinal cognitive behaviour therapy» in the search interface.* *C) I search for the keyword (thesaurus term or «Subject Terms») «cognitive behaviour therapy» and I search for «Longitudinal Empirical Study» in the field of the database that contains the information about the research methodology (Methodology). Then I link both searches using AND.*
Selecting and processing information	*After selecting it as interesting and useful, you have obtained an article on the subject you are dealing with for your final paper. How appropriate do you find the following ways of approaching it?* *A) After reading the title of the article, I read its «abstract» to see if I am really interested in its content.* *B) I analyse the bibliography of the article to see what studies it is based on.* *C) I read the entire article and underline it (if I have printed it) or comment on it (if I read it in pdf), in order to have the key ideas of it.*
Organizing and presenting information	*Once you have consulted several articles and prepared comments, tables and other notes, you are ready to write the theoretical framework of your final manuscript. How appropriate do you find the following procedures to do so?* *A) I begin by writing an introduction to the topic to be discussed, to organize my ideas.* *B) I review all the information that I have prepared and outline the structure of the theoretical framework (sections and content of each one approximately).* *C) I write the main characteristics of each article and present them in order of relevance in my theoretical framework.*

Data collection

A message was sent by email students from a fully online university, inviting them to participate in a pilot on information problem-solving skills, of which 850 students showed interested in taking part. In a synchronous-session webinar they were informed of the main researchers, the purpose of the test and about what their participation would entail. The test-taking process was also explained, as was how to withdraw their participation. By completing the online test, students voluntarily accepted to participate in the project. Ultimately, 700 students (82%) took the online test anonymously, 159 males and 541 females respectively. Their mean age was 36.42 (*SD*=8.073). Participants lived in Spain (20.1%), Colombia (36.1%), Ecuador (27.7%), and other countries (16.1%). The average completion time was 30 minutes. The process complied with data protection legal requirements.

Data analysis

First, we calculated the distribution of the items of the adapted test. Second, we studied the evidence of the test's internal validity by contrasting several validation models. Third, we studied the test's internal consistency. And finally, we studied the test's factorial invariance according to the origin of the sample and gender of the test-taker. These analyses were carried out with the statistical package SPSS V. 18.0 and with AMOS V. 7.0 (Arbuckle, 2006).

Confirmatory models require at least three measurement indicators to accurately estimate the latent factors, with a minimum of 100 participants, and 10 times the number of observed variables (Byrne, 2001). We used a sample of 700 participants and 28 indicators (items) for the model, i.e., 700/28= 25 participants per indicator. The procedure used to fit the models was maximum likelihood. The goodness of fit statistics used to evaluate the adequacy of the models were: 1) Absolute fit of the model to empirical data with the ratio $\chi 2/df$ (Bentler & Bonett, 1980), indicating a good fit with values less than 3, the Root Mean Square Error of Approximation (RMSEA) (Steiger, 1990) whose values below .05 indicate good fit, and the standardized residuals matrix (if there are few higher values $|\pm 1.96|$) (Byrne, 2001). 2) Incremental fit measures: Normed Fit Index (NFI) (Bentler & Bonett, 1980), Tucker Lewis Index (TLI) (Hu & Bentler, 2009) and Comparative Fit Index (CFI) (Bentler, 1990), if values above .95 indicate good fit, the empirical model is significantly different from the null model. 3) Parsimony fit measures evaluate the model fit versus the number of parameters estimated: Parsimony Goodness of Fit Index (PGFI) (Jöreskog & Sörbom, 1993) and Parsimony Normed Fit Index (PNFI) (James, Mulaik & Brett, 1982) whose values above .50 indicate good fit.

Results

Descriptive. Table 2 shows the descriptive statistics and the asymmetry and kurtosis of the items of the adapted test. According to West, Finch and Curran (1995), an asymmetry value less than 2 and a kurtosis less than 7, would be adequate in order to use estimation procedures of Maximum Likelihood, since they are robust to small variations in normality. These conditions are met in the item data.

Table 2. Descriptive statistics, asymmetry and kurtosis of the test items. Initial and final biserial-point correlation.

Items	Mean	SD	Asymmetry	Kurtosis	r biserial-point 1	r biserial-point 2
1.	0.99	1.00	0.41	-1.21	.198	.209

2.	1.11	1.03	0.39	-1.10	.346	.345
3.	1.02	0.99	0.42	-1.10	.383	.387
4.	0.83	0.86	0.60	-0.75	.234	.231
5.	0.80	0.97	0.84	-0.58	.355	.355
6.	1.55	1.09	-0.35	-1.23	.540	.547
7.	0.90	0.91	0.41	-1.21	.352	.353
8.	1.13	0.99	0.25	-1.17	.358	.354
9.	1.00	0.95	0.36	-1.15	.143	
10.	1.13	0.88	0.10	-1.04	.311	.292
11.	1.18	0.78	0.85	0.66	.105	
12.	1.28	0.81	0.84	0.24	.144	.128
13.	1.21	1.05	0.17	-1.30	.418	.423
14.	1.07	0.88	-0.06	-1.53	.544	.536
15.	1.65	0.99	-0.54	-0.80	.632	.629
16.	1.65	0.88	0.35	-1.03	.312	.325
17.	1.48	0.93	0.46	-0.82	.279	.288
18.	1.35	0.70	0.78	0.42	.243	.246
19.	0.88	0.93	0.56	-0.96	.326	.324
20.	0.78	0.78	0.59	-0.56	.348	.342
21.	0.63	0.76	0.90	-0.17	.227	.226
22.	1.03	0.91	0.26	-1.15	.361	.367
23.	0.92	0.94	0.59	-0.81	.224	.224
24.	1.42	0.96	0.52	-0.78	.156	.157
25.	1.05	1.02	0.37	-1.19	.415	.421
26.	1.05	1.00	0.35	-1.17	.364	.373
27.	1.20	1.01	0.15	-1.24	.451	.459
28.	1.82	0.88	-0.96	0.29	.533	.534

Evidence of internal validity. Two theoretical models were tested in order to study the test's internal structure: a general factor model that collects the variance of the different items and a hierarchical model with five first order factors that are grouped into a general one. To adjust the hierarchical model, it was necessary to eliminate items 9 and 11 because their factorial weight with the factors was less than .20, and it was not clear at a theoretical level which factor of the first order they should belong to. Eliminating number 9 did not introduce any risk because it was sufficiently similar to number 10. Similarly, we decided to eliminate item 11 because question 12 addressed the same topic, even more thoroughly. Item 24 also had a factorial weight of .20; however, this item was not eliminated because it covered a necessary topic of the fourth factor of the model (Selection-Processing). Table 2 shows the goodness indices of the adjustment of the two models, being better for the contrasted hierarchical model. In general, goodness of fit indexes is adequate except for the NFI and the TLI.

Internal consistency. We calculated the internal consistency of the final test. Table 3 shows the corrected biserial-punctual correlation of each item with the total of the initial test (all items) and final (after eliminating items 9 and 11). The initial internal consistency (Cronbach's alpha) of the test was .817 and the final .822. As can be seen, items 9 and 11 had correlation values lower than .20 with the total of the

test, which justifies their elimination (Abad, Olea, Ponsoda, & García, 2011). After eliminating these items, the internal consistency increases and there remain only two items that have values below .20 of the biserial-point correlation, proving that the internal consistency improves considerably by eliminating these two items.

Table 3. Goodness-of-fit indices of the two contrasted models.

Model	Chi^2/df	GFI	NFI	CFI	TLI	PGFI	PNFI	RMSEA	Residuals >|± 1.96|
One factor	1.922	.934	.765	.870	.860	.805	.708	.036	5.67%
Hierarchical	1.743	.946	.810	.908	.899	.795	.735	.033	2.28%

Figure 2 shows the hierarchical model contrasted with the different factorial weights. As can be seen, first-order factors saturate in most of the items in medium-high mode (between .17 and .75). Resulting from the items that configured the first-order factors, the following skills are identified: Defining the research question, Search strategies planning, Searching and locating sources, Selecting and processing information and Organizing and presenting information, according to previous literature review on information behaviour. In the case of the general factorial weights for second order factors, they present very high values (≥ .90), which justifies this factor at an empirical level, the Information problem-solving process.

Factorial invariance. Finally, the presence of factorial invariance was studied according to the participants' gender and origin (nationality) with a view to whether the factor structure of the hierarchical model was similar in these groups. The results obtained are shown in Table 4.

The Akaike (1987) information criterion (AIC) indicates which model has better fit when comparing them to each other, the lower the AIC the better fit. For both gender and nationality, the best model is D. If we compare the four measurements with each other, we can see that according to the $\chi 2$ test, the differences between the models are not statistically significant. Following the criterion of Cheung and Rensvold (2002) a $\Delta CFI < .01$ would indicate strict invariance according to gender: the structural weights, the variance-covariance matrix and the variances errors are equal in both genders. This criterion is satisfied with regard to gender. With regard to nationality, the ΔCFI was indeed < .01 when comparing models B and C, so we can say that there is strong metric invariance; that is, the structural weights and the variance-covariance matrix are equal for the three nationalities.

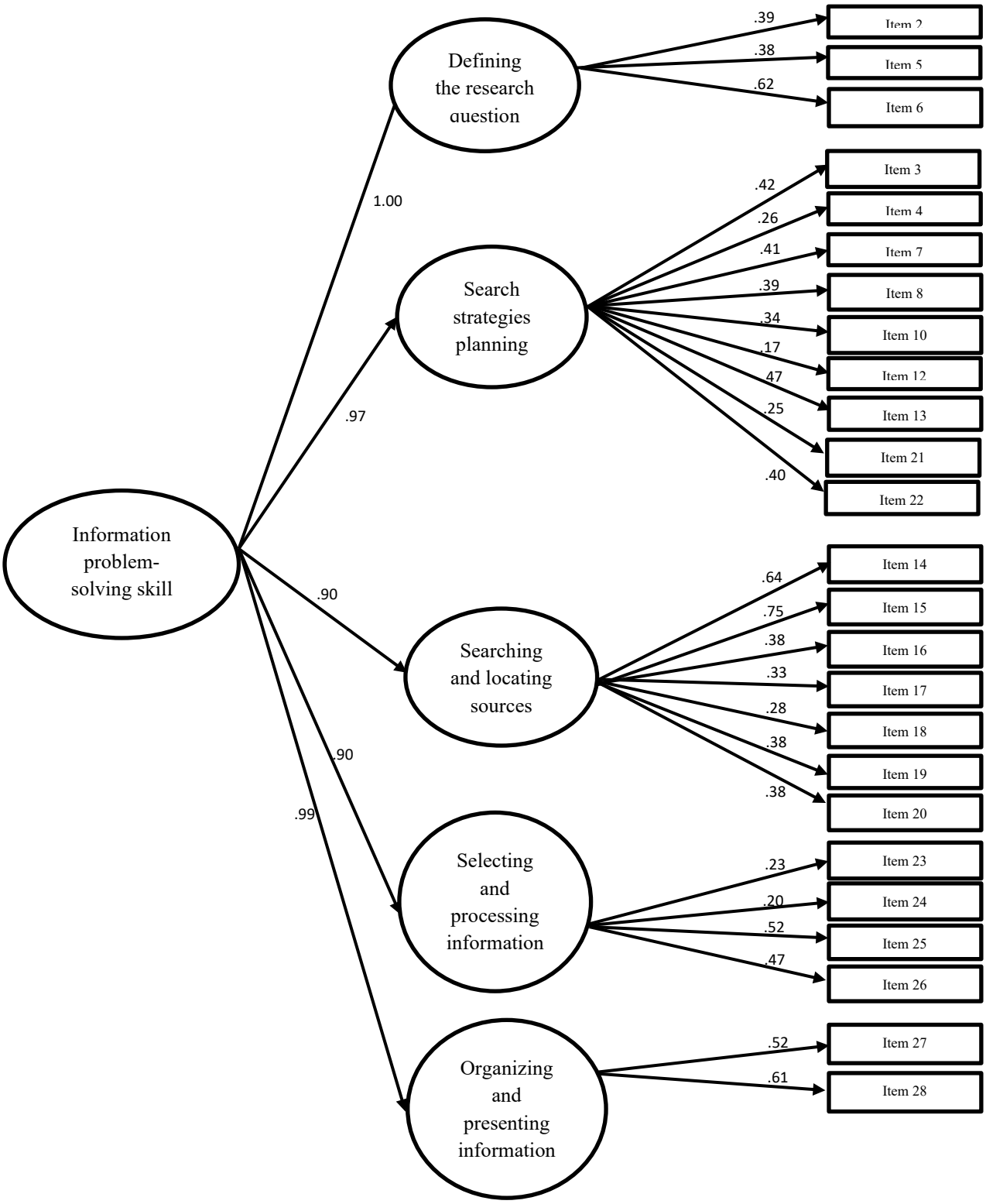

Figure 2. Hierarchical model contrasted

Table 4. Goodness-of-fit indices of the two factorial invariance models according to gender and nationality.

Models for gender	χ^2	*df*	χ^2/df	**GFI**	**CFI**	**RMSEA**	**AIC**
Model A. Unconstrained	29.690	10	2.969	.947	.960	.100	69.690
Model B. Equal structural weights	36.321	14	2.594	.933	.954	.090	68.321
Model C. Equal structural covariances	38.309	15	2.554	.931	.952	.089	68.309
Model D. Equal measurement residuals	41.604	20	2.080	.928	.956	.074	61.604
Comparison of models	$\Delta\chi^2$	Δdf	*p*		**ΔCFI**		
Models A and B (Metric invariance)	6.631	4	.843		.006		
Models B and C (Strong metric invariance)	1.988	1	.841		.002		
Models C and D (Strict metric invariance)	3.295	5	.345		.004		
Models for nationality	χ^2	*df*	χ^2/df	**GFI**	**CFI**	**RMSEA**	**AIC**
Model A. Unconstrained	1281.84	885	1.448	.861	.802	.028	1617.84
Model B. Equal structural weights	1342.43	937	1.433	.855	.798	.027	1574.43
Model C. Equal structural covariances	1352.35	945	1.431	.854	.797	.027	1568.35
Model D. Equal measurement residuals	1447.23	997	1.452	.846	.776	.028	1559.23
Comparison of models	$\Delta\chi^2$	Δdf	*p*		**ΔCFI**		
Models A and B (Metric invariance)	39.41	52	.100		.004		
Models B and C (Strong metric invariance)	9.92	8	.729		.001		
Models C and D (Strict metric invariance)	94.88	52	.999		.021		

Discussion

The aim of this research was to develop an easy-to-complete assessment tool, adapted to the context at which it was aimed, that meets the necessary reliability and validity criteria and that can assess the level of each information problem solving skill that students show when they need to carry out the review of the literature needed to write an academic text in educational sciences at the University. A main concern was to avoid the library biases that generalist tests show. The adaptation of a validated test such as the PIKE-P involved a double translation of the test (German-Spanish-German), the inclusion of six new items to analyse skills not presented originally and an adjustment to the educational context, all of which shapes the Procedural Information problem-solving Knowledge Evaluation-Education test (PIKE-E).

Therefore, a new statistical analysis of the internal validity of the test and its consistency was made. In order to facilitate students to complete it, the option of pairs of suitable possible answers from the initial test was dismissed, leaving three alternatives: adequate, neutral and inadequate. The final scale consisted of 28 items with the aim of covering the five skills included in the Brand-Gruwel et al model. (2005), therefore, the definition of the problem, the search for the information, its initial evaluation, the processing of the information and finally its elaboration. In the current test, we also used the situational judgment approach proposed by the authors of the PIKE-P to avoid ecological problems.

The results of our study are relevant for several reasons. First, they clearly establish five factors that evaluate the student's ability to solve an information problem related to the use of scientific educational documentation for an academic text. Besides, the presence of a single factor indicates that all the items in the validation study are included under the same general concept, the Information problem-solving process, as the different theoretical models on informational literacy and informational behaviour stand up for (Leichner, Peter, Mayer & Krampen, 2013; Machin-Mastromatteo, 2012). The specific skills identified by this test are therefore measurable and can be trained. Finally, the originality of the research lies in the international reliability analysis that the PIKE-P has shown in a context and in a language different from the original, avoiding ecological problems and library bias.

The internal validity of the scale showed a high index of adjustment to a hierarchical model with five first order factors and one of the second order. Likewise, the final reliability of the scale is high (α = 822) after eliminating two items. Thus, the factor weight data found in the questionnaire are acceptable in the first order factors and very high in the second order factor. The Defining the research question factor indicates the need for students to decide how their activity should start (Eisenberg and Berkowitz, 1990) and what are the initial terms of their search, that is, to identify what is needed (Argelagós and Pifarré, 2016; Brand-Gruwel et al., 2005; Frerejean et al., 2019) to start finding relevant scientific information in education. For example, in one of the items the student must decide which is the correct sequence of steps to undertake among: reading a task, identifying important concepts, using terms for searches, or linking related terms for search.

The second factor found in the test is Search strategies planning, that is, what the student thinks should be done to achieve a good result as soon as possible. We should not forget that students are offered different alternatives from which they must choose the one considered most appropriate. Hence, these items raise questions such as: what do you do to find more articles; what do you do to find more

educational bibliography; how would you rate the following modifications to the search query; how appropriate do you think the following tools are to find a certain article.

The third factor, Searching and locating sources, involves specific actions, that is, the application of searches (Rosman et al., 2015) to locate the information requested. These items focus on the search action itself, such as: what do you do to find the article as quickly as possible; do you want to use a specific database and how do you do it; or how appropriate do the following search queries seem to you which you could enter in the search interface. Therefore, it focuses on the application of generic or specific key terms in education, or the use of Booleans to locate the required information.

The next factor, Selecting and processing information brings together those items that evaluate how the student processes and values the information found (Catts & Lau, 2008). And once the results are obtained, to what extent a student is able to identify what is relevant from what is not, what procedure should be used to collect the information, and how certain ways of approaching the information needed are appropriate. The items propose questions such as: after conducting a search in a specialized database and based on the results obtained, what result do you choose according to your objective; what procedure do you follow to save the results that are useful to you; or considering the revision made, how suitable do you consider the following procedures to get with the information. Finally, the Organizing and presenting information factor raises questions related to the elaboration of a theoretical framework and how appropriate the procedure to write it is or, from what has already been written, how to proceed with the revision, reading and correction of texts. This factor is clearly identified with the ability to organize and present information (Brand-Gruwel et al., 2005).

Finally, the factor invariance is highlighted. This finding is important concerning possible linguistic differences between countries. The factor structure of the hierarchical model remains unchanged in the three Spanish-speaking countries and no significant differences are found at the gender level. Thus, the validity of the test for the linguistic and educational field to which it is directed is supported. The test, has shown a good reliability and an adequate internal consistency to measure students' information problem-solving skills regarding a bibliographic review task in educational sciences at the university.

Limitations and suggestions for future work

Although our study provides insights into assessing information problem-solving skills, the results need to be interpreted with caution, due to various drawbacks of this study.

On one hand, the elements considered in our research were collected by means of a single method. Our results are reliant on hypothetical situations where participants chose the best option that they believed showed how they would act. In further studies, these assessments could be complemented by objective data regarding performance, which could be recorded by log files, eye-tracking apparatus, or even by an integration of several techniques (Argelagós, Brand-Gruwel, Jarodzka, & Pifarré, 2018). On the second hand, data were collected completely online, which could raise some concern about their validity. Even so, studies like those of Casler, Bickel, and Hackett (2013) concluded that there should not be any differences with data collected in person. Additionally, the sample covered students from a single, private university and the imbalance in the sample's male - female ratio should also be taken into account. Future work could extend the sample across other universities or higher education contexts. However, given that the test was previously validated in another university (face-to-face), in another context (Psychology) and in another language (German) (Rosman, et al., 2015), we believe that it does not necessarily represent a significant limitation, given that our study largely corroborated the previous one's results within another type of university (online) and in another language (Spanish).

The results of this study open new lines of research, which would be of great interest to pursue. For instance, the test may contribute to the design and evaluation of a program to improve higher education students' skills to use digital information for carrying out academic tasks in educational sciences. As stated by Emmett and Emde (2007) measuring instruments can provide information for decision making on training content. Moreover, subsequent development of IPS skills could be measured by the PIKE-E and, due to the test's large scope and its good psychometric properties we expect such an approach to be of particularly high value.

References

Abad FJ, Olea J, Ponsoda V and García C (2011) *Medición en ciencias sociales y de la salud.* Madrid: Síntesis.

Akaike H (1987) Factor Analysis and AIC. *Psychometrika* 52: 317–332. DOI:10.1007/978-1-4612-1694-0_29

Al-Aufi and Al-Azri H (2013) Information literacy in Oman's higher education: A descriptive-inferential approach. *Journal of Librarianship and Information Science* 45(4): 335-346. DOI: 10.1177/0961000613486824

American Library Association (2000). Information literacy competency standards for higher education. Available at: hdl.handle.net/11213/7668

American Library Association (2012). Information literacy competency standards for higher education. Available at http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/ils_recomm.pdf

American Library Association (2016). Framework for information literacy for higher education. Available at http://www.ala.org/acrl/sites/ala.org.acrl/files/content/issues/infolit/Framework_ILHE.pdf

Arbuckle JL (2006). *Amos 7.0 User's Guide*. Chicago: SPSS.

Ardila A (2000). Testing Hispanic populations. *Texas Psychologist,* 1: 25-29.

Argelagós E and Pifarré M (2012). Improving information problem solving skills in secondary education through embedded instruction. *Computers in Human Behavior* 28(2): 515-526. DOI:10.1016/j.chb.2011.10.024

Argelagós E, Brand-Gruwel S, Jarodzka HM and Pifarré M (2018) Unpacking cognitive skills engaged in web-search: how can log files, eye movements, and cued-retrospective reports help? An in-depth qualitative case study. *International Journal of Innovation and Learning* 24(2): 152-175. DOI:10.1504/IJIL.2018.094069

Argelagós E and Pifarré M (2016) Key information-problem solving skills to learn in secondary education: A qualitative, multi-case study. *Journal of Education and Learning* 5(4): 1-14. DOI:10.5539/jel.v5n4p1

Badia A and Becerril L (2015) Collaborative solving of information problems and group learning outcomes in secondary education. *Journal for the Study of Education and Development* 38(1): 67-101. DOI: 0.1080/02103702.2014.996403

Barberá E, Quintana J, Galván C and Illera, JLR (2016) El portafolio electrónico como facilitador de una competencia informacional diversificada y reflexiva. *Textos universitaris de biblioteconomia i documentació* 36.

Bannert M and Reimann P (2011) Supporting self-regulated hypermedia learning through prompts. *Instructional Science* 40(1): 193-211. DOI: 10.1007/s11251-011-9167-4.

Bentler P M (1990) Comparative fit indexes in structural models. *Psychological Bulletin* 107: 238-246.

Bentler B M and Bonett D G (1980) Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* 88: 588-606. DOI: 10.1037/0033-2909.88.3.588

Booker L D, Detlor B and Serenko A (2012) Factors affecting the adoption of online library resources by business students. *Journal of the American Society for Information Science and Technology* 63(12): 2503–2520.

Boh Podgornik,B, Dolničar D, Šorgo A and Bartol T (2016) Development, testing, and validation of an information literacy test (ILT) for higher education. *Journal of the association for Information Science and Technology* 67(10): 2420-2436. DOI: 10.1002/asi.23586

Brand-Gruwel S, Wopereis I and Vermetten Y (2005) Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior* 21(3): 487-508. DOI: 10.1016/j.chb.2004.10.005

Brand-Gruwel S and Gerjets P (2008) Instructional support for enhancing students' information problem solving ability. *Computers in Human Behavior* 24(3): 615-622. DOI: 10.1016/j.chb.2004.10.005

Bregman RL (2012) Measuring and motivating student effort in an online version of the core class in operations management. *International Journal of Innovation and Learning* 11(4): 369-385. DOI: 10.1504/IJIL.2012.047138

Brislin RW (1970) Back-Translation for Cross-Cultural Research. *Journal of Cross-Cultural Psychology* 1(3): 185 216. DOI: 10.1177/135910457000100301

Byrne B M (2001) *Structural equation modeling with AMOS basic concepts, applications, and programming*. New Jersey: Lawrence Erlbaum.

Carlos P A and Ramírez I U (2017) Evaluación de la habilidad digital de los estudiantes universitarios: estado de ingreso y potencial educativo. *EDUTEC. Revista Electrónica de Tecnología Educativa* (61): 1-13.

Carr S, Iredell H, Newton-Smith C and Clark C (2011) Evaluation of Information Literacy Skill Development in First Year Medical Students. *Australian Academic and Research Libraries* 42(2): 136–148.

Casler K, Bickel L and Hackett E (2013) Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in human behavior* 29(6): 2156-2160. DOI: 10.1016/j.chb.2013.05.009

Catts R.and Lau J (2008) *Towards information literacy indicators*. Paris: UNESCO. Available at: hdl.handle.net/1893/2119

Cazco GHO, González MC, Abad FM, Altamirano JED and Mazón MES (2016) Determining factors in acceptance of ICT by the university faculty in their teaching practice. In: *Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality,* Salamanca, Spain 20-23 November 2016, pp. 139-146. DOI: 10.1145/3012430.3012509

Chan BS, Churchill D and Chiu T K (2017) Digital literacy learning in higher education through digital storytelling approach. *Journal of International Education Research* 13(1): 1–16. DOI:10.19030/jier.v13i1.9907

Cheung GW and Rensvold RB (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling* 9: 233–255. DOI: 10.1207/S15328007SEM0902_5

Detmering, R., McClellan, S., & Willenborg, A. (2019). A Seat at the Table: Information Literacy Assessment and Professional Legitimacy. *College & Research Libraries 80*(5), 720.

Eisenberg MB and Berkowitz RE (1990) *Information Problem Solving: The Big Six Skills Approach to Library & Information Skills Instruction*. Norwood: Ablex Publishing Corporation.

Emmett A and Emde J (2007) Assessing information literacy skills using the ACRL standards as a guide. *Reference Services Review* 35(2): 210-229. DOI: 10.1108/00907320710749146

Frerejean J, Velthorst GJ, van Strien JL, Kirschner PA and Brand-Gruwel S (2019) Embedded instruction to learn information problem solving: Effects of a whole task approach. *Computers in Human Behavior* 90: 117-130. DOI: 10.1016/j.chb.2018.08.043

Garcia C and Badia A (2017) Information problem-solving skills in small virtual groups and learning outcomes. *Journal of Computer Assisted Learning* 33(4): 382-392. DOI:10.1111/jcal.12187

Gisbert M, Espuny C and González J (2011) INCOTIC. Una herramienta para la @utoevaluación diagnóstica de la competencia digital en la universidad. *Profesorado, Revista de currículum y formación del profesorado* 15 (1): 75-90. Available at: https://www.redalyc.org/pdf/567/56717469006.pdf

Gilster P (1997) *Digital literacy*. New York: John Wiley & Sons.

González L, Marciales G, Castañeda-Peña H, Barbosa-Chacón J and Barbosa J (2013) Competencia informacional: desarrollo de un instrumento para su observación. *Lenguaje* 41(1): 105-131.

Hansen JP (1991) The use of eye mark recordings to support verbal retrospection in software testing. *Acta Psychologica* 76: 31-49. DOI: 10.1016/0001-6918(91)90052-2

Holmqvist K, Nyström N, Andersson R, Dewhurst R, Jarodzka H and van de Weijer J (2011) *Eye tracking: a comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.

Hsieh M L, Dawson P H and Carlin M T (2013) What Five Minutes in the Classroom Can Do to Uncover the Basic Information Literacy Skills of Your College Students: A Multiyear Assessment Study. *Evidence Based Library and Information Practice* 8(3): 34–57.

Hu LT and Bentler PM (2009) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 6(1): 1–55. DOI10.1080/10705519909540118

Hsu HP, Wenting Z and Hughes JE (2019) Developing elementary students' digital literacy through augmented reality creation: Insights from a longitudinal analysis of questionnaires, interviews, and projects. *Journal of Educational Computing Research* 57(6): 1400-1435.

Hwang GJ and Kuo FR (2011) An information-summarising instruction strategy for improving the web-based problem solving abilities of students. *Australasian Journal of Educational Technology* 27(2).

Hupfer ME.and Detlor B (2006) Gender and web information seeking: A self-concept orientation model. *Journal of the American Society for Information Science and Technology* 57(8): 1105-1115. DOI: 10.1002/asi.20379

Janssen J, Stoyanov S, Ferrari A, Punie Y, Pannekeet K and Sloep P (2013) Experts' views on digital competence: Commonalities and differences. *Computers & Education* 68: 473-481. DOI: j.compedu.2013.06.008

James LR, Mulai, S A and Brett JM (1982) *Causal analysis: models, assumptions and data*. Beverly Hills, CA: Sage.

Jöreskog KG and Sörbom D (1993) *LISREL 8: user's guide*. Chicago: Scientific Software International.

Katz IR (2007) Testing information literacy in digital environments: ETS's iSkills assessment. *Information technology and Libraries*, 26(3): 3-12. DOI: 10.6017/ital.v26i3.3271

Kurbanoglu, S. S., Akkoyunlu, B., & Umay, A. (2006). Developing the information literacy self-efficacy scale. *Journal of Documentation* 62(6): 730-743.

Lanning and Mallek J (2017) Factors influencing information literacy competency of college students. *The Journal of Academic Librarianship* 43(5): 443-450. DOI: 10.1016/j.acalib.2017.07.005

Lau J, Machin-Mastromatteo JD, Gárate A and Tagliapietra-Ovies AC (2016) Assessing Spanish-speaking university students' info-competencies with iSkills, SAILS, and an In-House instrument: challenges and benefits. In *European Conference on Information Literacy* (pp. 327-336). Springer, Cham.

Leichner N, Peter J, Mayer AK and Krampen G (2013) Assessing information literacy among German psychology students. *Reference Services Review* 41(4): 660-674. DOI: 10.1108/RSR-11-2012-0076

Lindroth T and Bergquist M (2010) Laptopers in an educational practice: Promoting the personal learning situation. *Computers & Education* 54(2): 311–320. DOI: 10.1016/j.compedu.2009.07.014

Machin-Mastromatteo JD (2012) Participatory action research in the age of social media: Literacies, affinity spaces and learning. *New Library World* 113 (11/12): 571-585. DOI:10.1108/03074801211282939

Martínez Abad F, Bielba Calvo M, Olmos Migueláñez S and Rodríguez-Conde M J (2014) Diseño y validación de un instrumento para la evaluación del nivel real en competencias informacionales para el futuro profesorado de educación secundaria. *Instituto Universitario de Ciencias de la Educación – Universidad de Salamanca.*

Marzal Martínez FJ, Solano Fernández JP, Vázquez G, Muñoz JP and Ros J (2011) Desarrollo y evaluación de la competencia gestión de la información en titulaciones técnicas: estudio de casos. In *Congreso Internacional de Innovación Docente*, Cartagena (España).

Mason L, Junyent AA and Tornatora MC (2014) Epistemic evaluation and comprehension of web-source information on controversial science-related topics: Effects of a short-term instructional intervention. *Computers and Education* 76: 143-157. DOI: 10.1016/j.compedu.2014.03.016

Mery Y, Newby J and Peng K (2011) Assessing the reliability and validity of locally developed information literacy test items. *Reference Services Review* 39(1): 98-122. DOI: 10.1108/00907321111108141

Mittermeyer D (2005) Incoming first year undergraduate students: How information literate are they? *Education for information* 23(4): 203-232. DOI: 10.3233/EFI-2005-23401

Motowidlo SJ, Hooper AC and Jackson HL (2006) A Theoretical Basis for Situational Judgment Tests. In J. A. Weekley & R. E. Ployhart (Eds.), *SIOP organizational series. Situational judgment tests:*

Theory, measurement, and application. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, pp. 57-81.

Mon FE and Cervera MG (2013) Competencia digital en la educación superior: instrumentos de evaluación y nuevos entornos. *Enl@ ce: Revista Venezolana de Información, Tecnología y Conocimiento,* 10(3): 29-43.

Oakleaf M (2009) Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the American Society for Information Science and Technology* 60(5): 969-983. DOI: 10.1002/asi.21030

O'Connor LG, Radcliff CJ and Gedeon JA (2002) Applying systems design and item response theory to the problem of measuring information literacy skills. *College & Research Libraries* 63(6): 528-543. DOI:10.5860/crl.63.6.528

Ondrusek A, Dent VF, Bonadie-Joseph I and Williams C (2005) A longitudinal study of the development and evaluation of an information literacy test. *Reference Services Review* 33(4): 388–417.

Pifarré M and Argelagós E (2020) Embedded Information Problem Solving Instruction to foster learning from digital sources: longitudinal effects on task-performance. *Sustainability* 12(19): 7919.

Pinto M (2010) Design of the IL-HUMASS survey on information literacy in higher education: A self-assessment approach. *Journal of information science* 36(1): 86-103.

Pinto, M (2011) An Approach to the Internal Facet of Information Literacy Using the ILHUMASS Survey. *The Journal of Academic Librarianship* 37(2): 145–154.

Pinto-Molina M and Puertas-Valdeiglesias S (2012) Autoevaluación de la competencia informacional en los estudios de Psicología desde la percepción del estudiante. In *Anales de Documentación* 15(2): 1-15. Murcia: Servicio de Publicaciones Universidad de Murcia. DOI: 10.6018/analesdoc.15.2.151661

Prastiwi M, Kartowagiran B and Susantini E (2020) Assessing Using Technology: Is Electronic Portfolio Effective To Assess the Scientific Literacy on Evolution Theory. *International Journal of Emerging Technologies in Learning (iJET)* 15(12): 230-243. Retrieved from https://www.learntechlib.org/p/217532/

Puteh M and Ibrahim M (2010) The usage of self-regulated learning strategies among form four students in the Mathematical problem-solving context: A case study. *Procedia-Social and Behavioral Sciences* 8: 446-452.

Puustinen M, Volckaert-Legrier O, Coquin D and Bernicot, J (2009) An analysis of students' spontaneous computer-mediated help seeking: A step toward the design of ecologically valid supporting tools. *Computers & Education* 53: 1040-1047. DOI: 10.1016/j.compedu.2008.10.003

Reed MJ, Kinder D and Cecile F (2007) Collaboration between Librarians and Teaching Faculty to Teach Information Literacy at One Ontario University: Experiences and Outcomes. *Journal of Information Literacy* 1(3): 1–19.

Rosman T, Mayer A.K and Krampen G (2015) Measuring psychology students' information-seeking skills in a situational judgment test format: Construction and validation of the PIKE-P Test. *European Journal of Psychological Assessment* 32: 220-229. DOI: 10.1027/1015-5759/a000239

Rosman T, Mayer AK and Krampen G (2015) Combining self-assessments and achievement tests in information literacy assessment: Empirical results and recommendations for practice. *Assessment & Evaluation in Higher Education* 40(5): 740–754. DOI: 10.1080/02602938.2014.950554

Salisbury F.and Karasmanis S (2011) Are they ready? Exploring student information literacy skills in the transition from secondary to tertiary education. *Australian Academic & Research Libraries* 42(1): 43-58. DOI: 10.1080/00048623.2011.10722203

Salmerón L, Naumann J, García V and Fajardo I (2017) Scanning and deep processing of information in hypertext: an eye tracking and cued retrospective think-aloud study. *Journal of Computer Assisted Learning* 33(3): 222-233. DOI: 10.1111/jcal.12152

Sapsford R (1999) *Survey research*. London, UK: Sage.

Şendurur E and Yildirim Z (2015) Students' web Search strategies with different task types: An eye-tracking study. *International Journal of Human-Computer Interaction* 31(2): 101-111. DOI: 10.1080/10447318.2014.959105

Sharma S (2007) From Chaos to Clarity: Using the Research Portfolio to Teach and Assess Information Literacy Skills. *Journal of Academic Librarianship* 33(1): 127–35. DOI: 10.1016/j.acalib.2006.08.014

Scharf D, Elliot N, Huey HA, Briller V and Joshi K (2007) Direct assessment of information literacy using writing portfolios. *The journal of academic librarianship* 33(4): 462-477.

Small R and Snyder J (2010) Research instruments for measuring the impact of school libraries on student achievement and motivation. *School Libraries Worldwide* 16(1): 61-72.

Schmeck A, Opfermann M, Van Gog T, Paas F and Leutner D (2015) Measuring cognitive load with subjective rating scales during problem solving: differences between immediate and delayed ratings. *Instructional Science* 43(1): 93-114. DOI: 10.1007/s11251-014-9328-3

Schwonke R, Berthold K and Renkl A (2009) How multiple external representations are used and how they can be made more useful. *Applied Cognitive Psychology* 23: 1227-1243. DOI: 10.1002/acp.1526

Sonley V, Turner D, Myer S and Cotton Y (2007) Information Literacy Assessment by Portfolio: A Case Study. *Reference Services Review* 35(1): 41–70. DOI: 10.1108/00907320710729355

Sousa VD and Rojjanasrirat W (2011) Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice* 17(2): 268-274. DOI: 10.1111/j.1365-2753.2010. 01434.x

Spante M, Hashemi SS, Lundin M and Algers A (2018) Digital competence and digital literacy in higher education research: Systematic review of concept use. *Cogent Education* 5(1): 1-21.

Sparks JR, Katz IR.and Beile PM (2016) Assessing digital information literacy in higher education: A review of existing frameworks and assessments with recommendations for next-generation assessment. *ETS Research Report* Series 2: 1-33.

Staley SM, Branch NA and Hewitt TL (2010) Standardised library instruction assessment: an institution-specific approach. *Information Research (Faculty Publications)* 15(3): 1–28.

Steiger JH (1990) Structural model evaluation modification: An interval estimation approach. *Multivariate Behavioral Research* 25: 173-180. DOI: 10.1207/s15327906mbr2502_4

Taylor A (2012) A study of the information search behaviour of the millennial generation. *Information research: an international electronic journal* 17(1): 1-20.

Thornton S (2006) Information literacy and the teaching of Politics. *Learning & Teaching in the Social Sciences* 3(1): 29–45.

Timmers C and Glas C (2010) Developing scales for information-seeking behaviour. *Journal of Documentation* 66(1): 46-69

Varlejs J and Stec E (2014) Factors Affecting Students' Information Literacy as They Transition from High School to College. *School Library Research* 17: 1-23.

Van Deursen, AJ and Van Diepen S (2013). Information and strategic Internet skills of secondary students: A performance test. *Computers & Education* 63: 218-226. DOI: /10.1016/j.compedu.2012.12.007

Van Deursen AJ and Van Dijk (2008) Measuring digital skills. Performance tests of operational, formal, information and strategic Internet skills among the Dutch population. *58th Conference of the International Communication Association*. Montreal, Canada.

Van Deursen AJ, Görzig A, Van Delzen M, Perik HT and Stegeman AG (2014) Primary school children's internet skills: A report on performance tests of operational, formal, information, and strategic internet skills. *International Journal of Communication* 8(23): 1343-1365. Available at: http://ijoc.org/index.php/ijoc/article/view/2407/1135V

Van Gog T, Paas F, van Merriënboer JJG and Witte P (2005) Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied* 11: 237-244. DOI: 10.1037/1076-898X.11.4.237

Van Laar E, Van Deursen AJ, Van Dijk JAGM, De Haan J (2017) The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior* 72: 577-588. DOI: 10.1016/j.chb.2017.03.010

van Strien JL, Kammerer Y, Brand-Gruwel S and Boshuizen HP (2016) How attitude strength biases information processing and evaluation on the web. *Computers in Human Behavior* 60: 245-252. DOI: 10.1016/j.chb.2016.02.057

Van de Vijver FJ and Leung K (1997) *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage Publications.

Velázquez SC (2010) Evaluación de la alfabetización informacional en el sector de la salud. *In Anales de documentación* (Vol. 13, pp. 41-51). Facultad de Comunicación y Documentación y Servicio de Publicaciones de la Universidad de Murcia.

West SG, Finch JF and Curran PJ (1995) Structural equation models with non-normal variables. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications.* Thousand Oaks, CA, USA: Sage, pp. 56-75.

Walhout J, Oomen P, Jarodzka H and Brand-Gruwel S (2017) Effects of task complexity on online search behavior of adolescents. *Journal of the Association for Information Science and Technology* 68(6): 1449-1461

Whitelock-Wainwright A, Laan N, Wen D and Gašević D (2020). Exploring student information problem solving behaviour using fine-grained concept map and search tool data. *Computers & Education* 145: 103-731.