



The Development and Face Validity Testing of the Open Test of Information Literacy with Context-Specific Add-ons: OTIL

Helena Hollis

University College London

helena.hollis.14@ucl.ac.uk, orcid.org/0000-0003-1950-0657

Dr. Marina Rachitskiy

Regent's University London

marina.rachitskiy@regents.ac.uk, orcid.org/0000-0002-8736-029X

Dr. Leslie van der Leer

Regent's University London

leslie.vanderleer@regents.ac.uk, orcid.org/0000-0001-7069-3760

Abstract

The OTIL is a new information literacy (IL) test, designed to be flexible and adjustable for use across a variety of contexts where IL is important. This paper discusses the development and face validation of OTIL and a Higher Education (HE) add-on that can be used to make the test HE specific. OTIL was developed in line with the new CILIP (2018) definition. It is comprised of a core set of general IL questions which can be combined with context-specific add-on question sets. Categories of IL form a template that can be used for additional question creation. Additionally, we have developed add-on questions for HE which do not contain in-built specificities to domain, region or institution in order to permit for IL research across different universities. The OTIL is designed for longitudinal research, with three

different test versions available. The categories, general questions, and HE questions have been face validity tested with a panel of experts. The expert panel found the OTIL to be relevant and to accurately represent what is meant by IL. Feedback from the panel was used to refine and improve the test questions. The OTIL aims to be adaptable to different information contexts, in recognition of the diverse ways in which IL can be expressed. We encourage researchers in IL to use the OTIL and create add-on question sets for their different information settings.

Keywords: information literacy; assessment; test development

1. Introduction

Information literacy (IL) has been defined in various ways, but is widely agreed to encompass one's ability to effectively find information, evaluate that information, store and manage it efficiently, and then use it in an ethical and meaningful manner, and finally to creatively make new information and share it (Owusu-Ansah, 2005). An individual's IL has been shown to impact their academic achievement (e.g. Bowles-Terry, 2012; Detlor, Julien, Willson, Serenko, & Lavallee, 2011), and therefore the importance of IL within academic contexts is well established. Beyond education, IL has also been argued to be of critical importance and to be a vital factor in society. Cope (2018) argues that IL is a means of questioning and reasoning within a social context, not merely a set of skills and precepts to be learned in a classroom. Indeed, the wider social importance of IL is emphasised in claims that it is a human right (Sturges & Gastinger, 2010; UNESCO, 2003). However, the testing of IL remains constrained to academic contexts (Hollis, 2018). Given the overt role of IL within educational contexts where students are taught and assessed on their use of information, especially at university level, this is not surprising. However, if we are to treat IL as a wider factor of importance to society at large, it would be useful to develop a means of assessing IL levels in individuals outside of educational institutions as well as within them.

IL assessment is a broad and varied field, encompassing both qualitative and quantitative approaches. We have focused on a quantitative approach, seeking to measure IL with a test that can generate IL scores and be used to make comparisons between groups, or with other factors. Various such measures

of IL already exist, but a previous evaluation of validated, freely available IL tests found them to be only designed for educational contexts, and to contain other limiting specificities which render them unusable outside of the domain, institution or region they were designed for (Hollis, 2018). To the best of our knowledge, at the time of this research, there is no test of IL designed for a general population. The only validated IL test that may have been appropriate for testing a general population was the iSkills test that was designed to be an IL test, but was later adapted to test ICT skills in the workplace (Katz, 2007), and is no longer available. There is also no validated and freely available IL test for Higher Education (HE) that is broad enough in focus that it could be used in any HE institution across the English-speaking world, with students of any subject. For instance, several existing HE tests are subject specific, e.g.: for chemistry (Emmett & Emde, 2007); education (Beile O'Neil, 2005); or psychology (Rosman, Mayer, & Krampen, 2015). Although there are HE IL tests that are not subject-specific, they have other inbuilt specificities which limit their usage. For example, some tests contain questions that are specific to the library of the institution it was designed for, in terms of resources and/or policies, and therefore cannot be used in other locations (Emmett & Emde, 2007; Ondrusek, Dent, Bonadie-Joseph, & Williams, 2013). Even very widely applicable IL tests contain some specificities, such as questions on a referencing style that not all institutions may use (Boh Podgornik, Dolničar, Šorgo, & Bartol, 2016). This is not a criticism of these existing tests; they are suitable for the specific purposes for which they were developed, typically for testing outcomes of IL interventions. Measuring IL so as to make comparisons between groups or with other factors is a different purpose that requires a different test development approach. Currently, researchers seeking to test IL with a validated instrument can do so with students, but not with other populations. Furthermore, without access to substantial funds to pay for tests, they will only be able to use these existing measures if they happen to be affiliated with an institution that has the same resources and policies as those for which the tests were designed. This is the gap that our new IL test aims to address. In addition, in line with Lloyd's (2011) notion that IL in the workplace is highly contextual and not generic, context-specific additions could be created for use in e.g. prisons, medical institutions, or different businesses.

A general test of IL as used in everyday life could be given to any participants in any context to compare the IL levels of different groups. This may yield

interesting findings, which would allow for stronger advocacy for promoting IL. For example, UNESCO and IFLA have both declared that information literacy is vital in furthering democratic engagement in society (IFLA, 2012; UNESCO, 2003). If IL levels could be measured in the general public, those IL scores could be compared with indicators of democratic engagement, and support these declarations with data (or if not supported, a further interesting debate could be raised). This is one example, but a wealth of other research questions could be explored using a test of general IL (e.g. *Do people with higher IL scores earn higher wages?*).

In HE, a vast range of research questions pertaining to IL are being addressed. However, as existing IL measures are designed to test particular cohorts (i.e. students within a subject area, or studying at one particular institution, or in one region), the comparisons that can be made are limited. A validated and freely available IL test that could be given to any student, studying any subject, in any institution would open the field for more research, yielding more interesting comparisons. Various research questions could be addressed (e.g. *Do higher university library budgets correlate with higher IL in the student body?*).

This paper presents the development and face validity testing of a new IL Open Test of Information Literacy (OTIL). The test has a core general IL subscale, which could be used as a standalone test for any participants in any context. We have also developed an HE subscale to be used alongside this. Used in conjunction, the OTIL with the HE add-on would allow researchers in HE to evaluate both their students' general IL levels as well as their IL as it relates to their studies. Additionally, the OTIL has the flexibility to add on additional subscales for different contexts. These have not currently been developed, but we would encourage information professionals working in other contexts to create add-ons to the OTIL.

2. Development of the Test

In order to develop a new IL test that has a core subscale of general IL questions, to which additional subscales can be added for different contexts, an underpinning definition of IL that would be broad enough to encompass

any context was needed. The Chartered Institute of Library and Information Professionals (CILIP) definition is specifically conceived to span across contexts, not only HE, and to include a wider base of IL concerns and practitioners. Furthermore, it fits closely with the way IL is described by other bodies (ACRL, 2016; ALA, 2000; IFLA, 2012; SCONUL, 2015; UNESCO, 2003). Therefore this definition was adopted for the development of the OTIL and the HE add-on.

In order to create a set of questions that would cover the full range of what is meant by IL, categories for IL questions were first created. To do so, the comprehensive CILIP definition of IL (CILIP, 2018) was first analysed and the language used was broken down, in order to identify themes, and used to create category headings. Table 1 demonstrates this process. First, the document was read closely and key language that specifically described what IL overall is was highlighted, and these highlights became extracts from the document that could be used for category creation. These extracts are in the first column of the table. The language extracted from the document was then re-grouped by topic, to put together the extracts that talk about the same type of IL activity or skill, with repetition deleted, so that the main themes can be seen in column two. Finally these themes were turned into categories for the creation of questions. The titles for the categories are in the third column of the table. The final theme of "Print; digital content; media; multiple sources" was not used as a category of questions, but rather as a criterion for question creation across the other categories (i.e. questions should ask about both physical and digital formats and different media types throughout the test).

Through this process of extraction, analysis, and categorisation, six distinct categories were identified, as outlined in Table 1. These categories have been used to create both the general OTIL and the HE add-on. To produce a balanced test that covers all facets of IL equally, an equal number of questions were developed for each category. It is not the aim of the test to comprehensively cover all aspects of each category (e.g. to fully measure how a participant uses and creates information in all possible ways), but rather to generate a score that is indicative of participants' IL overall as comprised of all categories. Therefore maintaining an equal number of questions per category provides a well-rounded overall score. As the aim for the test is to make it as widely useful as possible, the test was kept relatively short, making it less taxing on

Table 1: The creation of question categories from the CILIP definition of information literacy 2018 document.

Original language	
"how to discover, access, interpret, analyse, manage, create, communicate, store and share information." (p. 3)	
"make the best use of information and to interpret it judiciously" (p. 3)	
"critical thinking and awareness, and an understanding of both the ethical and political issues" (p. 3)	
"print, but also digital content, data, images and the spoken word" (p. 3)	
"Understand the ethical and legal issues associated with the use of information, including privacy, data protection, freedom of information, open access/open data and intellectual property" (p. 3)	
"understanding the limitation of these online resources, how they may be subject to manipulation and the need to be discerning about their value" (p. 4)	
"a need for awareness of internet security measures" (p. 4)	
"Information Literacy helps individuals to behave ethically in their online activities" (p. 4)	
"mindful of the information they use and share about themselves and others" (p. 4)	
"to reach informed views; where appropriate, to challenge, credibly and in an informed way assumptions or orthodoxies (including one's own), and even authority; to recognise bias and misinformation" (p. 4)	
"an ability to display critical judgement about multiple information sources" (p. 4)	
"information literacy helps to reach views about the reliability and authority of information sources" (p. 4)	
"In Higher Education, information literacy contributes to academic competencies, research methodologies and an understanding of plagiarism" (p. 5)	
"information literacy can be seen as the critical capacity to read between the lines" (p. 5)	
"perceiving relationships between important ideas, asking novel questions, and pursuing innovative lines of thought" (p. 5)	
"adopting a questioning approach, not only to solve problems, but also to frame problems and situations in new and ground-breaking ways" (p. 5)	
"interpret work-related information, share it (within organisations and with external stakeholders, such as clients or customers) and transform it into knowledge" (p. 5)	
"Information Literacy means working ethically, understanding the implications of data protection, intellectual property rights, such as copyright" (p. 5)	
Key themes	Categories for questions
Discover; access; research	Ability to discover and access information
Interpret; analyse; critical thinking and awareness; understand limitations; be discerning; recognise bias; questioning approach	Critical thinking ability
Manage; store; security	Ability to manage and store information effectively
Create; make the best use of; novel; innovative	Ability to use and create information

Table 1: (continued)

Key themes	Categories for questions
Share; communicate	Ability to share and communicate information
Understanding ethical and political issues; privacy; data protection; intellectual property; plagiarism	Understanding of ethical issues surrounding information
Print; digital content; media; multiple sources	

participants. Furthermore, we foresee the test being used in studies that make comparisons between IL and other factors (e.g. does IL correlate with job satisfaction?). Participants in such studies will be asked to complete not only the OTIL but also other sets of questions for the other factors being tested. Such studies, with various questionnaires, can become very lengthy, and short measures are thus desirable for this type of comparative research. With two questions per category, the general subscale has 12 questions in total. However, as there is no clear way in which 'the ability to manage and store information effectively' differs in HE compared to everyday life, this category was omitted in the HE add-on subscale, which therefore has 10 questions in total. In order to facilitate longitudinal testing, without asking participants identical questions at different times, three different versions of each subscale have been created. In total, there are therefore 36 general, and 30 HE questions.

Though many questions were written anew, some questions from existing IL tests have been adapted and incorporated (Beile O'Neil, 2005; Boh Podgornik et al., 2016; Brown & Brown, 2018; Catalano, 2015; Ferguson, Neely, & Sullivan, 2006; Geary, 2018; Mery, Newby, & Peng, 2011).

New questions were written, and existing questions edited, to be in line with best practice guidelines for multiple choice design in educational testing (Haladyna, Downing, & Rodriguez, 2002). Specifically, it was ensured that the question stems have been written to give precise instructions for participants; where a negative construction is used, the word NOT appears in upper case for clarity; there are no 'trick' questions included; all incorrect answers are plausible; all answer choices are similar in length and structure. Three answer choices to each question are given (Rodriguez, 2005), and the option of "I do not know" has been included as an additional answer choice in each question in order to reduce guessing (Zhang, 2013).

Across the 36 general questions and the 30 HE questions, difficulty varies. The questions have been divided into the three complete test versions balancing difficulty levels and ensuring each version is equally challenging. Questions deemed to be more challenging have been included in both subscales and in all three versions to avoid a ceiling effect.

General questions were designed with the aim of being comprehensible and reasonably answerable by any participant from most backgrounds or sectors of society. This is not to say that the test is appropriate for any possible participant; a good level of English language capability, reading ability, and broader verbal reasoning and cognitive capacity would be needed to meaningfully take the test. As a general guide, it is intended that most patrons of a public library would be able to complete it. Example scenarios were chosen to be such that even if a participant has no direct experience of the scenario, they ought to be able to imagine it. For example, question GU1 uses gardening as the theme for answers, but no actual knowledge of gardening is required, it is merely an illustrative topic. Topics were varied, so that no area of interest is privileged in the test.

HE IL questions were designed to be answerable by any student from foundation/access degree through to postgraduate level, in any discipline area. An average postgraduate should find the questions easier, and subsequently score higher, than an average undergraduate, but the test can be applied across levels. Questions may use examples from a certain subject area, but these are merely for illustrative purposes and no subject-specific knowledge is required to answer them. Similarly, some questions may use examples from specific systems that are not used across all programmes and institutions, such as HD1 which includes an example of Harvard style referencing, but the question can be answered without any knowledge of this specific system as the participant merely needs to read the example reference and find the key information within it.

HE questions were also reviewed against the Association of College and Research Libraries (ACRL) framework (ACRL, 2016). The frames were used as a checklist, and each HE question was marked off against these, ensuring that each frame is represented in each test version. Not every frame is equally represented in the test, nor has every one of the knowledge practices and dispositions been captured. However, overall the frames are covered by the questions and therefore the test is in line with the essence of the framework.

The ACRL framework was used as a checklist for the HE component of the test only. Due to the framework's academic context specificity, it was not suitable for checking questions in the general OTIL or as the fundamental underpinning basis for overall test design.

Full details of the 3 tests have been deposited in LIBER Quarterly's Dataset at Harvard Dataverse, where they can freely be consulted at <https://doi.org/10.7910/DVN/ZWOGMR>.

3. Face and Content Validity Testing

The full sets of general and HE questions were sent to library and information professionals for face and content validity assessment (i.e. the extent to which the scale appears to measure what it is designed to measure and the extent to which the scale appears to represent the whole construct, respectively; Coolican, 2014). As a key aim of the new test is wide applicability, respondents from a wide range of sectors were approached to be part of the expert panel. Their feedback was used to amend, exclude and replace items. Wider issues were raised by the panel which are also discussed.

3.1. Participants

Library and information professionals were recruited to form the expert panel for face validity testing. The individuals were selected based on their expertise in working with or providing training in IL or as Library and Information Science (LIS) researchers. In total, nine respondents provided feedback on the test questions. Their current sectors of work were HE (n=4), school libraries (n=1), public libraries (n=3), and consultancy (n=1). They had previous experience of work in HE (n=2), school libraries (n=2), public libraries (n=3), Further Education (FE) (n=2), career resource centres (n=1), special libraries (n=1) and product libraries (n=1). From the nine participants, five were also engaged in LIS research. Respondents were located in the UK, USA, and Singapore.

In addition to the nine respondents, the CILIP Information Literacy Group (ILG) were contacted, and the full set of questions was sent out to all

members of the group. Members of the ILG responded anonymously via a spokesperson for the group. They provided overall general feedback about the test.

3.2. Materials and Procedures

The complete set of general and HE questions were emailed to respondents. A brief description of the test and the rationale for its creation was included. Two sets of questions were given to respondents, firstly asking about their current and previous experience of library and information sectors, and secondly asking about the test. To assess the face validity of the new test both the overall representation of IL, and the relevance of individual questions, were addressed (Hardesty & Bearden, 2004). Questions asked whether the categories used are appropriate, and whether any questions were irrelevant or not part of IL. Respondents were asked about how well the construct of IL was represented, i.e. whether the questions adequately and fully cover what is meant by IL. Furthermore, respondents were asked to critique the test items by identifying any questions that were unclear or of an inappropriate level. Respondents were asked for their opinion on the inclusion of an “I do not know” answer option. Finally, respondents were asked if they would like to see any questions added, or if they had any further comments. The qualitative responses were received in writing and analysed using content analysis.

3.3. Results and Actions Taken

3.3.1. Face Validity and Construct Representation

No respondents felt that any aspect of IL was missing from the questions, and all were satisfied that the full concept was covered by the questions. One respondent suggested that “use and create” and “share and communicate” are too similar and should be combined into one category. However, as all other respondents felt the categories were clear and useful, it was decided to keep these distinct. The possible merging of these categories may be revisited in the future after factor analysis is conducted on the test.

One respondent felt that the equal number of questions for each category was inappropriate, as not all facets of information literacy are equally important within it as a concept. Specifically, the respondent felt critical thinking should be given a greater weighting:

“...some sections could contain fewer questions, whereas the critical thinking section could be expanded”

This view prompted a review of the rationale for keeping the categories of questions equally weighted, as the balance of these factors is not necessarily equally balanced within the construct of IL. Critical thinking does indeed feature most heavily in the definition used to develop our categories (CILIP, 2018). However, critical thinking is necessary to answer most questions in the test, as logical reasoning must be deployed to work out the correct answer. It was therefore decided that additional questions focused solely on critical thinking were not needed as it should be well captured by the questions as a whole, and therefore its predominance within the concept of IL is reflected in the test. Beyond critical thinking, some categories could have been given different weightings. However, which categories are more or less important within the construct of IL is a subject of much debate within the field, and different researchers will subscribe to different views on which aspects should be foregrounded (Owusu-Ansah, 2005). To make the test equally useful to all IL researchers, we therefore decided to preserve the even number of questions per category.

3.3.2. Specific Question Critique

In evaluating individual items, both criticism and praise of each item was tracked. Where an item was questioned, and the decision was made to remove the item, we replaced the removed item with an adapted version of a different, preferred item from the same category.

Initially, the OTIL included questions on Intellectual Property (IP). Following feedback from several respondents that the IP questions were unclear, and potentially not applicable across different countries where IP laws and practices may differ, these were removed. They were instead replaced with questions about the ownership of information more broadly.

In the wider feedback received from the ILG, one reviewer felt that questions on specific search strategies (using Boolean operators) are not part of IL. This could be seen as a very narrow technical type of knowledge that is not part of, or only peripherally part of, the concept of IL. However, we also received feedback from another respondent who felt the test ought to include *more* Boolean search questions. Ultimately, we concluded that, even if a participant does not know the operators, they should be able to deduce the right answer. This would then demonstrate critical thinking. Answering questions on search operators also demonstrates an understanding of the underlying mechanisms of search engines, which we consider a highly important part of modern-day IL. We decided that questions on search operators do tap into facets of IL and, therefore, have kept them in the test. However, these have been kept to a minimum so as not to dominate the 'discover and access' category.

3.3.3. Additional Questions

Two respondents commented that additional questions should be included about social media, especially in the general subscale. This was taken into consideration, and various question options to cover social media were explored. However, questions about a specific social media platform would exclude participants who do not use that particular platform, and these questions could also quickly become dated (e.g. a question on Myspace would once have seemed pertinent, but would now render the test outmoded). Creating social media questions that are not specific to any given platform led to overly vague and unclear items, as the rules and privacy options of various platforms differ greatly. Ultimately, it was decided that existing questions on the creation, sharing and ethics of information sufficiently cover the issues that apply in social media use, even though they do not use the overt vocabulary of social media. In this way, the test covers the skills and knowledge needed in navigating social media, without making direct reference to it.

Additionally, these two respondents suggested that questions on "fake news" be included, due to its importance and timeliness. As with social media, it was decided not to explicitly include this terminology as it may lose its popularity and become an outdated phrase. However, the assessment of sources, and discernment between fact and opinion are included in the test, and these

are the core skills needed to identify “fake news”. Therefore, the test does address the necessary IL abilities pertaining to news evaluation.

3.3.4. Including an “I do not know” Answer Option

Out of nine respondents, five were in favour of the inclusion of “I do not know” as an answer option; two respondents expressed no preference; and two respondents were against the inclusion of an “I do not know” answer. One respondent who was against it stated:

“I suppose when one search[es] for info or perform[s] a task, they would ultimately have to make a choice, no matter how unsure one is.”

This is an interesting point, as participants guessing answers could result in inaccurate scores through chance, while having participants select “I do not know” could result in scores that are less true to life, as in a real information activity the participants would most likely have taken a guess. Removing the “I do not know” option may make the test closer to real-life information activities, but in fact this test is not measuring behaviour, it is measuring knowledge and understanding of IL. As such, although participants may make guesses in their information behaviour in real life, we aimed to minimise guessing in the measurement of their knowledge and understanding. Ultimately it was decided to retain the “I do not know” option in the test, as it provides participants with the option to honestly report their uncertainty. It is therefore an interesting potential data point in its own right; The levels of “I do not know” answers could be analysed, and trends in gaps in knowledge could be identified.

3.4. Broader Issues Raised

3.4.1. Single Correct Answers

One respondent from the ILG questioned the prescriptive nature of the test, as there is always one correct answer, which they felt was not appropriate in IL testing:

“It isn’t about ‘right answers’ and ‘right ways’ of doing things (...) but negotiated practices that are fluid, very open-ended, and take place in a socially constructed environment where absolutes and universals are very questionable (...) the ‘single right answer’ approach taken here is deeply at odds with a critical and situated view of IL. It seems to be based on a highly normative approach that does imply a universal ‘right answer’ or ‘right behaviour’ that is in the gift of librarians.”

As the purpose of this test is to generate an IL score that can be analysed (e.g. to compare groups), the questions necessarily need to have a correct answer. However, we acknowledge that IL does take place in a variety of environments and different practices will be more or less effective dependent upon the context and aims of the individual. Therefore, questions have been devised to give an indication of the context the participant should imagine, so that they can select an answer that is appropriate for that context, rather than one universal answer. Further, following this feedback, questions were carefully reviewed to remove highly prescriptive or absolute language, and rephrase it with language that is more contextualised (e.g. GM4 asks which method would make it easier to find the information later, not which method is best). Though this is ultimately a quantitative test that does score answers as right or wrong, it aims to give sufficient information in the question stems to enable participants to select the most appropriate answer for the described need.

3.4.2. Use of Multiple-Choice Question Format

Further feedback from an ILG member questioned the use of a multiple-choice question (MCQ) format for assessing IL:

“...the MCQ format is highly problematic in any learning or assessment context. It notoriously does not test a thoughtful critical response, actual behaviours, or the potential to apply newly-learned practices in other situations (...) what is [sic] actually tests is simple memorisation of what the teacher has said...”

It is true that the MCQ format is not a measure of actual behaviour and there are many more applied methods of testing IL. However, the purpose of this IL

test is not to measure IL *behaviour*, but rather IL *knowledge and understanding*. To assess this, the MCQ format was considered the most suitable as it allows for quick, easy and impartial assessment resulting in a clear score. Easy distribution, scoring and analysis makes the test useful to a broader range of researchers who may not have the time or resources for more complex methods. Part of the future research could separately investigate whether participants with high IL scores do, in fact, behave in an information-literate way.

Furthermore, educational testing literature suggests MCQ testing is effective over time (Madaus & O'Dwyer, 1999), and scores on MCQs correlate strongly with descriptive tests (Brown, 2001). Many of the disadvantages attributed to MCQ tests (e.g. the claim it only tests memory, not analytical thinking) have been strongly disputed (e.g. Roberts, 2006; Woodford & Bancroft, 2004).

4. Discussion

Although numerous IL measures are available, they tend to be narrow and specific to the area for which they were developed. We aimed to develop a more broad and general IL measure which could be combined with context-specific add-ons, such as HE-specific IL. To accomplish this, we developed a scale in line with the CILIP definition of IL. Two sets of questions were developed. The general scale included questions that are more general in nature and reflects IL knowledge of an average adult. The HE subscale included questions that are specific to HE and evaluate IL knowledge that is commonly learned through HE. The face validity of the scales was evaluated by nine experts in the area of IL and CLIP ILG members.

Based on the feedback of our expert panel, we are satisfied that the categories used to underpin the OTIL are valid in their face value representation of IL, and that the content matches IL as a construct. These categories can be used in the creation of any additional subscales for other contexts. Furthermore, having replaced and refined individual items following the panel's feedback, the questions designed for each category for both the general and the HE subscale also satisfy the face validity of IL. These questions will now need to be tested and assessed for statistical reliability and validity; this will be the next stage of the development of the OTIL and our statistical findings will also be disseminated.

Some comments raised an interesting ontological and epistemological debate. Having single correct answers, and using the multiple-choice format, can be seen to fit best with a positivist, instruction-focused approach to IL. Cope (2018) and Lloyd (2011) posit that IL is entirely contextual and socially constructed. This view may reject the notion of IL as a coherent, singular construct. In designing this test, we envisage it as a research tool enriching the scope of IL investigation, not as a standardised examination. We recognise the way IL is constructed in different situations, but also assume there is an underlying construct of IL that is manifested across contexts. For example, if someone is very good at discerning authors' opinions in the news, or at summarising information concisely for their job, or utilising multiple sources to find a good deal, the person is, in fact, displaying their underlying IL in each of these highly-contextualised activities. Furthermore, we would expect IL to manifest itself across different contexts, including while answering the questions of the OTIL, and thus a singular and coherent construct of IL is measurable.

However, this may not be the case. It may be that each information context demands a separate set of knowledge and abilities, which do not translate over into other settings. If so, we would find a low correlation between scores on the general and HE subscale, and we would expect the OTIL to show low internal validity. This is not our current prediction, but it would certainly be an interesting finding, and the creation of context-specific add-ons to the OTIL will allow for further research into this question. Because all existing validated IL tests are context specific (Hollis, 2018), it has not been possible to gather this kind of quantitative evidence and test this hypothesis previously. The opportunity to create more add-ons also acknowledges that the test can be as fluid as IL itself.

4.1. Limitations and Future Directions

In testing the face validity of the OTIL, we recruited information professionals from a range of fields. Our panel had a rich breadth of experience in information work and research, and responded that our question categories and questions were valid and representative of IL. However, only limited conclusions on the validity of the test can be drawn from their assessment, as not all library and information contexts were represented

within our expert panel. We hope future evaluation of the OTIL and the HE add-on will provide more support for the tests' validity from information professionals in other fields and contexts (e.g. health, prison or law librarians) and invite experts to feedback to us on the face validity of OTIL and the HE add-on.

Further validity testing (i.e. discriminant, convergent and predictive validity) must also be conducted for the OTIL to be considered a validated instrument that adequately tests the construct of IL (Coolican, 2014; Hardesty & Bearden, 2004; Messick, 1979). This will be the next phase in our project, and we aim to collect and analyse data from varied participants, the results of which we will disseminate in future publication. We would encourage other researchers to independently validate our test.

The aim of the OTIL development was to create a flexible IL scale where more specific subscales could be added to make it more specialised to a particular context. According to the CILIP extensive definition, IL is observed across a variety of contexts: Everyday life; Citizenship; Education; Workplace; and Health (CILIP, 2018). The proposed OTIL targets 'everyday life', and the HE add-on partially targets the 'education' context. Further subscales could be developed for the other contexts, to be used alongside the general scale. We hope to inspire library and information professionals working in these various sectors to follow our process and develop subscales for their specialised contexts.

In addition, we hope to inspire the development of further add-ons in order to adjust the difficulty level of OTIL. Currently, the OTIL may be very challenging, or easy, for some participant groups. This can facilitate testing for comparisons between groups. However, further add-ons could be combined with the general scale to make it more challenging, or simpler, for particular populations.

Finally, though the general and HE add-on aim to be as widely applicable as possible, there are limitations. This is an English language test, and though non-native speakers with a good level of English should be able to complete it, it does exclude non-English speakers. The test uses British English, which may be off-putting to those accustomed to different spelling styles, though it should not prevent comprehension. The OTIL is designed for use with adults. It might be suitable for young adults, perhaps with a

secondary school add-on, but it would not be suitable for testing children. Though the level of OTIL is pitched at a general public library audience, this does potentially make it inappropriate for some settings (e.g. for participants with a lower reading age, or participants who do not use digital media).

5. Conclusion

The OTIL is a new IL test that can be used outside of educational contexts, enriching the scope of future IL research. It has been tested for face validity, and further validity testing will be undertaken. The OTIL offers a simple set of categories that form a flexible template for future add-ons to be created, adapting it for any library and information context. We encourage anyone researching IL to use the OTIL and create add-ons to suit their needs. The test can be consulted at LIBER Quarterly's Dataset <https://doi.org/10.7910/DVN/ZWOGMR>.

References

- ACRL. (2016). *Framework for information literacy for higher education* [Text]. Retrieved September 21, 2018, from <http://www.ala.org/acrl/standards/ilframework>.
- ALA. (2000). *Information literacy competency standards for higher education*. Retrieved September 29, 2018, from <http://www.ala.org/Template.cfm?Section=Home&template=/ContentManagement/ContentDisplay.cfm&ContentID=33553>.
- Beile O'Neil, P. (2005). *Development and validation of the Beile test of information literacy for education (b-tiled)* (Doctoral thesis). University of Central Florida. Retrieved February 5, 2019, from <https://stars.library.ucf.edu/etd/530/>.
- Boh Podgornik, B., Dolničar, D., Šorgo, A., & 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. <https://doi.org/10.1002/asi.23586>.
- Bowles-Terry, M. (2012). Library instruction and academic success: A mixed-methods assessment of a library instruction program. *Evidence Based Library and Information Practice*, 7(1), 82–95. <https://doi.org/10.18438/B8PS4D>.

- Brown, R.W. (2001). Multi-choice versus descriptive examinations. In *31st Annual Frontiers in Education Conference: Impact on Engineering and Science Education*, Volume III (pp. T3A-13). Reno, NV, USA: IEEE. <https://doi.org/10.1109/FIE.2001.963903>.
- Brown, E., & Brown, H. (2018). *What you don't know might hurt you: Using pre-testing to assess information literacy knowledge*. Presented at the Quantitative and Qualitative Methods in Libraries (QQML), Chania, Greece.
- Catalano, A. (2015). The effect of a situated learning environment in a distance education information literacy course. *The Journal of Academic Librarianship*, 41(5), 653-659. <https://doi.org/10.1016/j.acalib.2015.06.008>
- CILIP. (2018). *CILIP definition of information literacy 2018*. Information Literacy Group. Retrieved December 1, 2018, from <https://infolit.org.uk/ILdefinitionCILIP2018.pdf>.
- Coolican, H. (2014). *Research methods and statistics in psychology* (6th ed.). Hove, England: Psychology Press.
- Cope, J. (2018). Information literacy and social power. In M.T. Accardi, E. Drabinski, & A. Kumbier (Eds.), *Critical library instruction: Theories and methods* (pp. 13-27). Sacramento, CA: Library Juice Press.
- Detlor, B., Julien, H., Willson, R., Serenko, A., & Lavallee, M. (2011). Learning outcomes of information literacy instruction at business schools. *Journal of the American Society for Information Science and Technology*, 62(3), 572-585. <https://doi.org/10.1002/asi.21474>.
- Emmett, A., & Emde, J. (2007). Assessing information literacy skills using the ACRL standards as a guide. *Reference Services Review*, 35(2), 210-229. <https://doi.org/10.1108/00907320710749146>.
- Ferguson, J.E., Neely, T.Y., & Sullivan, K. (2006). A baseline information literacy assessment of biology students. *Reference & User Services Quarterly*, 46(2), 61-71. <https://doi.org/10.5860/rusq.46n2.61>.
- Geary, J. (2018). *Measuring learning outcomes of new library initiatives*. Presented at the Quantitative and Qualitative Methods in Libraries (QQML), Chania, Greece.
- Haladyna, T.M., Downing, S.M., & Rodriguez, M.C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-333. https://doi.org/10.1207/S15324818AME1503_5.
- Hardesty, D.M., & Bearden, W.O. (2004). The use of expert judges in scale development: Implications for improving face validity of measures of unobservable constructs. *Journal of Business Research*, 57(2), 98-107. [https://doi.org/10.1016/S0148-2963\(01\)00295-8](https://doi.org/10.1016/S0148-2963(01)00295-8).
- Hollis, H. (2018). Information literacy as a measurable construct: A need for more freely available, validated, and wide ranging instruments. *Journal of Information Literacy*, 12(2), 76-88. <https://doi.org/10.11645/12.2.2409>.
- IFLA. (2012). *The Moscow declaration on media and information literacy*. Retrieved December 1, 2018, from <https://www.ifla.org/files/assets/information-literacy/publications/moscow-declaration-on-mil-en.pdf>.

- Katz, I.R. (2007). Testing information literacy in digital environments: ETS's iSkills assessment. *Information Technology and Libraries*, 26(3), 3–12. <https://doi.org/10.6017/ital.v26i3.3271>.
- Lloyd, A. (2011). Trapped between a rock and a hard place: What counts as information literacy in the workplace and how is it conceptualized? *Library Trends*, 60(2), 277–296. <https://doi.org/10.1353/lib.2011.0046>.
- Madaus, G.F., & O'Dwyer, L.M. (1999). A short history of performance assessment: Lessons learned. *Phi Delta Kappan*, 80(9), 688–695.
- Mery, Y., Newby, J., & Peng, K. (2011). Assessing the reliability and validity of locally developed information literacy test items. *Reference Services Review*, 39(1), 98–122. <https://doi.org/10.1108/00907321111108141>.
- Messick, S. (1979). Test validity and the ethics of assessment. *ETS Research Report Series*, 1979, i–43. <https://doi.org/10.1002/j.2333-8504.1979.tb01178.x>.
- Ondrusek, A., Dent, V.F., Bonadie-Joseph, I., & Williams, C. (2013). A longitudinal study of the development and evaluation of an information literacy test. *Reference Services Review*, 33(4), 388–417. <https://doi.org/10.1108/00907320510631544>.
- Owusu-Ansah, E.K. (2005). Debating definitions of information literacy: enough is enough! *Library Review*, 54(6), 366–374. <https://doi.org/10.1108/00242530510605494>.
- Roberts, T.S. (2006). The use of multiple choice tests for formative and summative assessment. In *Proceedings of the 8th Australasian Conference on Computing Education – Volume 52* (pp. 175–180). Sydney, NSW: Australian Computer Society, Inc.
- Rodriguez, M.C. (2005). Three options are optimal for multiple-choice Items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2), 3–13. <https://doi.org/10.1111/j.1745-3992.2005.00006.x>.
- Rosman, T., Mayer, A.-K., & Krampen, G. (2015). Measuring psychology students' information-seeking skills in a situational judgment test format. *European Journal of Psychological Assessment*, 32(3), 220–229. <https://doi.org/10.1027/1015-5759/a000239>.
- SCONUL. (2015). *Perceptions of the seven pillars of information literacy: A brief review*. Retrieved February 5, 2019, from <https://www.sconul.ac.uk/sites/default/files/documents/Seven%20Pillars%20Review%202015.pdf>.
- Sturges, P., & Gastinger, A. (2010). Information literacy as a human right. *Libri*, 60(3), 195–202. <https://doi.org/10.1515/libr.2010.017>.
- UNESCO. (2003). *The Prague declaration: Towards an information literate society*. Retrieved December 1, 2018, from <http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/PragueDeclaration.pdf>.
- Woodford, K., & Bancroft, P. (2004). Using multiple choice questions effectively in information technology education. In R. Atkinson, C. McBeath, D. Jonas-Dwyer &

R. Phillips (Eds.), *Beyond the comfort zone: Proceedings of the 21st ASCILITE Conference* (pp. 948–955). Perth. Retrieved December 1, 2018, from <https://www.ascilite.org/conferences/perth04/procs/pdf/woodford.pdf>.

Zhang, X. (2013). The 'I Don't Know' option in the vocabulary size test. *TESOL Quarterly*, 47(4), 790–811. <https://doi.org/10.1002/tesq.98>.