DOI: https://doi.org/10.57125/ELIJ.2025.09.25.02

How to cite: Ibrahim, S. E., & Ismail, M. R. (2025). Item analysis of information literacy test for open, distance and digital education new learners. *E-Learning Innovations Journal, 3*(2). 31–54. https://doi.org/10.57125/ELIJ.2025.09.25.02

Item Analysis of Information Literacy Test for Open, Distance and Digital Education New Learners

Shahril Effendi Ibrahim*

Master of Library and Information Science, Assistant Manager Digital Library, Faculty of Education, Open University Malaysia, Malaysia, https://orcid.org/0009-0002-7901-8053

Md Rosli Ismail

PhD and Master of Philosophy, Faculty of Education, Open University Malaysia, Malaysia, https://orcid.org/0000-0003-4765-5074

Corresponding author: *shahril_effendi@oum.edu.my*.

Received: April 19, 2025 | **Accepted:** August 17, 2025 | **Published:** September 25, 2025

Abstract: Assessing information literacy is key to understanding students' ability to locate, evaluate, and use information effectively. This study aimed to develop and validate multiple-choice questions (MCQs) to assess the information literacy skills of new undergraduate learners in open, distance, and digital education (ODDE), using the SCONUL Seven Pillars of Information Literacy Model as the framework for the construct. A total of 60 MCQs were tested with 388 new learners. Item analysis evaluated the difficulty index (DIF I), discrimination index (DI), and distractor effectiveness (DE). Validity was established through expert review and construct alignment, while reliability was measured using Cronbach's alpha and KR-20. Analysis via Excel and JASP showed that 41 items (68.3%) had acceptable difficulty levels, with a mean DIF I of 55.59 ± 10.63. Two items (3.2%) were too complicated, while 17 items (28.3%) were too easy. Based on DI, 54 items (90%) were rated as acceptable to excellent, and six (10%) were rated as poor. A weak positive correlation was found between DIF I and DI ($r = 0.222$, $p > 0.05$). Educators in various ODDE settings can adapt the refined-

items instrument to assess learners' information literacy levels based on local needs contexts. Its use supports more precise diagnostic practices, enabling targeted instructional interventions and resource planning in digital learning environments. The study not only introduces a psychometrically validated IL test based on the SCONUL model but also provides actionable evidence to guide higher education institutions and policymakers in embedding information literacy into national digital literacy strategies, curriculum design, and quality assurance frameworks for ODDE.

Keywords: information literacy, information literacy assessment, SCONUL Seven Pillars of Information Literacy, reliability management, test and testing.

Introduction

In an era where information is readily available, accessible, and abundant, every learner must be equipped with skills for accessing, retrieving, evaluating, utilising, and presenting information effectively. These skills and competencies are defined as information literacy (IL). By acquiring those skills, learners will be able to participate actively in their learning process, engage in critical thinking, problem-solving, and collaboration with colleagues. In addition, the skills empower students to differentiate between scholarly, non-scholarly, and misleading materials (e.g., fabricated news, alternative facts) (Jones & Mastrorilli, 2022).

In recent years, information literacy assessment has undergone considerable development due to shifts in educational settings, the influence of digital technologies, and the increasingly complex nature of the information environment. Globally, there is a shift away from content-oriented approaches towards frameworks based on competencies and context, with increased attention to critical thinking, responsible information use, and lifelong learning. This change can be observed in the use of several international models, including the ACRL Framework for Information Literacy for Higher Education (Association of College and Research Libraries [ACRL], 2015) and the SCONUL Seven Pillars of Information Literacy (The Society of College, National and University Libraries [SCONUL], 2011).

One prominent trend is the expansion in the types of assessment methods used. Around the world, educational institutions are combining both perception-based measures—such as information literacy self-efficacy scales—and performance-based instruments, including multiple-choice items and tasks simulating real-world scenarios. In the American context, there is widespread application of standardised tools, for example, Project SAILS (Standardised Assessment of Information Literacy Skills) and the Information Literacy Test (ILT), particularly in academic library environments.

To address this issue, the present study has developed and analysed a set of multiple-choice questions (MCQs) based on the SCONUL Seven Pillars model. The item analysis focused on the difficulty index, discrimination index, distractor effectiveness, and internal consistency. Through this

approach, the study aims to ensure the test items are valid and reliable for use in assessing information literacy among new learners in the ODDE setting.

According to the Association of College and Research Libraries (ACRL, 2015), IL comprises an integrated set of skills that include the reflective identification of information, understanding the production and valuation of information, and using information to generate new knowledge and engage ethically within learning communities. Like other competencies and skills, the level of IL skills and knowledge needs to be determined and assessed. Various methods and instruments have been introduced and implemented to assess students' competencies in IL skills. The assessment can be divided into two common types: a) Information literacy perception-based assessment and b) Information literacy test-based assessment (Walters et al., 2020).

The Open University Malaysia, a pioneer in open, distance, and digital education (ODDE) in Malaysia, is committed to providing its new learners with information literacy competencies and knowledge. However, there is a need to assess the level of IL competencies among new learners. MCQs, one method in IL test-based assessment, have been selected as an assessment format to evaluate the new undergraduate learners. According to Kumar et al. (2021), MCQs are frequently employed in assessments, particularly in universities, due to their efficiency, reliability, and ease of standardisation. A group of researchers, comprising two academics and two librarians, has embarked on an attempt to develop a new set of IL scenario-based test assessments in MCQ format to evaluate the IL skills and competencies of new learners. The test will be coordinated as an educational diagnostic assessment. Results and data from the assessment will be used to determine the strengths and weaknesses of the new learners. Interventions such as library instructions, information literacy workshops, learning objects, and special tutorials shall be organised accordingly, based on the results of the test, to provide the new learners with the necessary IL skills.

However, test-based assessments need to be appropriately developed and organised. This is to ensure that the results and data gained from the assessment fulfil the objectives of the assessment, namely to determine the strengths and weaknesses of the new learners' IL skills and ultimately provide new undergraduate learners with the necessary IL skills and competencies for their education in the ODDE environment. Hence, item analysis should be implemented to provide input on the validity and reliability of the items in the MCQs. It helps to identify items that can be revised and discarded, resulting in a quality question bank (Kumar et al., 2021).

Research Problem

Although various IL assessments have been introduced globally, many of them are not adapted to the local ODDE environment and lack proper psychometric validation. Most available instruments rely on self-perception, which may lead to an overestimation of ability, such as the Dunning-Kruger effect, and are not always suitable for the Malaysian educational and cultural context. This situation

presents a challenge for ODDE institutions to accurately measure IL skills and plan appropriate teaching and learning interventions based on trustworthy data.

The paper addresses the issue that new learners in Open, Distance and Digital Education (ODDE) environments often enter without a clear understanding of their own information literacy (IL) skill levels. While test-based assessments are more objective than self-assessment tools, many IL tests lack proper validation and reliability checks, and there is no standardised IL test tailored for ODDE contexts in Malaysia. Therefore, the problem is the lack of a reliable, valid, and locally developed IL test instrument to assess the IL competencies of new learners accurately.

Research Focus

The primary focus of this study is the development and psychometric evaluation of a multiple-choice question (MCQ)-based information literacy (IL) test tailored for new undergraduate learners in an Open, Distance, and Digital Education (ODDE) environment, specifically at Open University Malaysia (OUM). Recognising the lack of standardised, contextually appropriate IL assessment tools in Malaysian ODDE settings, this study seeks to establish an evidence-based, valid, and reliable instrument that reflects the cognitive and practical IL needs of adult learners.

The theoretical foundation of the study is grounded in the SCONUL Seven Pillars of Information Literacy model, which serves both as the conceptual framework for item development and the basis for construct validity assessment. Each test item is mapped to the core competencies outlined by the SCONUL model, namely a) Identify, b) Scope, c) Plan, d) Gather, e) Evaluate, f) Manage, and g) Present, ensuring that the assessment captures a comprehensive range of IL skills. The research is also informed by Bandura's Social Cognitive Theory, particularly his construct of self-efficacy, acknowledging the interplay between learners' knowledge, confidence, and performance in information-handling tasks. The psychometric evaluation includes detailed item analysis focusing on:

- Difficulty Index (DIF I), to determine item appropriateness across ability levels.
- Discrimination Index (DI), to assess each item's ability to differentiate between high- and low-performing students.
- Distractor Effectiveness (DE), to evaluate the functionality of incorrect options.
- Test reliability, using both Cronbach's alpha and Kuder-Richardson Formula 20 (KR-20).
- Construct validity, via expert judgement and alignment with the SCONUL framework.

Research Aims

The aims of the study are to:

- Develop a reliable and valid MCQ-based information literacy test for new undergraduate learners.

- Evaluate the quality of each item using psychometric analysis (difficulty, discrimination, and distractor effectiveness).
- Determine the overall reliability of the test instrument.
- Use item analysis outcomes to improve and finalise the IL test for future implementation.

Research Questions

The research questions of the study are:

1. How difficult are the items in the newly developed IL MCQ test for new ODDE learners?
2. How well do the items discriminate between high and low performers?
3. How effective are the distractors in the test items?
4. What is the reliability of the IL test using internal consistency measures?
5. Can the IL test developed based on the SCONUL model be used as a valid assessment instrument in the ODDE context?

This study aims to address several key research questions related to the development of a multiple-choice information literacy (IL) test for new learners in ODDE environments. The first question examines the difficulty level of the test items, a fundamental concept in Classical Test Theory (CTT). Analysing item difficulty helps determine whether the items are appropriate for the intended learners' ability levels. The second question examines how well the test items discriminate between high and low performers. In CTT, the discrimination index is a key indicator of item quality, reflecting how effectively an item distinguishes between learners with stronger and weaker IL skills. The third question focuses on the effectiveness of distractors used in the MCQ items. From a test development perspective, distractors are considered functional if they attract a reasonable proportion of respondents with low performance. This shows that incorrect options are plausible and contributes to the test's overall validity. The fourth question assesses the internal consistency of the test using reliability measures such as the Kuder-Richardson Formula 20 (KR-20) and Cronbach's alpha. These statistics help confirm whether the items are consistently measuring the same construct. Finally, the fifth research question examines whether the test, based on the SCONUL Seven Pillars of Information Literacy, can be used as a valid assessment tool in the ODDE context. The SCONUL model provides a theoretical foundation for defining IL competencies, and evaluating construct validity ensures that the test aligns with this model and the needs of adult distance learners.

Research Hypotheses

H1: The overall test will exhibit high internal consistency, with KR-20 and Cronbach's alpha values exceeding 0.80, indicating strong reliability.

H2: The majority of test items will fall within the acceptable difficulty range (30%–70%), indicating appropriate item challenge levels for the target population.

H3: At least 70% of the items will demonstrate good to excellent discrimination (DI ≥ 0.2), reflecting the test's ability to differentiate learners' performance.

H4: The majority of distractors will function effectively (selected by at least 5% of respondents), demonstrating plausible and contextually relevant incorrect choices.

Literature Review

The literature review is divided into three parts, namely: a) IL assessment, b) IL standards and models, and c) Analysis of test items.

a) Information Literacy Assessment

In the context of information literacy (IL), the primary aim of an assessment is to evaluate students' skills in IL. Examples of utilising IL assessments include evaluating the efficiency of IL instructional programs, aligning instruction with IL learning objectives, assessing the impact of modifications in instructional programs, and enhancing the assessment process itself. According to Yu (2023), locally created assessments can be tailored to the specific learning objectives, curriculum, and context of an institution, ensuring greater relevance and alignment with local needs and instructional goals. Pinto et al. (2024) agree that a locally developed test can demonstrate how information literacy programs affect student learning outcomes and provide valuable data for enhancing instructional programs. Numerous efforts have been undertaken to evaluate information literacy through an objective test evaluation method. One of the ways involves fixed-choice assessments, including multiple-choice (MCQs), matching, and true-false formats. They are widely used for their objectivity, efficiency, and scalability in evaluating knowledge and skills (Jones, 2020). Walsh (2009) asserts that, following his evaluation of nine types of assessment instruments, the multiple-choice question is the predominant approach for testing students' competencies in information literacy. According to Rosman et al. (2015), the MCQ type of test is "easy to administer while also maximising scoring objectivity" (p. 1).

In comparison to other assessment instruments, tests that are developed locally and consist of fixed-choice questions, such as MCQs, matching, and true-false tests, are. If questions in MCQs are well-constructed, MCQs can be "the most convenient, fair, and inexpensive way to judge an individual's level of performance or readiness" (Elmas et al., 2018, p. 14). Locally developed tests are also capable of quantifying student learning outcomes and facilitating comparisons between groups when utilised as pre- and post-assessments in a course (Oakleaf, 2008). MCQs developed locally also have a drawback. One of them is in terms of know-how. This type of test is unable to address higher-order abilities since it only captures declarative knowledge. When know-how is expected as a learning outcome, MCQs have limited capability to evaluate. However, Xu et al. (2016) argue that well-written tests are practical, versatile, and can measure and evaluate both higher-order and lower-order thinking abilities. Consequently, Babo and Suhonen (2018) recommend adopting a complementary approach for evaluating these skills and competencies. They also agree that test-

based assessment can prevent overestimation through deliberate over-reporting of abilities. This overestimation behaviour is also known as the Dunning-Kruger Effect and is frequently exhibited by students who were assessed based on perception-based assessments (Zhou & Jenkins, 2020).

b) Information Literacy Standards and Models

A variety of tests and questionnaires are employed as tools for evaluating information literacy knowledge and skills. The majority adhere to the recommendations and subjects outlined in many information literacy standards, models, and guidelines established by recognised bodies and agencies in IL (Al-Qallaf, 2019; Boh Podgornik et al., 2016). Each of these standards and guidelines has its own learning outcomes and objectives. One of the IL standards and guidelines most frequently referenced when establishing IL assessment instruments is the Information Literacy Competency Standards for Higher Education (Association of College and Research Libraries [ACRL], 2000). Three IL Tests developed based on the Information Literacy Competency Standards for Higher Education are the Standardised Assessment of Information Literacy Skills (SAIL), the Madison Information Literacy Test, and the iSkills™ Assessment (Boh Podgornik et al., 2016). Gross and Latham (2012) employed an ACRL-based assessment to identify students with subpar skills and discovered that these students frequently overestimate their competencies.

Another IL model, which served as a guide and reference in developing the IL test, is the SCONUL Seven Pillars of Information Literacy model (Society of College, National, and University Libraries Association [SCONUL], 2011). The IL model consists of seven pillars, which, along with other awareness statements (understands) and performance challenges (can), constitute the basis for information literacy practices in higher education in the UK and Ireland. Several studies have developed questionnaires that are directly based on the SCONUL Seven Pillars Model. These questionnaires often employ Likert scales or multiple-choice formats to assess the abilities of respondents in each pillar or competency, as demonstrated in research conducted with university students and academic staff (Malik et al., 2022).

Another IL standard used in developing IL assessments is the Framework for Information Literacy for Higher Education (ACRL, 2015). Due to the changing landscape in information, data, media, and technology, the Framework was introduced in 2016 to replace the Information Literacy Competency Standards for Higher Education. Your Information Literacy Practices (YILP) is one of IL's online measurements, which utilises the Framework as a reference in developing the assessment items (Kennette & McIntosh, 2022).

Although various IL tests and questionnaires have been developed and applied across diverse environments, target groups, and educational systems, there appears to be no general agreement on the most effective method or specific IL test format for assessing the information literacy of students in higher education.

c) Analysis of Test Items

Analysing test items is one of the key elements in developing a high-quality assessment instrument. According to Walsh (2009), many studies need to make more attempts to evaluate the reliability or validity of their assessment instruments for information literacy. In the context of information literacy (IL), the primary aim of an assessment is to evaluate students' skills in IL. Examples of utilising IL assessments include evaluating the efficiency of IL instructional programs, aligning instruction with IL learning objectives, assessing the impact of modifications in instructional programs, and enhancing the assessment process itself (Walters et al., 2020). According to Gratch-Lindauer (2003), developing a test locally offers the advantages of creating an assessment that aligns with the specific curriculum of a course or program, the flexibility to employ diverse formats that cater to local requirements, and a heightened probability that instructors will utilise the data. Oakleaf (2008) agrees that a locally developed test has other advantages, such as local control of assessment, evaluation, and interpretations of the assessment. Numerous efforts have been undertaken to evaluate information literacy through an objective test evaluation method. One way involves fixed-choice assessments, including multiple-choice (MCQs), matching, and true-false formats (Mery et al., 2011). Walsh (2009) asserts that, following his evaluation of nine types of assessment instruments, the multiple-choice question is the predominant approach for testing students' competencies in information literacy. According to Rosman et al. (2015), the MCQ type of test is "easy to administer while also maximising scoring objectivity" (p. 1).

In comparison to other assessment instruments, tests that are developed locally and consist of fixed-choice questions, such as MCQs, matching, and true-false tests, are easier and straightforward to create (Walsh, 2009). If questions in MCQs are well-constructed, MCQs can be "the most convenient, fair, and inexpensive way to judge an individual's level of performance or readiness" (Elmas et al., 2018, p. 14). Locally developed tests are also capable of quantifying student learning outcomes and facilitating comparisons between groups when utilised as pre- and post-assessments in a course (Oakleaf, 2008). MCQs developed locally also have a drawback. One of them is in terms of know-how. This type of test is unable to address higher-order abilities since it only captures declarative knowledge.

This is unsurprising, as numerous brief assessments are primarily intended to measure knowledge and skills acquired during library instruction sessions, particularly given the challenges associated with assessing reliability in short multiple-choice tests. However, some studies discuss how their locally-developed test items were tested and improved their overall reliability and validity. One is an online MCQ test, the Information Literacy Test (ILT), developed by the James Madison University (JMU) Library. Cameron et al. (2007) assert that their objective was to create a dependable and valid assessment that other institutions could utilise to evaluate information literacy competencies in accordance with the ACRL Information Literacy Standards. Following the preliminary pilot testing and the removal of problematic items, a 60-item assessment was administered to 524 students. The test analysis indicated strong overall reliability; however,

individual questions assessing a specific IL standard exhibited lower reliability. Content validity was assessed by three reference librarians, who evaluated the test items against the ACRL standards, categorising them as matching, not matching, or unclear. Content validity was assessed as "substantial," with inter-rater agreement on 42 of the 60 questions, and two librarians concurred on 59 out of 60 items (p. 232). O'Connor et al. (2002) employed a singular cohort of 554 students, following preliminary pilots, to evaluate the reliability of their instrument, which was conducted as the initial phase of Project SAILS. This project has developed a standardised information literacy platform available for licensing by other entities. They employed a Rasch model to analyse the distribution of test scores. They determined that the majority of their test items were reliable, asserting that the items collectively assessed at least a segment of the information literacy trait (p. 540). Analysis of test items is not only applicable to test-based assessment. A study by Kurbanoglu et al. (2006) elaborates on the process of item analysis and item discrimination indices used to address the validity of the items on the perception-based Information Literacy Self-Efficacy Scale.

Materials and Methods

The SCONUL Seven Pillars model was selected as the theoretical foundation for this study due to its widespread acceptance in higher education and its structured alignment with core information literacy behaviours, especially in higher education settings. This model provided a clear basis for identifying relevant competencies to be assessed, particularly for new learners in ODDE environments. The use of three-option multiple-choice questions (MCQs) was methodologically justified as it supports cognitive clarity, reduces random guessing, and limits distractor fatigue, without compromising the validity of the assessment. The decision to adopt a scenario-based format was intended to increase ecological validity, ensuring that items reflected real-life problem-solving situations that ODDE learners are likely to encounter in their daily lives. This approach enhances the relevance and authenticity of the test. Classical Test Theory (CTT) was selected as the analytical framework due to its suitability for pilot-stage test development and its manageable requirements in terms of sample size and statistical modelling, especially when Item Response Theory (IRT) was not yet practical.

Several validation stages were conducted to assess the quality of the test. Face and content validation were first conducted through expert review. A panel of four specialists in information literacy and education evaluated the clarity, relevance, language, and conceptual alignment of each item using a structured evaluation form. Following expert review, the test was pilot-tested with a sample of 35 new undergraduate learners at Open University Malaysia (OUM), representing diverse demographics in terms of age, gender, and faculty. The item analysis was conducted using CTT techniques, which included calculating the Difficulty Index (DIF I) to determine the proportion of correct responses, the Discrimination Index (DI) to assess how well items differentiated between higher- and lower-performing learners, and the Distractor Effectiveness (DE) to identify the functionality of incorrect options. Reliability testing was performed using the Kuder-Richardson Formula 20 (KR-20) and Cronbach's alpha, both of which indicated strong internal consistency.

Based on the item analysis findings and expert input, items with weak discrimination, extreme difficulty levels, or non-functional distractors were revised or removed, contributing to the refinement of a more valid and reliable test instrument.

Several meeting sessions were conducted and attended by two academics and two librarians from the Faculty of Education and the Faculty of Applied Sciences, respectively. Items were developed based on the Society of College, National and University Libraries (2011) Seven Pillars of Information Literacy model. The model consists of "a series of statements relating to a set of skills/competencies and a set of attitudes/understandings" of information literacy for higher education (SCONUL, 2011, p. 4). The seven pillars of information literacy are (1) Identify, (2) Scope, (3) Plan, (4) Gather, (5) Evaluate, (6) Manage, and (7) Present. For this study, the first three (3) pillars are selected and used as constructs in developing the scenario-based, three-option multiple-choice questions (MCQ). The number of options in MCQ tests often relies on various aspects, including the examined content, the cognitive level necessary for responding, test creation resources, and historical question statistics (Garcia et al., 2024).

Sample and Participants

The study was conducted with new undergraduate learners at Open University Malaysia (OUM), comprising 388 participants from one cohort in May 2024. The sample size was determined based on recommendations for item analysis in test validation studies. As the assessment instrument consisted of 60 multiple-choice items, the minimum recommended sample size is at least 5–10 respondents per item (Lakens, 2022). Therefore, a sample size of at least 300 was required, and the actual number of participants (n = 388) exceeded this threshold, ensuring adequate statistical power for reliability and validity testing.

Participants were drawn from four faculties:

- Faculty of Education (n = 102; 26.3%)
- Faculty of Science and Technology (n = 88; 22.7%)
- Faculty of Business and Management (n = 113; 29.1%)
- Faculty of Social Sciences and Humanities (n = 85; 21.9%)

The majority of respondents were adult learners aged between 25 and 65 years, reflecting the demographic composition typical of the institution's open and distance education model. The age distribution was as follows:

- 25–34 years: 98 participants (25.3%)
- 35–44 years: 141 participants (36.3%)
- 45–54 years: 106 participants (27.3%)
- 55–65 years: 43 participants (11.1%)

Gender representation was reasonably balanced, with 210 female learners (54.1%) and 178 male learners (45.9%). Most participants were employed full-time, consistent with the university's profile of adult learners. To ensure representativeness, stratified random sampling was applied across faculties. This approach helped reflect the distribution of new learners in each academic division while minimising sampling bias. The sample's demographic characteristics closely mirror those of the overall new student population at OUM, according to internal enrolment data for the May 2024 cohort. Participants were recruited through official faculty communication channels. The researchers collaborated with faculty administrators to distribute invitations via email and the university's learning management system. Participation was voluntary, and informed consent was obtained digitally before administering the instrument.

Instruments and Procedures

The test administration was conducted in two phases: a pilot test phase and the main test phase. Participants were informed about the voluntary nature of their involvement, and electronic informed consent was obtained before participation. All responses were anonymised, and data were stored securely to protect participants' confidentiality.

The pilot test was conducted with a sample of 35 undergraduate learners from a similar demographic profile to that of the main study participants. This was done to evaluate the reliability and clarity of the test items. The pilot study results indicated acceptable internal consistency, with a Cronbach's alpha value of 0.82, suggesting that the items were suitably reliable for measuring information literacy in an open and distance learning context. Several minor revisions were made based on feedback and item analysis, which included rewording ambiguous items and adjusting distractors in multiple-choice questions.

For the central test administration, the finalised version of the instrument—comprising 60 multiple-choice questions based on the SCONUL Seven Pillars of Information Literacy—was distributed online using the university's Learning Management System (LMS). The online multiple-choice questions (MCQs) consist of three possibilities, including one correct answer and two incorrect options, known as distractors. A correct answer earns one point, whereas an incorrect answer receives zero points. The top 27% (105) are classified as high achievers (H), whereas the bottom 27% are classified as poor achievers (L) (Torres et al., 2014). Each item was analysed for four indices, namely, difficulty index (DIF I), discrimination index (DI), distractor effectiveness (DE), and internal consistency reliability.

Post-validation of the paper was done through item analysis. Using Truman Kelley's "27% of sample" group size, the scores of all learners were arranged in order of merit. The upper 27% (n = 105) of the students were considered high achievers, and the lower 27% (n = 105) were considered low achievers. Each item was analysed for:

a) Difficulty Index and P-value

The item difficulty index (DIF I) refers to the percentage of the total number of correct responses of the upper 27% and lower 27% of the respondents to the total of the two groups. The calculation is performed using the formula DIF I = (H + L)/N, where H represents the number of correct responses among the upper 27% (n = 105) of the respondents and L is the total number of incorrect responses among the lower 27% (n = 105) of the respondents. N represents the total number of individuals in both groups. A higher index value indicates lower difficulty of the item, while a lower index signifies higher difficulty.

It is using formula DIF I = (H + L)/N ×100

H = Number of students answering the item correctly in the high-achieving group (upper 27%)

L = Number of students answering the item correctly in the low-achieving group (lower 27%)

N = Total number of students in the two groups

The DIF-I is represented as a P-value. It indicates the percentage of learners who accurately respond to questions on a test. The DIF-I can vary from 0 to 100 per cent. Items scoring above 70% are considered easy; those between 30% and 70% are regarded as having average or acceptable difficulty, while items scoring below 30% are classified as difficult (Velou & Ahila, 2020).

b) Discrimination Index (DI)

The discrimination index (DI) or d-value is the ability of an item to differentiate between respondents of higher and lower abilities. It ranges between +1 and -1. According to Truman Kelley, values of 0.4 and above are regarded as high, and values less than 0.2 are considered low (Ebel, 1954).

DI is calculated using the formula of DI = 2 x (H-L)/N

H = Number of students answering the item correctly in the high-achieving group (upper 27%)

L = Number of students answering the item correctly in the low-achieving group (lower 27%)

N = Total number of students in the two groups

c) Distractor Effectiveness (DE) or Functionality

Distractor effectiveness (DE) is calculated for each item according to the quantity of non-functional distractors (NFDs). NFDs are distractors with a selection rate of 5% or less of the top 27% and the lower 27% of the respondents. The items were classified according to the number of non-functional distractors (NFDs) in multiple-choice questions (MCQs). Two NFDs corresponding to a distractor effectiveness (DE) of 0% (low DE), one NFD of 50% as moderate DE, and zero NFD of 100% as high DE (Walvoord, 2010).

d) Internal Consistency Reliability

Cronbach's alpha is a metric that measures the internal consistency of a set of items, indicating the degree to which they are closely related to one another. It serves as an indicator of dependability. Alpha values between 0.8 and 0.9 are considered to be within a satisfactory range (Cronbach, 1951). The Kuder–Richardson Formula 20 (KR-20) is an index of reliability for assessments utilising binary variables or dichotomous data (outcome having only two values). Its score range is from 0 to 1, with 0 indicating no dependability and 1 signifying excellent reliability (Kuder & Richardson, 1937). A score approaching 1 indicates greater test reliability. The level of skewness and kurtosis was also determined for the item set. Skewness is a measure of lack of symmetry. Kurtosis assesses whether data exhibit heavy or light tails compared to a normal distribution. Skewness and kurtosis values between -2 and +2 are deemed adequate for demonstrating normal distribution.

Data Analysis

Data analysis was conducted using MS Excel 2021 MSO and JASP version 0.19.0.0. Both of which are suitable for descriptive and inferential statistical procedures commonly employed in educational research. Before analysis, data were screened for missing values. Cases with more than 10% missing responses were excluded from the analysis to preserve data quality. For cases with minor missing data (less than 10%), pairwise deletion was applied where applicable, particularly in correlation analysis, to maximise the use of available data without introducing bias. This approach ensured that the integrity of the dataset was maintained while minimising the impact of missing responses on the overall results.

Results

The result of the study shows that the scores of 388 new learners range from 23% to 90%, indicating a wide distribution of ability levels within the cohort. This variability is desirable in norm-referenced tests as it allows effective differentiation among learners with different levels of IL competence. As shown in Table 1, the test scores of the class ranged from 14 to 54 (with a total score of 60) and had a mean test score of 36.88 ± 8.97. The median was 38. As displayed in Figure 1, the kurtosis value (-0.174) is close to zero, indicating a distribution that approximates normality, with no significant presence of outliers or extreme scores. This supports the appropriateness of using parametric tests for further statistical analysis, such as correlation or t-tests. The internal consistency reliability of the instrument is strong, with a Cronbach's alpha of 0.85 and a Kuder-Richardson 20 (KR-20) value of 0.86. These values exceed the commonly accepted threshold of 0.70, indicating that the test items are homogeneous and consistently measure the intended construct—namely, information literacy knowledge. Hence, the hypotheses are as follows: H1, The overall test will exhibit high internal consistency, with KR-20 and Cronbach's alpha values exceeding 0.80, indicating strong reliability.

These findings suggest that the test demonstrates satisfactory psychometric properties. It provides a reliable measure of IL performance across a diverse learner population and is capable of distinguishing varying levels of ability. The near-normal distribution and broad score range further support its suitability for large-scale diagnostic purposes in ODDE settings.

Figure 1

Distribution of test scores

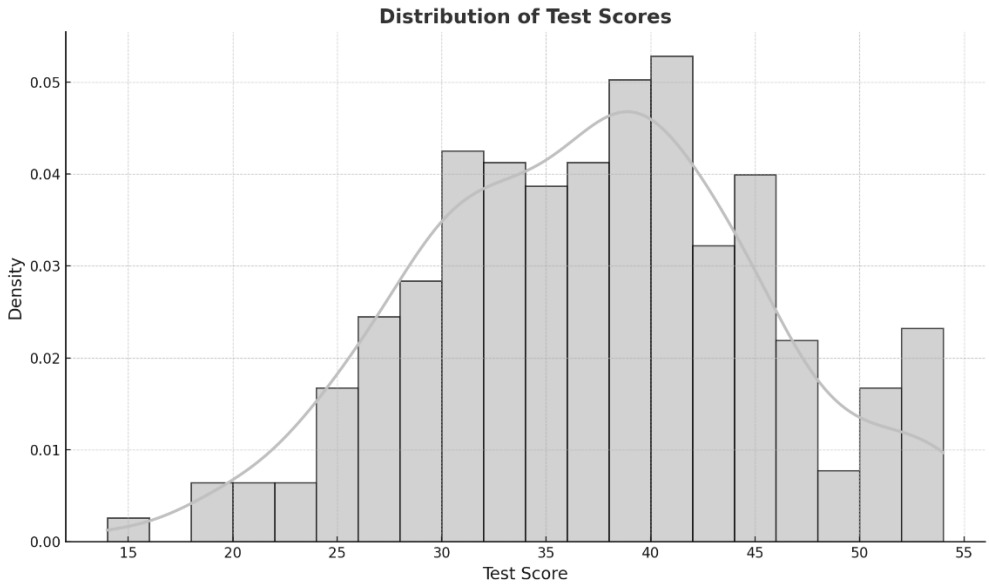

Table 1

Descriptive results of the test score

Parameters	Result
No. of items	60
Mean of test score	36.88 ±8.97
Percentage of mean test score (%)	61.32±14.99
Median	38
Range of test score	14-54
Range of test score (%)	23-90
Skewness	-0.503
Kurtosis	-0.174
Cronbach alpha	0.85
Kuder-Richardson 20 (KR-20)	0.86

Source: Fieldwork, 2024

As shown in Table 2, of the total 60 multiple-choice questions, 41 items (68.3%) fall within the recommended range of 30–70%, with a mean difficulty index of 55.59% (SD = 10.63). This indicates that the majority of items are of good or acceptable difficulty, suitable for assessing learners with varying levels of information literacy. Notably, 10 of these items (16.7%) fall within the narrower and more desirable range of 50–60%, with a mean difficulty of 56.33% (SD = 3.16), suggesting that these items are especially effective at discriminating around the average level of ability. Based on the findings, Hypothesis 2 (H2): The majority of test items will fall within the acceptable difficulty range (30%–70%), indicating that appropriate item challenge levels for the target population are supported.

Only two items (3.2%) were found to be too difficult (DIF I < 30%), with a mean difficulty index of 25.0% (SD = 3.03). These items may be too challenging for the intended learners and need further review to identify potential issues related to content clarity, alignment, or instructional relevance. In contrast, 17 items (28.3%) were classified as too easy (DIF I > 70%), with a mean of 70.74% (SD = 3.46). While easy items can support learner confidence and ensure broad content coverage, a high proportion may reduce the overall discriminative power of the test, particularly for high-performing individuals.

Overall, the test demonstrates a well-balanced distribution of item difficulty, with the majority falling within acceptable bounds. However, the presence of a relatively high percentage of easy items suggests the potential for refinement to improve the test's capacity to differentiate among learners, especially those at the upper end of the performance spectrum. Future iterations may benefit from the replacement or modification of overly easy or difficult items, guided by ongoing item analysis and expert review.

Table 2

Classification of items according to the difficulty index

DIF I Difficulty index (%)	Interpretation	Items (%)	Difficulty index (mean ± SD)
< 30	Too difficult	2 (3.2)	25.0 ± 3.03
30 - 70	Good / Acceptable	41 (68.3)	55.59 ± 10.63
50 - 60	Excellent	10 (16.7)	56.33 ± 3.16
> 70	Too easy	17 (28.3)	70.74 ± 3.46

Source: Fieldwork, 2024

The discrimination index (DI) is a key metric in evaluating the quality of test items, particularly in determining how well each item differentiates between high- and low-performing learners. A higher DI indicates that an item is effective in distinguishing learners who possess the targeted knowledge or skills from those who do not. The analysis of the 60-item multiple-choice information literacy test reveals promising results in terms of item discrimination.

As indicated in Table 3, 90% of the questions demonstrate good to excellent discrimination (DI ≥ 0.2). Hence, hypothesis 3, H3: At least 70% of the items will demonstrate good to excellent discrimination (DI ≥ 0.2), reflecting the test's ability to differentiate learners' performance is accepted.

Over half of the items (28 items or 56.7%) fall into the "excellent" category (DI ≥ 0.40), with a mean DI of 0.47 (SD = 0.06). This suggests that these items are highly effective at distinguishing between learners with higher and lower information literacy competencies. The relatively low standard deviation within this group reflects consistency in item performance. Another 26 items (43.3%) are classified as "good" (DI between 0.20 and 0.39), with a mean DI of 0.33 (SD = 0.10). These items are still considered acceptable for assessment purposes, particularly in formative evaluations, though they are slightly less discriminative compared to the excellent items. The higher standard deviation in this range may indicate greater variability in the performance of individual items. Only six items (10%) fall into the "poor" category (DI < 0.20), with an average DI of 0.08 (SD = 0.05). Items in this category do not meaningfully differentiate between high and low scorers, which may be due to ambiguous wording, flawed distractors or a mismatch with the test construct. These items should be reviewed and either revised or discarded in future test iterations to improve the overall psychometric quality of the instrument.

Overall, the data show that 85% of the test items possess acceptable to excellent discriminatory power. This indicates a strong capacity of the test to distinguish learner performance levels, supporting its reliability and usefulness in measuring information literacy competencies among new undergraduate learners in Open, Distance, and Digital Education (ODDE) settings.

Table 3

Classification of items according to the discrimination index (DI)

Discrimination Index (DI)	Interpretation	Items (%)	Discrimination index (mean ± SD)
< 0.20	Poor discrimination	6 (10.0)	0.08 ± 0.05
≥ 0.2	Good to excellent discrimination	54 (90.0)	0.38 ± 0.08
0.20 – 0.39	Good discrimination	26 (43.3)	0.33 ± 0.10
≥ 0.4	Excellent discrimination	28 (56.7)	0.47 ± 0.06

Source: Fieldwork, 2024

The distractor analysis of the 60-item multiple-choice test offers essential insights into the quality and diagnostic value of the assessment instrument. As shown in Table 4, out of 120 distractors for 60 items, the majority, 102 (85%), were classified as functional distractors (FDs), while 18 (15%) were identified as non-functional distractors (NFDs). Based on the findings, hypothesis 4, H4: The majority of distractors will function effectively (selected by at least 5% of respondents), demonstrating that plausible and contextually relevant incorrect choices are accepted. In addition,

of all 60 items, 2 (3.3%) had a low distractor effectiveness (DE), 36 (60%) had moderate DE, and 22 (36.7%) had a high DE.

This is a positive indication of item quality, as functional distractors play a key role in differentiating between high- and low-performing test-takers. As only 15% of the 120 distractors were classified as non-functional distractors (NFDs), this indicates that they were either never selected or chosen by fewer than 5% of respondents. The presence of NFDs reduces the overall discriminatory power of an item, suggesting opportunities for improvement through item revision.

Further analysis of item-level DE reveals that 22 items (36.7%) had high distractor effectiveness, with all three distractors functioning well. This indicates strong item performance and effective option construction. A majority of the items, 36 (60%), had moderate distractor effectiveness, containing one NFD. While these items still retain diagnostic value, revising or replacing the non-functional distractors could enhance their quality. Only 2 items (3.3%) were identified as having low distractor effectiveness, with each having two non-functional distractors. These items may fail to distinguish between learners of differing ability levels effectively and warrant closer scrutiny for improvement or replacement.

Overall, the findings demonstrate that the test is reasonably well-constructed, with a high proportion of functioning distractors and relatively few poorly performing items. However, systematic refinement of the items with NFDs—particularly those with low DE—would further improve the psychometric soundness of the test. Continuous monitoring and revision based on distractor performance are crucial for maintaining the test's validity and ensuring its ability to assess information literacy among diverse learners accurately.

Table 4

Distractor analysis and distractor effectiveness (DE) of MCQs test items

Parameter	Number (%)
Number of items	60
Total distractors	120
Functional distractors (FDs)	102 (85)
Non-functional distractors (NFDs)	18 (15)
Number of items with 2 NFDs (Low DE)	2 (3.3)
Number of items with 1 NFD (Moderate DE)	36 (60.0)
Number of items with 0 NFD (High DE)	22 (36.7)

Source: Fieldwork, 2024

Pearson's correlation of the difficulty index (DIF I) and discrimination index (DI) was also analysed for each of the MCQ items. Table 5 shows the correlation between DIF I and DI. The results show Pearson's r = 0.222 and a p-value of 0.089. This reflects a weak positive linear relationship,

suggesting that, in general, items that are slightly more difficult tend to have marginally better discrimination power. However, the association does not reach statistical significance at the conventional alpha level of 0.05. This means that the likelihood of the observed relationship occurring by chance is relatively high (8.9%).

From a psychometric perspective, this result suggests that there is no clear relationship between the difficulty of an item and its ability to differentiate between high- and low-performing students. Ideally, well-constructed test items should show moderate difficulty (DIF between 0.4 and 0.6) and high discrimination (DI ≥ 0.4). In this analysis, the lack of a significant correlation may signal inconsistencies in item construction or that some items are not functioning as intended psychometrically.

Table 5

Pearson's Correlations between difficulty index, DIF I and discrimination index, DI

Variable		DIF I	DI
DIF	Pearson's r	-	
	p-value	-	
DI	Pearson's r	0.222	-
	p-value	0.089	-

Source: Fieldwork, 2024

The 60-item MCQ-based information literacy test has overall good psychometric quality, although improvements are still needed for some items. The high internal consistency (α = 0.85; KR-20 = 0.86) and near-normal distribution support the reliability of the instrument for assessing information literacy among new undergraduate learners. Most items had acceptable difficulty levels (68.3%) and demonstrated strong discrimination ability (85%), with 85% of distractors performing effectively. Although the correlation between item difficulty and discrimination was weak and not statistically significant, the test overall shows good psychometric quality for measuring information literacy knowledge. These results align with the research aims of developing a reliable and valid MCQ-based test and evaluating item quality through psychometric analysis. In addition, the item analysis provides clear direction for improving low-performing items, especially those with poor discrimination or weak distractors, to finalise a better version of the test for future use. These findings provide a solid foundation for interpreting how well the test captures the intended construct of information literacy knowledge. The following discussion will explore these results in greater depth, linking them to existing literature and highlighting their implications for assessment practices in open and distance learning contexts.

Discussion

Item analysis is a straightforward yet effective process that facilitates the examination of observations, interpretation of students' acquired knowledge, and assessment of the quality of the

test item (Kumar et al., 2021). In this study, out of 60 items, the majority, specifically 41 items (68.3%), exhibited a good or acceptable level of difficulty index, DIF I, with a mean of 55.59 ± 10.63. Conversely, two items (3.2%) were classified as too difficult (DIF I < 30%), while 17 items (28.5%) were classified as too easy (DIF I > 70%). The good or acceptable level of difficulty index (68.3%) is considered low compared to previous studies, which reported 75% and 90% as the acceptable rates of items based on the difficulty index (Boh Podgornik et al., 2016; Lockhart, 2015). This low percentage of good and acceptable levels is due to a high percentage of 'too easy' items. Those items will be reviewed and improved.

In the study, 54 of the 60 items (90%) had an acceptable to excellent discrimination index (DI). It suggests that the majority of the items are good to excellent in differentiating the ability of the new learners with higher scores from those with lower scores. The DI values of the study are also comparable to those of a study by Lockhart (2015), which reported a high of 89% of the 100 test items with a difficulty index of good to excellent. According to Kumar et al. (2021), discrimination indices should be evaluated within the context of the specific test being analysed. DI values are often lower for assessments encompassing a broad spectrum of content areas compared to more homogeneous assessments. They also suggested that items with poor DI are often ambiguously worded and recommended a re-examination of the particular items.

The study also shows that two items of the test (3.3%) had two NFDs (low distractor effectiveness). It indicates that the two 'wrong options' are selected by less than 5% of the upper 27% and lower 27% of the respondents (210 learners). Those options should be reviewed and enhanced to ensure that they can optimally distract the learners' attention. Identifying potential functional distractors and minimising NFDs are crucial for the quality of multiple-choice questions.

The skewness and kurtosis values were -0.503 and -0.174, respectively. It indicates that the data distribution is slightly left-skewed, with slightly lighter tails and a peak, compared to a normal distribution. The numbers, however, indicate that the data distribution approximates normality, exhibiting only slight deviations from normality. The study also shows a weak positive correlation between the difficulty index, DIF I and the discrimination index, DI ($r = 0.222$, $p > 0.05$).

The study has several limitations. First, the sample, although large (n = 388), was drawn from a single cohort at one of the ODDE institutions in Malaysia, limiting its generalizability to other contexts. Cultural and educational backgrounds may influence how learners interpret questions, particularly in IL, where conceptual understandings can differ across systems. The use of a multiple-choice format, while efficient, may not fully capture higher-order IL skills such as synthesis or critical evaluation (Haataja et al., 2023). Additionally, items were developed in English, which may have introduced language-related barriers for some learners, especially in a multilingual country like Malaysia.

Theoretically, the study contributes to the growing body of IL assessment research by offering an empirically validated test based on the SCONUL Seven Pillars model, contextualised for open,

distance and digital education (ODDE). This addresses a critical gap in assessment tools that are both psychometrically sound and culturally appropriate. While many international IL tests exist, few are designed specifically for adult learners in flexible learning environments or reflect the realities of post-secondary learners in Southeast Asia. By tailoring test items to the Malaysian ODDE context and applying rigorous item analysis, the study bridges the gap between theory and practice, paving the way for better-informed IL interventions and more equitable learner support systems.

In future work, cross-cultural validation of the instrument and longitudinal tracking of IL development could further enhance its robustness and utility. Adjustments to item language and inclusion of open-ended formats might also enrich the assessment's depth and inclusivity.

Conclusions and Implications

Item analysis is a valuable procedure that should be conducted before a locally developed test is approved and becomes an assessment instrument. This is not an exception for information literacy test-based assessment. It provides essential information on the reliability and validity of test items by calculating the item difficulty index (IDI), item discrimination index (IDI), and distractor effectiveness (DE). According to Kumar et al. (2021), for MCQ-type assessments, a well-constructed item with four options should ideally have a moderate difficulty index (DIF I = 30–70%), a strong discrimination index (DI ≥ 0.4), and maximum distractor effectiveness (DE = 100%) with three functional distractors (FDs). These standards can also be applied to three-option MCQs with two FDs.

In this study, item analysis showed that 19 items met all three criteria. The difficulty index analysis found that 68% of the items were categorised as acceptable or good, while 90% had good or excellent discrimination. This indicates that the test is balanced—not too easy or too tricky—and can distinguish between learners with higher and lower ability levels.

The findings support the use of this test to offer a high-quality information literacy assessment for new undergraduate learners. The test can help measure the level of information literacy knowledge and identify areas of weakness among new students. The university and library can use this data to guide instructional interventions and improve learners' IL skills. The study also enables the university to offer a high-quality information literacy assessment through MCQs for new learners. The test will evaluate the level of information literacy knowledge and areas of weakness among incoming undergraduate learners. The university and library will implement interventions to enhance the level of information literacy knowledge and competencies of new undergraduate students, following improvements to the selected test items.

Suggestions for Future Research

To address the lack of a valid and reliable information literacy (IL) assessment tool for new learners in Open, Distance, and Digital Education (ODDE) environments, a targeted research agenda is needed. Specific research questions may include: (1) How can an IL test grounded in the SCONUL Seven Pillars be developed to suit learners in ODDE settings? (2) What are the

psychometric properties (validity, reliability, item difficulty, and discrimination) of the developed test? and (3) How does performance on the test correlate with learners' IL self-efficacy and academic outcomes? A mixed-methods design is recommended. The quantitative phase should involve test construction, pilot testing, and item analysis (difficulty index, discrimination index, and distractor effectiveness) using tools such as JASP. Reliability should be evaluated with KR-20 and Cronbach's alpha, and construct validity should be tested via expert review and exploratory factor analysis (EFA). The qualitative phase may involve interviews or focus groups to explore learners' experiences with the test and IL instruction. To support implementation, institutions should integrate the IL test into the student induction programme. This enables the early identification of skill gaps and allows for timely interventions, such as modular tutorials or librarian-led workshops. Capacity building should also be prioritised: academic librarians and instructional designers must be trained in test development, item writing, and psychometric evaluation. Over time, the question bank should be reviewed and refined through continuous data collection and analysis to ensure cultural and contextual relevance.

Acknowledgements

The authors would like to thank the Open University Malaysia (OUM) for providing a research grant (OUM-IR-2023-003) to support the conduct of this study. The authors also wish to acknowledge Thirumeni Subramaniam, PhD, for her valuable insights and foundational ideas that contributed significantly to the development of this article.

Conflict of Interest

None.

Funding

The authors declares that no financial support was received for this study.

References

Al-Qallaf, C. L. (2019). Information literacy assessment of incoming students in an information studies graduate program. *Global Knowledge, Memory and Communication*, *68*(3), 223–241. https://doi.org/10.1108/GKMC-07-2018-0062

American Colleague and Research Libraries (ACRL). (2000). *Information literacy competency standards for higher education*. https://alair.ala.org/server/api/core/bitstreams/ce62c38e-971a-4a98-a424-7c0d1fe94d34/content

Association of College and Research Libraries. (2015). *Framework for information literacy for higher education*. https://www.ala.org/acrl/standards/ilframework

Babo, R., & Suhonen, J. (2018). E-assessment with multiple choice questions: A qualitative study of teachers' opinions and experience regarding the new assessment strategy. *International Journal of Learning Technology*, *13*(3), 220-248. https://doi.org/10.1504/IJLT.2018.095964

Bent M., & Stubbings, R. (2011). *The SCONUL seven pillars of information literacy core mode*. The Society of College, National and University Libraries. https://senseandreference.wordpress.com/wp-content/uploads/2016/12/coremodel.pdf

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

Cameron, L., Wise, S. L., & Lottridge, S. M. (2007). The development and validation of the information literacy test. *College & Research Libraries*, *68*(3), 229–237. https://doi.org/10.5860/crl.68.3.229

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Ebel, R. L. (1954). Procedures for the analysis of classroom tests. *Educational and Psychological Measurement*, *14*(2), 352–364. https://doi.org/10.1177/001316445401400215

Elmas, R., Bodner, G. M., Aydogdu, B., & Saban, Y. (2018). The inclusion of science process skills in multiple choice questions: Are we getting any better? *European Journal of Science and Mathematics Education*, *6*(1), 13–23. https://doi.org/10.30935/scimath/9519

Garcia, A. N., Ebert, I. S., Sun, H.-P., Davis, J., & Wu, B. (2024). *Comparison of 3-option and 4-option multiple choice questions and related issues and practices*. Credentialing Insights. https://www.credentialinginsights.org/Article/comparison-of-3-option-and-4-option-multiple-choice-questions-and-related-issues-and-practices

Gross, M., & Latham, D. (2012). What's skill got to do with it? Information literacy skills and self-views of ability among first-year college students. *Journal of the American Society for Information Science and Technology*, *63*(3), 574–583. https://doi.org/10.1002/asi.21681

Haataja, E. S. H., Tolvanen, A., Vilppu, H., Kallio, M., Peltonen, J., & Metsäpelto, R.-L. (2023). Measuring higher-order cognitive skills with multiple choice questions –potentials and pitfalls of Finnish teacher education entrance. *Teaching and Teacher Education*, *122*, Article 103943. https://doi.org/10.1016/j.tate.2022.103943

Jones, G. (2020). Designing multiple-choice test items. In P. Winke & T. Brunfaut (Eds), *The Routledge Handbook of Second Language Acquisition and Language Testing* (1st edn, pp. 90–101). Routledge. https://doi.org/10.4324/9781351034784-11

Jones, W. L., & Mastrorilli, T. (2022). Assessing the Impact of an Information Literacy Course on Students' Academic Achievement: A Mixed-Methods Study. *Evidence Based Library and Information Practice*, *17*(2), 61–87. https://doi.org/10.18438/eblip30090

Kennette, L. N., & McIntosh, E. (2022). Your information literacy practices (YILP): Information literacy and resourcefulness. *Partnership: The Canadian Journal of Library and Information Practice and Research*, *17*(2), 1–22. https://doi.org/10.21083/partnership.v17i2.6680

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160. https://doi.org/10.1007/BF02288391

Kumar, D., Jaipurkar, R., Shekhar, A., Sikri, G., & Srinivas, V. (2021). Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Medical Journal Armed Forces India*, *77*, S85–S89. https://doi.org/10.1016/j.mjafi.2020.11.007

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), Article 33267. https://doi.org/10.1525/collabra.33267

Lockhart, J. (2015). Using item analysis to evaluate the validity and reliability of an existing online information literacy skills assessment instrument. *South African Journal of Libraries and Information Science*, *80*(2), 36–42. https://catalogue.leidenuniv.nl/discovery/fulldisplay?vid=31UKB_LEU:UBL_V1&docid=alma9939086066902711&lang=en&context=L&adaptor=Local%20Search%20Engine

Malik, A., Ali, S., Batool, S. H., & Ameen, K. (2022). Linking information literacy with research productivity: A survey of mathematicians in Pakistan. *Portal: Libraries and the Academy*, *22*(2), 475–498. https://doi.org/10.1353/pla.2022.0026

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

O'Connor, L. G., Radcliff, C. J., & Gedeon, J. A. (2002). Applying systems design and item response theory to the problem of measuring information literacy skills. *College & Research Libraries*, *63*(6), 528–543. https://doi.org/10.5860/crl.63.6.528

Oakleaf, M. (2008). Dangers and Opportunities: A Conceptual Map of Information Literacy Assessment Approaches. *Portal: Libraries and the Academy*, *8*(3), 233–253. https://doi.org/10.1353/pla.0.0011

Pinto, M., Garcia-Marco, J., Caballero, D., Manso, R., Uribe, A., & Gomez, C. (2024). Assessing information, media and data literacy in academic libraries: Approaches and challenges in the research literature on the topic. *The Journal of Academic Librarianship*, *50*(5), Article 102920. https://doi.org/10.1016/j.acalib.2024.102920

Rosman, T., Mayer, A.-K., & 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. https://doi.org/10.1080/02602938.2014.950554

Serap Kurbanoglu, S., Akkoyunlu, B., & Umay, A. (2006). Developing the information literacy self-efficacy scale. *Journal of Documentation*, *62*(6), 730–743. https://doi.org/10.1108/00220410610714949

Velou, M., & Ahila, E. (2020). Refine the multiple choice questions tool with item analysis. *International Archives of Integrated Medicine*, *7*(8), 80-85. https://www.iaimjournal.com/storage/2020/08/iaim_2020_0708_13.pdf

Walsh, A. (2009). Information literacy assessment: Where do we start? *Journal of Librarianship and Information Science*, *41*(1), 19–28. https://doi.org/10.1177/0961000608099896

Walters, W. H., Sheehan, S. E., Handfield, A. E., López-Fitzsimmons, B. M., Markgren, S., & Paradise, L. (2020). A multi-method information literacy assessment program: Foundation and early results. *Libraries and the Academy*, *20*(1), 101–135. https://preprint.press.jhu.edu/portal/sites/default/files/20.1walters.pdf

Walvoord, B. E. F. (2010). *Assessment clear and simple: A practical guide for institutions, departments, and general education* (2nd ed). Jossey-Bass.

Xu, X., Kauer, S., & Tupy, S. (2016). Multiple-choice questions: Tips for optimising assessment in-seat and online. *Scholarship of Teaching and Learning in Psychology*, *2*(2), 147–158. https://doi.org/10.1037/stl0000062

Yu, C. (2023). Integrating information literacy and academic writing: Developing a self-assessment scale of information-based academic writing. *The Journal of Academic Librarianship*, *49*(6), Article 102804. https://doi.org/10.1016/j.acalib.2023.102804

Zhou, X., & Jenkins, R. (2020). Dunning–Kruger effects in face perception. *Cognition*, *203*, Article 104345. https://doi.org/10.1016/j.cognition.2020.104345