Communications in Information Literacy

Volume 16 | Issue 2 Article 1

12-15-2022

Metacognitive awareness for IL learning and growth: The development and validation of the Information Literacy Reflection Tool (ILRT)

Sara Robertson
Portland Community College, sara.robertson@pcc.edu

Michele Burke
Chemeketa Community College, michele.burke@chemeketa.edu

Kimberly Olson-Charles
University of Western States, kolsonch@gmail.com

Reed Mueller
Bushnell University, rmueller@bushnell.edu

Recommended Citation

Robertson, S., Burke, M., Olson-Charles, K., & Mueller, R. (2022). Metacognitive awareness for IL learning and growth: The development and validation of the Information Literacy Reflection Tool (ILRT). *Communications in Information Literacy, 16* (2), 58–89. Retrieved from https://pdxscholar.library.pdx.edu/comminfolit/vol16/iss2/1

Metacognitive Awareness for IL Learning and Growth: The Development and Validation of the Information Literacy Reflection Tool (ILRT)

Sara Robertson, Portland Community College
Michele Burke, Chemeketa Community College
Kimberly Olson-Charles, University of Western States
Reed Mueller, Bushnell University

Abstract

This article describes the development and validation of the Information Literacy Reflection Tool (ILRT), a metacognitive self-assessment for use with undergraduate researchers. It was developed as a teaching and learning tool with the intent to help students recognize and engage the metacognitive domain as a step toward developing personal agency and self-regulation as lifelong, metaliterate learners. Throughout the scale development, three studies were conducted with nine expert reviewers and 44 community college students to consider content and face validity and 542 community college students as part of an item-reduction and construct validation effort. The resulting scale is most appropriately construed through a bifactor model and is made up of 57 items comprising a strong information literacy general factor and six specific factors modestly aligning with each of the threshold concepts outlined in the ACRL *Framework for Information Literacy for Higher Education.* As a part of the ongoing development of this instrument, work is needed to more fully assess reliability and validity.

Keywords: assessment, information literacy, self-reflection, scale development, threshold concepts, metacognition, undergraduates, metaliteracy, self-regulated learning, Framework for Information Literacy for Higher Education

Robertson, S., Burke, M., Olson-Charles, K., & Mueller, R. (2022). Metacognitive awareness for IL learning and growth: The development and validation of the Information Literacy Reflection Tool (ILRT). *Communications in Information Literacy, 16*(2), 58–89.

Metacognitive Awareness for IL Learning and Growth: Development and Validation of the Information Literacy Reflection Tool (ILRT)

This paper describes the development of the Information Literacy Reflection Tool (ILRT), a flexible and transparent self-assessment instrument which represents the six threshold concepts and the range of knowledge practices and dispositions laid out in the Association of College and Research Libraries (ACRL, 2015) *Framework for Information Literacy for Higher Education*. Because the *Framework* is written for IL educators and experts, it can be challenging to translate into meaningful language for undergraduate students. The ILRT uses clear language to provide a meaningful and representative description of the threshold concepts in the *Framework*, in a way that students can connect to their existing knowledge and experience. The ILRT was developed as a teaching and learning tool with the intent to help students recognize and engage the metacognitive domain as a step toward developing personal agency and self-regulation as lifelong, metaliterate learners.

Within metaliterate learning (Mackey & Jacobson, 2014), the *Framework* emphasizes metacognition as a domain that is central to lifelong learning within a dynamic information ecology (ACRL, 2015). In a systematic review of the intersections and opportunities for IL assessment and metacognitive strategies, Hostetler et al. (2018) found that metacognitive scales are an effective alternative to skills-based assessments. When used intentionally and with coaching from teachers, metacognitive self-assessments have been found to encourage various types of cognition, self-reflection, and self-regulated learning across a range of contexts and populations (Hostetler et al., 2018). Self-regulated learners use metacognition to monitor their cognitions, motivations, behaviors, and affects in order to make progress toward learning goals (Craig et al., 2020; Schunk & Greene, 2017).

Practice in metacognitive self-reflection asks students to identify where they already have facility or exposure and can be a mechanism for growth as information researchers (Doyle et al., 2019; Fulkerson et al., 2017). Metacognitive self-reflection is also important for the transfer of learning skills from one context to another (Billing, 2007; Kuglitsch, 2015). The development of this instrument builds on previous work (Catalano, 2017; Doyle et al., 2019; Schraw & Dennison, 1994) and initiates work on ILRT validation as a trustworthy and useful self-report measure.

The ILRT is not intended to measure proficiency or readiness for research; rather, it is centered as a teaching tool and a formative assessment. The ILRT is designed to promote self-reflection on one's own awareness and perceived use of IL knowledge practices and dispositions, and to support self-regulated learning and growth for individuals in and outside of classroom settings. In one aspect of this study, we attempted to verify whether the ILRT can measure and define each discrete threshold concept laid out in the *Framework* as well as overall information literacy as a construct. Ultimately, the researchers' goal is to provide a valid instrument that educators can use to help put students at the center of their own learning and make the complexities of the IL learning process more transparent.

This study draws upon literature in education, library and information science, and psychology. In the literature review, we introduce IL as a metaliteracy and examine the intersection of research on metacognition and IL. We then review the metacognitive self-report instruments that informed this study. Finally, we establish the need for valid IL teaching tools to use with first-year college students that align with the *Framework* and support metacognition.

Literature Review

IL as a Metaliteracy

While the rescinded *Information Literacy Competency Standards for Higher Education* (Association of College and Research Libraries, 2000) focused on skills required to find and use information effectively, the *Framework* claims IL as a metaliteracy described by Mackey and Jacobson (2011) by positioning the learner as both consumer and producer in dynamic information environments where multiple literacies overlap. Metaliteracy, or lifelong learning within modern information ecologies, necessitates reflecting and engaging in the intentional construction and evolution of multiple interrelated literacies (Mackey & Jacobson, 2014). As a teaching and learning tool, the ILRT works within traditional or metaliterate contexts to break down IL into descriptive prompts that assist the user in identifying and approaching IL as a contextual, personally evolving literacy. In positioning IL as a metaliteracy, the *Framework* notes that a learner will increasingly rely on metacognition to monitor, regulate, and coordinate engagement across the learning domains (the affective, cognitive, and behavioral), and it highlights metacognition as crucial to becoming more self-directed in rapidly changing information ecosystems . Metacognition in general is the awareness of and regulation of one's own thinking and cognitive assets, and in this sense, it acts as both engine and navigation system in directing momentum toward self-knowledge and critical reflection. Metacognitive awareness relates to thinking about the content being learned as well as about oneself as a thinker and learner. In the original definitions of metacognition, Flavell (1979) made the following distinction: "Cognitive strategies are invoked to *make* cognitive progress, metacognitive strategies to *monitor* it" (p. 909). This relies on a learner to first be aware of their own cognitive processes and then to develop strategies to monitor their learning over time.

People who have facility in information literacy rely on metacognition to draw on their existing knowledge, adjust a line of inquiry, and critically evaluate their own thinking. Fulkerson et al. (2017) asserted that metacognition ought to be central to teaching and learning environments as it instills in learners the long-term value of knowing how to plan and monitor their own processes. In the rapidly changing information environment, it means that educators must normalize the necessity of adaptation and critical thinking.

Metacognition is a level above mastery of content because it includes the awareness of the strategies, thought processes, and approaches that are used while learning (e.g., while navigating a task, puzzling through a problem, decision making, or following a process). Jacobson et al. (2021) emphasized this by stating, "Metacognition is a core concept in metaliteracy, just as it is in Self-Directed Learning (SDL) and Assessment as Learning (AaL). This congruence provides clear avenues for using metaliteracy's framework in ways that support SDL" (p. 87). Supporting SDL is another reason for educators to incorporate metacognition as an essential learning goal in the IL classroom. As a metacognitive tool, the ILRT provides an initial step towards building IL concepts and reflection into teaching and assessment and has the potential to support increased self-awareness.

Education Research on Self-Assessments

The ultimate goal for educators who incorporate metacognitive teaching practices is to create a learning environment that encourages enhanced awareness so that learners can become more independent and are more likely to transfer knowledge practices and dispositions in future contexts (Pintrich, 2002). Andrade's (2019) comprehensive review of research on self-assessment tools established that for first-year students, especially those who are new to academic environments, formative assessments can be particularly valuable because they provide an effective self-evaluation opportunity to monitor and better understand their own learning. Andrade (2019) also described how the focus of the self-assessment is worth considering, for example, on one's own competence, judgments, or ability to follow an individual's progress towards targets, or an assessment of a final product. The ILRT is a self-assessment that builds awareness of the practices and knowledge applied to situations relying on IL. The scale developed here is an example of a self-assessment that promotes awareness of and reflection on IL concepts and strategies, but does not measure competency, skill, or achievement.

Existing Scales for IL

Two self-report instruments exist that measure metacognitive awareness for IL: the Metacognitive Strategies for Library Research Skills Scale (MS-LRSS; Catalano, 2017) and the Perception of Information Literacy Scale (PILS; Doyle et al., 2019). The MS-LRSS was tested with adult learners and consists of a 21-item inventory using a 5-point scale (*not at all, slightly, moderately so, very much so, extremely*). Four subscales were identified that represent metacognitive constructs, including awareness, self-checking and debugging, planning, and cognitive strategy. After the initial item list was developed and reviewed by two IL experts, the MS-LRSS was tested with a diverse population of adults enrolled in undergraduate and graduate courses (n = 207 total, n = 74 first-year students). The subscales were developed using Shraw and Dennison's (1994) Metacognitive Awareness Inventory (MAI). The ILRT study also builds on the MAI subscales to operationalize metacognition within the domain of IL for both "knowledge of cognition" (which includes declarative knowledge, procedural knowledge, and conditional knowledge) and "regulation of cognition" (which includes planning, information management, monitoring, debugging, and evaluation). The metacognitive constructs defined by the MAI and retested by Catalano in the development of the MS-LRSS were used as guides for the methodology steps taken for our studies of the ILRT. When writing and revising the initial item set for the ILRT, the researchers looked to the MAI items as examples, and aimed to capture the essential concepts covered by each disposition and knowledge practice of the *Framework.*

After the initiation of this research study, the Perception of Information Literacy Scale (PILS) was developed by Doyle et al. (2019) using a graduate student population from a diverse range of academic areas of study. The initial items were written based on the dispositions and knowledge practices articulated in the *Framework*, and content validation was carried out by expert review. PILS uses a self-assessment 7-point scale from novice to expert, appropriate for adult learners and in alignment with Dreyfus's five-stage model of skill acquisition. Confirmatory factor analysis established construct validity with a good fit for seven factors, including subscales that align with the six threshold concepts in the *Framework,* where the Searching as Strategic Exploration frame is defined in two subscales (tool & tasks; mindset) for a 36-item scale. Doyle et al. (2019) also uses the knowledge practices and dispositions of the *Framework* as a base for item development, and tested content validity with an expert review.

Outside of IL there are well established metacognitive scales that can serve as models for IL researchers, such as the Metacognitive Awareness Reading Strategies Inventory (MARSI), which is for use in high school and early college reading instruction (Mokhtari & Reichard, 2002). Speth et al. (2007) noted that the Approaches and Study Skills Inventory for Students (ASSIST) is another example of a self-assessment scale that promotes deeper and more strategic learning, but it is a more generalized approach to learning habits of mind and strategies. It is clear there are few established instruments specific to IL learning for novice researchers and more rigorous research is required to better establish validity and reliability of the existing tools (Hostetler et al., 2018; Mahmood, 2017).

Establishing an IL Metacognitive Scale for First Year College Students

While Catalano's MS-LRSS (2017) and Doyle et al.'s PILS (2019) are much-needed additions to this body of research, they each have limitations that the ILRT attempts to address. The MS-LRSS has gaps in the scope and content that the *Framework* demands of learners. Individual items in the MS-LRSS are written based on the ACRL *Information Literacy Competency Standards* (2000), which are skills-based and more process-focused than the *Framework.* Additionally, the MS-LRSS is positioned as "domain dependent" because it "asks the respondent to consider their responses in the context of a specific assignment" (Catalano, 2017, p. 180). Rather than framing IL as skills used to complete a specific assignment, the ILRT encourages self-reflection around the transferable IL metaliteracy concepts described in the *Framework.*

Neither the MS-LRSS nor the PILS focus on first-year college students. Although the ILRT may be applied with broader audiences, the intent is to provide an instrument with established reliability and validity tested with a population of first-year college students, with a developmental approach to building metacognitive awareness around IL. The ILRT articulates and tests the foundational concepts and practices represented in the *Framework* to provide a flexible and appropriate reflection scale aligned with core metacognitive constructs.

The ILRT uses a Likert-type scale that prompts respondents to reflect on themselves using frequency; participants self-report a percentage of time the statement item is "true of me," from *very untrue of me (0% of the time)* to *very true of me (100% of the time)*. This differs from a novice-to-expert scale or a scale that asks respondents to self-assess proficiency. Instead of assigning value to their abilities, respondents are simply defining their own relationship to that statement, which aims to promote critical reflection. The statement prompts also serve as reference points for future discussion, for example, with peers or after further instruction and practice. The emphasis on reflection is especially important for novice researchers because the unskilled are more likely to inflate responses when self-assessing ability (Hostetler et al., 2018; Kruger & Dunning, 1999). Moreover, the researchers were motivated by the need for a reliable formative self-assessment tool that could be used in the course of teaching (instead of a pre- or post- test, for example) to further specific learning goals.

Based on the needs identified in the review of existing metacognitive assessment tools, we developed the ILRT to prioritize student-centered self-reflection in alignment with the *Framework* as a metaliteracy.

Method and Results: Studies 1, 2, and 3

Overview

Development of the ILRT proceeded in three phases, each with unique samples recruited after receiving IRB approvals. Across these phases, we refined individual item drafts and reduced the total number of items in the ILRT through expert review (Study 1), pilot testing with a student sample (Study 2), and construct validation efforts using Exploratory Structural Equation Modeling (ESEM) methods (Study 3).

Study 1: Expert Review

This study focused on reviewing draft items for possible inclusion in the pilot version of the ILRT to be used in Study 2. Specifically, we were interested in an initial assessment of content validity, which "concerns the degree to which a sample of items, taken together, constitute an adequate operational definition of a construct…evaluate[d] through expert assessment" (Polit & Beck, 2006, p.409). Toward this end, Study 1 was comprised of two stages.

First, three of the researchers developed 101 provisional item prompts. When creating these prompts, the authors worked to map draft prompts across the six information literacy frames defined by the ACRL (2015): Authority Is Constructed and Contextual (16 prompts), Information Creation as a Process (14 prompts), Information Has Value (13 prompts), Research as Inquiry (24 prompts), Scholarship as Conversation (15 prompts), and Searching as Strategic Exploration (19 prompts).

Second, the researchers utilized regional library science networks to recruit content areas experts. Each expert participant reviewed each prompt for appropriate or inappropriate mapping to proposed frames, inter-item breadth within frames, redundancy with other items, and lack of clarity or precision. Expert feedback for each prompt included narrative responses and Likert-type ratings to assess item-to-frame fit.

Sample

Inclusion criteria for entry into the study required content experts to hold either a position as an information literacy instruction librarian at any academic level or a writing instruction position at the college level. Altogether, 18 potential participants were invited to participate in the study, 13 agreed to participate, and of those 13, nine entered into the study; thus, 50% of those recruited enrolled as participants in Study 1.

Within this purposive sample of nine experts, one (11.1%) worked at the high school level, one (11.1%) worked at a state library, five (55.5%) worked in two-year college settings, and two (22.2%) worked in four-year college settings. Within this sample, one participant reported that their primary role in their work setting was that of writing instructor, while the other eight participants indicated that their primary role was in a library rather than a classroom setting.

Data Collection

When prompting feedback on potential items, expert participants were first provided frame definitions (e.g., Authority Is Constructed and Contextual: "Information resources reflect their creators' expertise and credibility, and are evaluated based on the information need and the context in which the information will be used. Authority is constructed in that various communities may recognize different types of authority. It is contextual in that the information need may help to determine the level of authority required." [ACRL, 2015, p. 12]). Experts then provided written qualitative feedback collected in individual anonymized documents for each potential item regarding strength of relationship to the frame, word choice, clarity, and identification of gaps in conceptual coverage. Written feedback also included suggestions for ensuring all essential concepts are covered within frames, addressing potential misinterpretations of the intent of frames, as well as observations related to the set of prompts as a representation of the *Framework* as a whole. In addition to qualitative written feedback, reviewers also provided Likert-type "fit" ratings for each of the candidate prompts for later use in constructing item content validity index (I-CVI) scores (Lynn, 1986; Polit & Beck, 2006). All nine participants provided ratings for the candidate prompts using a scale from 1 (*Extremely weak relationship to the Frame*) to 11 (*Extremely strong relationship to the Frame*).

Analysis

Qualitative Analysis. For analysis, we used a consensual qualitative research (CQR) process (Hill et al., 1997), which aligned nicely with the constructivist perspectives (Morrow, 2005) of the research team. Further, this was viewed as a way to enhance the reflexivity of our work as a research team and the trustworthiness of the qualitative analysis itself. On these points, Hill et al. (1997) articulated the following process for teams conducting CQR: "Team members first examine the data independently and then come together to present and discuss their ideas until they reach a single unified version that all team members endorse as the best representation of the data" (p. 523). This analytical process was designed to a) decrease group think, because individuals have first reviewed the data apart from others, b) reduce individual bias, and c) provide a check on the decisions of the coding team by using an auditor who was not a part of the initial coding work or discussions.

Using this CQR process, three of the authors independently coded the corpus of 33 pages of expert feedback on the prompt candidates. The coding phase included the use of a minimal initial codebook, with allowance for the creation of new, inductive codes as needed. The codes utilized in this aspect of the qualitative analysis included the following to capture participant sentiment about prompt candidates:

- Clear statement/clear language
- Unclear statement/unclear language
- Wording suggestion or verbiage change, belongs in a different frame/misplaced
- Overlap/redundancy
- Belongs in more than one frame/covers more than one

- Misleading
- Additional valence codes

Each response was assessed in sentence-based units and could be associated with multiple codes.

After coding the corpus individually, the three authors then worked together to arrive at a consensus about prompt candidate codes. Subsequently, the fourth author audited their work and provided feedback about perceived consistencies and inconsistencies. Using this feedback, the first three authors then derived their final, consensus understanding of expert input about the proposed prompts. At this point, some items were tentatively marked for revision, re-mapping, or removal; however, final decisions were not made until the quantitative analysis was completed.

Quantitative Analysis. In addition to providing qualitative feedback about the prompt candidate items, expert participants also responded to each item using a Likert-type scale. Using these data, I-CVIs were computed using the methodology described by Polit and Beck (2006):

1. Experts individually rated all draft prompts regarding their apparent relevance to an associated construct (here, an ACRL frame) using an 11-point Likert-type scale of relatedness/relevance to the construct.
2. Next, for each prompt across each expert, the original Likert-type scale was bifurcated into "adequate or strong relatedness/relevance" and "weak or no relatedness/relevance" categories; in this study, any potential prompt that was rated at a 7 or above by an expert on the 11-point scale was considered adequately related and relevant to the frame onto which it was initially mapped for that expert.
3. After this, an I-CVI was computed for each item using the proportion of experts who rated an item as at least adequately related to a frame or better. Using Lynn (1986) as a standard, Polit and Beck (2006) noted that when there are more than six raters, items should be considered for removal or re-mapping if their I-CVI is less than 0.78; this is the cutoff score we used in this study, which resulted in 38 prompt candidates being highlighted for review and possible removal or remapping.

Results

Using both qualitative and quantitative analyses, the research team reviewed each of the original 101 prompt candidates. Where both the quantitative and qualitative results aligned, an item was either retained, moved, or removed, depending on the alignment. Where results were mixed, in that the qualitative and quantitative results suggested different actions to be taken on a prompt, the team discussed the merits of each line of data to arrive at a decision on whether to retain a prompt for inclusion in the pilot instrument for Study 2 or not. In all, 14 prompts were kept for use in the Authority Is Constructed and Contextual frame [AUTHORITY], 12 in the Information Creation as a Process frame [INFO CREATION], 13 in the Information Has Value frame [INFO VALUE], 23 in the Research as Inquiry frame [INQUIRY], 14 in the Scholarship as Conversation frame [CONVERSATION], and 18 in the Searching as Strategic Exploration frame [SEARCHING]. In total, seven prompt candidates were removed and the remaining 94 were included in the pilot measure used in Study 2.

Study 2: Student Pilot Testing

Study 2 was our pilot test of the 94-item pilot instrument resulting from Study 1. In this study, we shifted from expert reviews related to content validity to college student reviews to assess the face validity of individual items and item-mapping to information literacy frames from the perspective of the target demographic of the ILRT, which is first-year college students.

Sample

The research team recruited participants using a convenience sampling method. In total, 44 students opted to participate in the study: 12 from one section of a standard college success course (CG 100) and 32 across two sections of the second course in an introductory writing sequence (WR 122) at a community college. The sample included 24 males (55.8%), 18 females (41.9%), 1 with a self-described gender designation, and 1 who did not report a gender category. The mean age in years for the sample was 24.74 (SD 8.18; Range 17–52) with 58% of the students being 22 years of age or younger. The sample included White (37.5%), Hispanic/Latinx (25%), Asian (15%), and Black (10%) participants; 12.5% of the students preferred to self-describe or not report their ethnicity. Two-thirds of the respondents indicated that English was their first language, and 19% of the sample indicated Spanish as their first language. Other first languages reported included Arabic, Dari/Farsi, Korean, Malayan, and Vietnamese. Regarding educational attainment, two-thirds of the students reported that they had 40 or fewer quarter credits, indicating that most students would be categorized as first-year college students; 88.1% of the sample had 70 or fewer quarter credits.

Data Collection

After obtaining informed consent from each participant, the 94-item pilot version of the ILRT was administered during class time. Students rated each item (e.g., "I have strategies for coming up with search terms and keywords.") on a six-point Likert-type scale anchored for self-reflection as follows:

1 = Very untrue of me (0% of the time)

2 = Untrue of me (about 20% of the time)

3 = Somewhat untrue of me (about 40% of the time)

4 = Somewhat true of me (about 60% of the time)

5 = True of me (about 80% of the time)

6 = Very true of me (100% of the time)

Once everyone in the class had completed the self-reflection rating, students were asked to review each item and provide written feedback on any item they thought was confusing or ambiguous.

After administering the tool and obtaining written feedback, the authors conducted group-level discussions to obtain additional feedback from the students; notes were taken as a record of these discussions. Our objective was to obtain additional feedback about participants' understanding of the items and the relative importance of items within each of the six frames represented by the ILRT frames. This was important to gain insight into the face validity of pilot instrument items, supporting decisions about which items to retain and which items to remove in the next iteration of the instrument

Results

The authors then reviewed individual- and group-level feedback related to each of the 94 items on this form of the ILRT. Where no issues had been raised, an item was retained.

Where concerns were raised either through individual written feedback or group-level verbal feedback, an item was either revised and retained, or removed from the item pool.

At this stage, 10 prompts were kept for use in the Authority Is Constructed and Contextual frame [AUTHORITY], 10 in the Information Creation as a Process frame [INFO CREATION], 10 in the Information Has Value frame [INFO VALUE], 15 in the Research as Inquiry frame [INQUIRY], 11 in the Scholarship as Conversation frame [CONVERSATION], and 16 in the Searching as Strategic Exploration frame [SEARCHING]. In total, 22 prompt candidates were removed and the remaining 72 were included in the version measure used in Study 3.

Study 3: Item-Reduction and Construct Validity

In Study 3, our objective was twofold: to reduce the item pool resulting from Study 2 (Appleton et al., 2016; Marsh et al., 2010) and to assess the factor structure of the reduced-item ILRT through an Exploratory Structural Equation Modeling (ESEM) process (Marsh et al., 2014; van Zyl & ten Klooster, 2022).

Sample

The pool of respondents for Study 3 was obtained from a convenience sample of community college students across 16 sections of WR 121: College Composition. In total, 542 students consented to participate in the study. Of the 498 respondents who reported gender, 267 (53.6%) were females, 208 (41.8%) were males, 9 (1.8%) were non-binary, 5 (1.0%) used a self-described gender designation, and 9 (1.8%) conveyed that they did not want to report a gender category. The mean age in years for the sample was 20.56 (SD 5.88; Range 15–50) with 80.6% of the students being 22 years of age or younger. Of those reporting race and ethnicity, the sample was 49.0% White, 24.2% Hispanic/Latinx, 9.5% Asian, 6.8% Black, 3.5% Hawaiian/Pacific Islander, and 2.8% Native American; 4.4% of the respondents specified that they were choosing not to report their race or ethnicity. Of those reporting their first language, 75.1% of the students reported English as their first language and 11.0% noted Spanish as their first language; more than 20 other languages were indicated. Regarding educational attainment, 91.4% of the students had 36 quarter credits or fewer; thus, most would be categorized as college first-year students.

Instrument

Participants completed the 72-item draft version of the ILRT resulting from Study 2 using the Qualtrics web platform. After reading and agreeing to a click-to-consent screen, participants reported demographic information and responded to each of the items. Upon completion, participants were notified that their response had been recorded and were provided contact information should they have follow-up questions or comments related to the study.

Planned Analyses

Item Reduction. For analyses related to item reduction, we utilized Mplus 8.4 (Muthén & Muthén, 2019) and followed procedures advanced by Appleton et al. (2016) and Marsh et al. (2010). Our intent was to reduce the total item count from 72 while also retaining breadth of measurement of each of the ACRL's (2015) IL frames, maintaining satisfactory subscale reliability estimates, and providing an ILRT factor structure with acceptable goodness-of-fit indices. The researchers had no pre-analysis target for the number of items to remove, and preferred the item-reduction to be driven primarily by empirical results using the following criteria:

1. Flag any item that had a subscale confirmatory factor analysis (CFA) factor loading of < .5 within its respective frame.

2. Flag any additional items where the Mplus modification indices were > 10.

3. Team review of all items targeted for removal to assure the breadth of the reduced subscale would adequately represent the associated information literacy frame; where the team determined the breadth would be inappropriately reduced, a flagged item was retained.

4. Assess coefficient α of the resulting (reduced) frame subscales for adequacy (.70 or greater, an often-cited rule of thumb; Taber, 2018).

Model Fit and Measurement Quality. Subsequent to item-reduction efforts, the factor structure of the resulting reduced-item ILRT was modeled in a stepwise manner, moving from restrictive models to decreasingly restrictive models (van Zyl & ten Klooster, 2022) and comparing relative fit across competing models for a six factor (six subscale) ILRT: Independent Cluster Model Confirmatory Factor Analysis (ICM-CFA), Hierarchical Confirmatory Factor Analysis (H-CFA), Bifactor Confirmatory Factor Analysis (B-CFA), Exploratory Structural Equation Modeling (ESEM), Hierarchical Exploratory Structural Equation Modeling (H-ESEM), Bifactor Exploratory Structural Equation Modeling (B-ESEM). Set-ESEM was considered (Marsh et al. 2020); however, because we construe the ILRT frames as conceptually related, set-ESEM modeling was not included in our analysis (Alamer, 2022).

Researchers have differentiated various CFA and ESEM models in the literature. ICM-CFA models constrain all items "to have zero factor loadings on all factors other than the one they are designed to measure" and are seen by some as too restrictive in many cases, failing to attain acceptable fit in many multidimensional measures (Marsh et al., 2014, p. 86). H-CFA and B-CFA models are differentiated from ICM-CFA models in that each has at least one additional factor beyond the *n* specific factors in an ICM-CFA. In an H-CFA, the original ICM-CFA factors are effect indicators of constructs (in this case, aspects of IL) that then themselves become indicators of second- or higher-order constructs (Cucina & Byle, 2017). B-CFA can be differentiated from H-CFA in that each item loads on its specific ICM-CFA factor as well as general factor simultaneously (Cucina & Byle, 2017). Finally, in response to the critique by some that CFA-based models are too restrictive (e.g., Marsh et al., 2014) and that they fail to allow for observed factor correlations (Appleton et al., 2016), ESEM models (including H-ESEM and B-ESEM) have been developed to test models where individual scale items can be modeled to load onto specific factors (like ICM-CFA) while also being allowed to load more minimally onto other factors (unlike ICM-CFA). In the work described below, we utilized each of the model types noted above.

For analysis, we used Mplus 8.4 (Muthén & Muthén, 2019) to test model fit and measurement quality and were assisted in code generation using de Beer and van Zyl's (2019) code generator. Because the response scale used in the ILRT includes six anchor points, we treated the resulting subscales as continuous rather than ordinal (Sass et al., 2014). Given this, we used robust maximum estimation (MLR) rather than maximum likelihood or weighted least squares mean and variance estimation methods (van Zyl & ten Klooster, 2022).

Model fit was evaluated using several goodness of fit indices. Approximate fit indices used were the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR). Incremental fit indices used were the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Akaike Information Criterion (AIC), Bayes Information Criterion (BIC), and the sample-size Adjusted Bayes Information Criterion (aBIC). When evaluating a model for fit, chi-square is known “to be oversensitive” (Appleton et al., 2016, p. 58) and thus was not primary in our evaluation of model fit. For fit comparison, guidelines were taken from van Zyl and ten Klooster (2022, p. 10); see Table 1.

Table 1: Fit Indices and Thresholds

Fit Index	Cut-off Criterion
RMSEA	0.06 - 0.80 (marginally acceptable fit); 0.01-0.05 (excellent fit); Model comparison: Retain model where ΔRMSEA ≤ 0.015
SRMR	0.06 - 0.80 (marginally acceptable); 0.01-0.05 (excellent); Model comparison: Retain model where ΔSRMR ≤ 0.015
CFI	0.90 to 0.95 (marginally acceptable); 0.96 to 0.99 (excellent) Model comparison: Retain Model with Highest CFI (ΔCFI > 0.01)
TLI	0.90 to 0.95 (marginally acceptable); 0.96 to 0.99 (excellent) Model comparison: Retain Model with Highest TLI (ΔTLI > 0.01)
AIC	Lowest value in comparative measurement models
BIC	Lowest value in comparative measurement models
aBIC	Lowest value in comparative measurement models

For models emerging out of relative fit and comparative fit thresholds, measurement quality was assessed in several ways. Results were assessed through inspection of standardized factor loadings ($\lambda > .35$), standard errors, item uniqueness (δ; > .1, < .9), explained common variance (I-ECV; < .7), and levels of reliability (α [> .7]; ω [> .7]; ωH [> .7]) (Morin et al., 2016; van Zyl & ten Klooster, 2022).

Results

Item Reduction. Using the four-step process noted above, all 72-items of the penultimate form of the ILRT were subjected to empirical review. On the basis of the >= 0.5 factor loading threshold (step 1) five items were flagged for team review. Subsequently, the remaining 67 items were reviewed for high factor cross-loadings using the modification indices (> 10) from Mplus output (step 2); an additional 14 items were flagged. Team review then proceeded (step 3) and 4 of those 19 flagged items were deemed essential in accurately capturing the breadth of the *Framework* and thus were not removed. This resulted in a total reduction of 15 items. Finally, in step 4 each of the resulting subscales of this 57-item ILRT form had adequate or better coefficient α results (0.77 - 0.88; see Table 2), suggesting the move toward alternative model testing was appropriate with this reduced pool of items. Item descriptive statistics are found in Appendix A.

Table 2: Coefficient Alpha (α) by Subscale after Item-Reduction

Subscale	α
Authority Is Constructed and Contextual [AUTHORITY]	0.85
Information Creation as a Process [INFO CREATION]	0.82
Information Has Value [INFO VALUE]	0.77
Research as Inquiry [RES INQUIRY]	0.86
Scholarship as Conversation [CONVERSATION]	0.83
Searching as Strategic Exploration [SEARCHING]	0.88

Model Fit and Measurement Quality. Each of the six competing models of the final 57-item form of the ILRT were assessed for relative fit, with fit indices available for review in Appendix B. In evaluating these results, only the three ESEM models had goodness-of-fit that met at least marginal thresholds for RMSEA, SRMR, CFI, and TLI. Though the bifactor ESEM model showed slightly better fit than either the ESEM model or the hierarchical ESEM model, change thresholds (ΔRMSEA ≤ 0.015; ΔSRMR ≤ 0.015; ΔCFI > 0.01; ΔTLI > 0.01) were not surpassed in any combination of comparisons among the three models. This noted, inspection of the factor loadings of the ESEM model indicated that 24 of 57 items failed to load on their assigned factor at > .35. Further, factor loadings of the H-ESEM model were similarly poor, with 25 of 57 items loading on their assigned factor at < .35. Thus, observing this as well as multiple moderate to strong cross-loadings in both models, they were rejected for further exploration.

In such situations, van Zyl and ten Klooster (2022) provide important guidance that: "should there be multiple medium to large cross-loadings in the ESEM model, it could indicate support for the presence of a larger global factor, and therefore the bifactor ESEM model should be explored"; if bifactor models are explored, "there should be a well-defined G-Factor (where all items load significantly on such), and *reasonably* well-defined S-Factors (cross- and non-significant loadings are permitted)" (p. 10). In inspecting the B-ESEM model, all but one of the 57 items loads at > .35 on the general factor, and all residuals were within a priori acceptability ranges (see Appendix C). To examine the degree of unidimensionality in bifactor models, Rodriguez et al. (2016) promote an examination of explained common variance (ECV) which "indexes variance specific to a general factor by taking the ratio of variance explained by a general factor and dividing it by the variance explained by a general and group factors where factors are assumed to be uncorrelated" (p.

144). In the present bifactor model, ECV is .803 suggesting a strong general factor and relatively weaker specific factors.

Both general and specific factor reliability estimates were adequate, with α and ω exceeding 0.7 in all instances (see Table 3). Further, ωH was 0.93, indicating that the majority of the variance in total scores can be attributed to an individual's differences on a general factor (Rodriguez et al., 2016), rather than specific factors. As such, almost all of the reliable variance in total scores can be attributed to the general factor, which is assumed to reflect individual differences regarding their self-reflection of "general information literacy" in this sample. Thus, and critically, total scores might best be interpreted as an essentially unidimensional reflection of information literacy.

Table 3: Reliability Estimates (Bifactor ESEM Model)

Factor	α	ω
General	0.96	0.96
Authority Is Constructed and Contextual [AUTHORITY]	0.85	0.84
Information Creation as a Process [INFO CREATION]	0.82	0.81
Information Has Value [INFO VALUE]	0.77	0.76
Research as Inquiry [INQUIRY]	0.86	0.84
Scholarship as Conversation [CONVERSATION]	0.83	0.81
Searching as Strategic Exploration [SEARCHING]	0.89	0.88

Finally, specific factor correlations in the bifactor ESEM model ranged from very small to moderate in size (see Table 4).

Table 4: Reliability Estimates (Bifactor ESEM Model)

	Authority	Info Creation	Info Value	Inquiry	Conversation
Info Creation	0.10				
Info Value	0.24	0.09			
Inquiry	0.12	0.06	0.06		
Conversation	0.03	0.06	0.02	0.13	
Searching	0.06	0.06	0.11	0.32	0.05

The results of Study 3 represent an initial step in understanding the factor structure of the ILRT, resulting in some models being deemed inappropriate (e.g., ICM-CFA, H-ESEM) that were originally considered theoretically viable by the research team. Further, there appears to be modest evidence for the viability of a bifactor understanding of the ILRT based upon these results. See Supplemental 1 for the final 57-item set, which is available at https://osf.io/jvcwd.

Discussion

The goal of this research was to develop and validate the Information Literacy Reflection Tool (ILRT), which incorporates metacognitive awareness by using the constructs established by Schraw and Dennison (1994). To capture the essential concepts described in the *Framework,* the initial item set was written to align with the dispositions and knowledge practices within each of the six frames (see Supplemental 2, which is available at https://osf.io/jvcwd). Based on Study 1, Study 2, and the item reduction and modeling-testing work in Study 3, we have initial evidence that the 57 ILRT items cover essential aspects of the *Framework* as a whole. Item reduction was successful in that the item count was reduced from 72 to 57 while maintaining concept coverage, reducing time-to-complete, and making the tool simpler for administration in a college classroom setting. Taking the ILRT is an opportunity for an individual to reflect on their metacognitive awareness of IL dispositions and knowledge practices and have exposure to the essential concepts across the six IL frames.

While the results of our work at this point provide some initial support for the ILRT, we have only provided evidence of the content (Study 1), face (Study 2), and construct (Study 3) validity of the instrument. In Mahmood's (2017) systematic review of self-efficacy-focused informational literacy tools, these considerations are important. However, we note that more work remains to be done to fully establish the ILRT as a trustworthy instrument, namely efforts to better assess its test-retest reliability, concurrent, predictive, convergent, and discriminant validity.

Implications for Use

That the bifactor model emerged from Study 3 as best-fitting overall has implications for use. Practitioners should be aware that the general factor (i.e., overall reflection on one's own information literacy) is the most salient result obtained from the ILRT while the specific factors (e.g., Information Creation as a Process, Scholarship as Conversation) represent minor residual detail not completely captured by the general factor (Markon, 2019). As such, calculating a total score might best be interpreted as a unidimensional interpretation of information literacy awareness. Put another way, these results suggest that the ILRT measures metacognitive awareness of IL as a primary construct, and further research is needed to establish the usefulness of each of the threshold concept subscales incorporated in the ILRT.

An additional implication emerging from these studies is the student audience for which the ILRT is most appropriately used at this time. Our work here was with a sample of early undergraduate-level writers. The intent of the instrument is for use with first-year college students across academic disciplines and in everyday student research contexts; the instrument has not been studied with other student groups.

Finally, when using the ILRT with appropriate populations, we contend that while each section is a subscale that aligns with one frame of the *Framework,* these do not stand alone, even if pedagogically relevant for classroom discussion. Further, the individual items from which both the general and specific factors are comprised on their own have no demonstrated reliability or validity; however, when used together to assess overall information literacy, the ILRT offers unique value in helping students learn about their own information literacy. Emerging self-perception scales, such as the Informed Learning Scale (Flierl et al., 2021) that are based on the *Framework*, are intended to measure and communicate IL self-efficacy and achievement. As an alternative approach, the ILRT establishes a metacognitive awareness self-report measure that is fully representative of IL as described by the *Framework*. The value of the ILRT for student learning will ultimately be determined by how the instrument is used in teaching contexts and how students transfer IL self-reflection practices outside of the classroom. These aspects of the potential impact of the ILRT on student learning need further research.

Limitations and Future Directions

Without replication of the results of Study 3, the present preference for the bifactor structure of the ILRT remains tentative; thus, similar studies should be conducted on similar populations with the final 57-item version of the ILRT that emerged in our work in these studies. Such replications would enhance confidence in the construct validity of the instrument. As noted above, other considerations of validity, such as convergent and discriminant validity, should be pursued in future work. Additionally, we have provided no evidence at this point regarding sensitivity to change or measurement invariance across groups. Thus, further study is required to explore the ILRT and the assessment of perceived developmental progression in information literacy practices as well as measurement invariance (Putnick & Bornstein, 2016) within emerging college-writers (e.g., gender identifications) and across other populations (high school students, upper-division undergraduates beginning discipline-specific research, graduate students, public library settings, etc.).

Though supplements, such as the *Companion Document to the ACRL Framework for Information Literacy for Higher Education: Women's and Gender Studies*, have been approved by the ACRL Board of Directors, the *Framework* itself is static while the environment has changed radically since 2015 when it was published, and we began this study. Further research of the ILRT is needed to incorporate advances in antiracist/anti-oppressive pedagogy, including a survey of critical reviews of the *Framework*, and could identify individual ILRT items that are potential candidates for addition, revision, or elimination. Critical praxis continues to advance, propelled by events such as the pandemic, civil unrest, the rise of dis/mis/mal-information, and social justice movements. We are embarking on further research to identify conceptual gaps in the ILRT that if addressed would better represent the assets and experiences that students carry with them into learning contexts.

Critical review of the ILRT is needed to identify language and assumptions indicative of deficit model approaches to education and inform use of the ILRT in teaching contexts. Deficit models are problematic because they embed prejudice by framing achievement level as a result of qualities personal to the individual, like grit (Tewell, 2020), or as particular to a demographic, such as first-generation college students (Ilett, 2019). Tewell (2020) has argued that "grit and other deficit models are fundamentally about how best to maintain the functioning of our existing systems, without requiring significant changes or sacrifices on the part of privileged classes" (p. 147). Deficit framing allows educators to ignore the impact of privilege, systemic inequality and racism, and the accompanying discrepancies in access to resources. Consider this example from the ILRT, where the item "I value the skills, time and effort needed to produce knowledge" does not mention access to other important resources, like money.

In Study 2 we gathered and used student feedback to clarify language and refine the ILRT items but recommend even more direct work with undergraduates to ground the ILRT in student languaging and lived experiences. We are interested in learning more about how the ILRT may be used within asset-based information literacy instruction to position students as experts on their own information use and learning (Morrison, 2017). In the context of asset-based education, it is worth noting yet again that the ILRT is not intended to generate a grade or to imply a proficiency or deficit in IL. The ILRT uses a time-based frequency scale so individuals can evaluate how "true of me" each item reflects their current IL practice. As a formative assessment a score generated by the ILRT can represent something pedagogically valuable while not providing an assessment of the respondent's worth against some proficiency standard. The value of engaging with the ILRT is in conceptual exposure and building an awareness of one's own IL practices and how they change over time and context.

Self-regulated learning (SRL) events involve a complex interplay of metacognition, affect, motivation, and an individual's self-concept (a broad construct that includes, among other things, self-efficacy) (Efklides et al., 2017). In their work developing an IL self-efficacy scale, Kurbanoglu et al. (2006) highlighted the crucial role of self-efficacy as it relates to motivation and lifelong learning. While some ILRT items use language related to confidence (e.g., "I am good at keeping information organized while I research"), the relation to a student's self-efficacy is indirect and the ILRT is not intended to measure self-efficacy. That said, engaging with the ILRT is a metacognitive experience and as such, it involves affect and conveys information related to cognition that has implications for the self-concept (Efklides et al., 2017; Yan & Oyserman, 2018), which therefore has the potential to impact self-efficacy. As Efklides et al. (2017) noted, "Metacognitive experiences trigger control decisions, update self-concept, and offer the ground for [self] appraisals that give rise to achievement emotions" (p. 73). Connecting the ILRT to motivation, affect, and the self-efficacy necessary for life-long learning is a broad, rich area for exploration. Further testing is needed to determine how the ILRT as a metacognitive scale might be used in conjunction with SRL to surface and connect elements such as student reflective practices, confidence, awareness of the affective and other learning domains, motivation, and the development of personal agency.

Finally, we are particularly interested in testing for critical moments for using the ILRT in teaching contexts (Craig et al., 2020; Rosman et al., 2015). Studying the ideal context for metacognitive reflection and formative assessment approaches during IL instruction is an established need (Craig et al., 2020; Hostetler et al., 2018), and the ILRT is a tool that can help test that line of inquiry. The ILRT is available in various formats under Creative Commons licensing at https://ilreflection.org/.

References

Alamer, A. (2022). Exploratory structural equation modeling (ESEM) and bifactor ESEM for construct validation purposes: Guidelines and applied example. *Research Methods in Applied Linguistics, 1*(1), Article 100005. https://doi.org/10.1016/j.rmal.2022.100005

Andrade, H. (2019). A critical review of research on student self-assessment. *Frontiers in Education, 4*, Article 87. https://doi.org/10.3389/feduc.2019.00087

Appleton, P. R., Ntoumanis, N., Quested, E., Viladrich, C., & Duda, J. L. (2016). Initial validation of the coach-created Empowering and Disempowering Motivational Climate Questionnaire (EDMCQ-C). *Psychology of Sport and Exercise, 22*, 53–65. https://doi.org/10.1016/j.psychsport.2015.05.008

Association of College and Research Libraries. (2000). *The information literacy competency standards for higher education.* https://alair.ala.org/handle/11213/7668

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

Billing, D. (2007). Teaching for transfer of core/key skills in higher education: Cognitive skills. *Higher Education, 53*(4) 483–516. https://doi.org/10.1007/s10734-005-5628-5

Catalano, A. (2017). Development and validation of the Metacognitive Strategies for Library Research Skills Scale (MS-LRSS). *The Journal of Academic Librarianship, 43*(3), 178–183. https://doi.org/10.1016/j.acalib.2017.02.017

Craig, K., Hale, D., Grainger, C., & Stewart, M. E. (2020). Evaluating metacognitive self-reports: systematic reviews of the value of self-report in metacognitive research. *Metacognition and Learning, 15*(2), 155–213. https://doi.org/10.1007/s11409-020-09222-y

Cucina, J., & Byle, K. (2017). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. *Journal of Intelligence, 5*(3), Article 27. https://doi.org/10.3390/jintelligence5030027

de Beer L. T., & van Zyl, L. E. (2019). *ESEM code generator for Mplus.* https://www.surveyhost.co.za/esem/

Doyle, M., Foster, B., & Yukhymenko-Lescroart, M. A. (2019). Initial development of the Perception of Information Literacy Scale (PILS). *Communications in Information Literacy, 13*(2), 205–227. https://doi.org/10.15760/comminfolit.2019.13.2.5

Efklides, A., Schwartz, B. L., & Brown, V. (2017). Motivation and affect in self-regulated learning: Does metacognition play a role? In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 64–82). Routledge.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906–911 https://doi.org/10.1037/0003-066X.34.10.906.

Flierl, M., Maybee, C., & Bonem, E. (2021). Developing the Informed Learning Scale: Measuring information literacy in higher education. *College & Research Libraries, 82*(7), 1004–1016. https://doi.org/10.5860/crl.82.7.1004

Fulkerson, D. M., Ariew, S. A., & Jacobson, T. E. (2017). Revisiting metacognition and metaliteracy in the ACRL *Framework*. *Communications in Information Literacy, 11*(1), 21–41. https://doi.org/10.15760/comminfolit.2017.11.1.45

Hill, C. E., Thompson, B. J., & Williams, E. N. (1997). A guide to conducting consensual qualitative research. *The Counseling Psychologist, 25*(4), 517–572. https://doi.org/10.1177/0011000097254001

Hostetler, K., Luo, T., & Stefaniak, J. E. (2018). Aligning information literacy assessment with metacognitive strategies. *Journal of University Teaching and Learning Practice, 15*(5), Article 2. https://doi.org/10.53761/1.15.5.2

Ilett, D. (2019). A critical review of LIS literature on first-generation students. *portal: Libraries and the Academy, 19*(1), 177–196. https://doi.org/10.1353/pla.2019.0009

Jacobson, T. E., Mackey, T. P., & Olivier, J. (2021). Aligning metaliteracy with self-directed learning to expand assessment opportunities. In E. Mentz & A. Lubbe (Eds.), *Learning through assessment: An approach towards self-directed learning* (pp. 71–97). AOSIS.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Kuglitsch, R. Z. (2015). Teaching for transfer: Reconciling the *Framework* with disciplinary information literacy. *portal: Libraries and the Academy 15*(3), 457–470. https://doi.org/10.1353/pla.2015.0040.

Kurbanoglu, S. 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

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382–385. https://doi.org/10.1097/00006199-198611000-00017

Mackey, T. P., & Jacobson, T. E. (2011). Reframing information literacy as a metaliteracy. *College & Research Libraries*, *72*(1), 62–78. https://doi.org/10.5860/crl-76r1

Mackey, T. P., & Jacobson, T. E. (2014). *Metaliteracy: Reinventing information literacy to empower learners.* ALA Neal-Schuman.

Mahmood, K. (2017). Reliability and validity of self-efficacy scales assessing students' information literacy skills: A systematic review. *The Electronic Library, 35*(5), 1035–1051. https://doi.org/10.1108/EL-03-2016-0056

Markon, K. E. (2019). Bifactor and hierarchical models: Specification, inference, and interpretation. *Annual Review of Clinical Psychology, 15*, 51–69. https://doi.org/10.1146/annurev-clinpsy-050718-095522

Marsh, H. W., Guo, J., Dicke, T., Parker, P. D., & Craven, R. G. (2020). Confirmatory factor analysis (CFA), exploratory structural equation modeling (ESEM), and set-ESEM: Optimal balance between goodness of fit and parsimony. *Multivariate Behavioral Research, 55*(1), 102–119. https://doi.org/10.1080/00273171.2019.1602503

Marsh, H. W., Martin, A. J., and Jackson, S. (2010). Introducing the short version of the Physical Self-Description Questionnaire: New strategies, short-form evaluative criteria, and applications of factor analyses. *Journal of Sport & Exercise Psychology, 32*(4), 438–482. https://doi.org/10.1123/jsep.32.4.438

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, 85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94*(2), 249–259. https://doi.org/10.1037//0022-0663.94.2.249

Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 116–139. https://doi.org/10.1080/10705511.2014.961800

Morrison, K. L. (2017). Informed asset-based pedagogy: Coming correct, counter-stories from an information literacy classroom. *Library Trends*, 66(2), 176–218. https://doi.org/10.1353/lib.2017.0034

Morrow, S. L. (2005). Quality and trustworthiness in qualitative research in counseling psychology. *Journal of Counseling Psychology, 52*(2), 250–60. https://doi.org/10.1037/0022-0167.52.2.250

Muthén, L .K. & Muthén, B. O. (2019). *Mplus* (Version 8.4) [Computer software]. StatModel. https://www.statmodel.com/

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice*, *41*(4), 219–225. https://doi.org/10.1207/s15430421tip4104_3

Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489–497. https://doi.org/10.1002/nur.20147

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137–150. https://doi.org/10.1037/met0000045

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

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(2), 167–180. https://doi.org/10.1080/10705511.2014.882658

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033

Schunk, D. H., & Greene, J. A. (2017). Historical, contemporary, and future perspectives on self-regulated learning and performance. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 1–15). Routledge.

Speth, C. A., Namuth, D. M., & Lee, D. J. (2007). Using the ASSIST short form for evaluating an information technology application: Validity and reliability issues. *Informing Science: The International Journal of an Emerging Transdiscipline, 10*, 107–119. https://doi.org/10.28945/459

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tewell, E. (2020). The problem with grit: Dismantling deficit thinking in library instruction. *portal: Libraries and the Academy*, 20(1), 137–159. https://doi.org/10.1353/pla.2020.0007

van Zyl, L. E., & ten Klooster, P. M. (2022). Exploratory structural equation modeling: Practical guidelines and tutorial with a convenient online tool for Mplus. *Frontiers in Psychiatry, 12*, Article 795672. https://doi.org/10.3389/fpsyt.2021.795672

Yan, V. X., & Oyserman, D. (2018). The world as we see it: The culture–identity–metacognition interface. In J. Proust, & M. Fortier (Eds.), *Metacognitive Diversity: An Interdisciplinary Approach* (pp. 225–244). Oxford University Press. https://doi.org/10.1093/oso/9780198789710.003.0011

Appendix A: Item Descriptive Statistics

Factor	Item	Mean	Skewness	Kurtosis	SD
1 - Authority	AUTH1	4.46	-0.57	-0.05	1.12
	AUTH2	4.61	-0.53	0.21	0.99
	AUTH3	4.53	-0.56	0.39	1.01
	AUTH4	4.34	-0.27	-0.19	1.03
	AUTH5	4.76	-0.74	0.71	1.02
	AUTH6	4.52	-0.84	0.95	1.03
	AUTH7	4.26	-0.45	-0.19	1.15
	AUTH8	4.48	-0.55	0.34	1.02
	AUTH9	4.36	-0.62	0.19	1.16
2 - Info Creation	INFOCR1	4.60	-0.37	0.02	0.95
	INFOCR2	4.66	-0.84	0.66	1.07
	INFOCR3	4.48	-0.46	0.13	1.00
	INFOCR4	4.32	-0.43	-0.08	1.09
	INFOCR5	4.48	-0.46	-0.02	1.01
	INFOCR6	3.82	-0.10	-0.67	1.32
	INFOCR7	4.79	-0.85	0.73	1.03
	INFOCR8	4.20	-0.37	-0.26	1.15
3 - Info Value	INFOVAL1	4.88	-0.74	0.68	0.91
	INFOVAL2	5.07	-1.00	1.40	0.89
	INFOVAL3	4.94	-0.92	0.98	0.97
	INFOVAL4	4.65	-0.69	0.17	1.08
	INFOVAL5	4.29	-0.58	-0.12	1.25
	INFOVAL6	4.85	-1.14	0.86	1.26
	INFOVAL7	4.71	-0.84	0.18	1.18
	INFOVAL8	4.40	-0.61	-0.14	1.23
4 - Inquiry	RESINQ1	4.07	-0.34	-0.37	1.18
	RESINQ2	4.21	-0.25	-0.07	1.02
	RESINQ3	4.14	-0.29	-0.11	1.09
	RESINQ4	4.06	-0.28	0.01	1.10
	RESINQ5	4.94	-0.75	0.43	0.95
	RESINQ6	4.48	-0.25	-0.31	0.98
	RESINQ7	4.43	-0.63	-0.14	1.21
	RESINQ8	4.66	-0.70	0.53	1.01
	RESINQ9	4.15	-0.25	0.06	1.05
	RESINQ10	4.51	-0.42	0.17	0.95
	RESINQ11	4.79	-0.47	0.30	0.87

Factor	Item	Mean	Skewness	Kurtosis	SD
5 - Conversation	SCHOL1	4.59	-0.46	0.17	0.97
	SCHOL2	4.27	-0.48	0.02	1.15
	SCHOL3	3.81	-0.27	-0.49	1.27
	SCHOL4	4.09	-0.27	-0.42	1.15
	SCHOL5	4.37	-0.54	0.03	1.12
	SCHOL6	4.38	-0.52	0.00	1.14
	SCHOL7	4.04	-0.37	-0.23	1.19
	SCHOL8	4.76	-0.69	0.41	1.02
6 - Searching	SEARCH1	4.63	-0.72	0.61	1.03
	SEARCH2	4.58	-0.55	-0.23	1.13
	SEARCH3	4.16	-0.30	-0.60	1.24
	SEARCH4	4.79	-0.49	0.15	0.85
	SEARCH5	4.67	-0.78	0.38	1.08
	SEARCH6	4.41	-0.58	0.26	1.03
	SEARCH7	4.48	-0.73	0.72	1.04
	SEARCH8	4.38	-0.50	0.21	1.07
	SEARCH9	4.74	-0.58	0.51	0.93
	SEARCH10	4.33	-0.56	-0.31	1.27
	SEARCH11	4.78	-0.69	0.35	0.97
	SEARCH12	4.67	-0.65	0.49	0.99
	SEARCH13	4.33	-0.23	-0.25	1.04

Appendix B: Goodness-of-Fit Indices

	χ2 (all *)	df	RMSEA	RMSEA [90% CI]	SRMR	CFI	TLI	AIC	BIC	aBIC	Meets Criteria
Model 1 – ICM-CFA	2355.960	1524	0.035	[0.032 – 0.038]	0.045	0.892	0.897	66325.058	66494.924	66462.829	No
Model 2 – H-CFA	2413.757	1533	0.036	[0.034 – 0.040]	0.047	0.886	0.881	66378.293	66539.939	66505.512	No
Model 3 – B-CFA	2362.966	1482	0.037	[0.034 – 0.040]	0.045	0.886	0.877	66382.475	67314.259	66590.697	No
Model 4 – ESEM	1902.706	1269	0.031	[0.028 – 0.034]	0.031	0.927	0.909	75101.292	76973.825	75574.009	Marginal
Model 5 – B-ESEM	1782.540	1218	0.030	[0.027 – 0.033]	0.028	0.935	0.915	75029.966	77119.051	75557.051	Marginal
Model 6- H-ESEM	2102.702	1278	0.035	[0.033 – 0.038]	0.030	0.926	0.907	75084.704	76916.022	75547.774	Marginal

Appendix C: Bifactor-ESEM Model Item-level Measurement Indicators

		G factor		Authority		InfoCreation		InfoValue		Inquiry		Conversation		Search		
Factor	Item	λ	S.E.	λ	S.E.	λ	S.E.	λ	S.E.	λ	S.E.	λ	S.E.	λ	S.E.	δ
1 - Authority	AUTH1	**0.66**	0.04	-0.02	0.17	-0.06	0.21	0.02	0.10	-0.06	0.11	-0.22	0.08	-0.10	0.08	0.50
	AUTH2	**0.54**	0.04	**0.38**	0.12	-0.06	0.07	0.06	0.08	0.06	0.07	0.02	0.09	-0.01	0.08	0.55
	AUTH3	**0.49**	0.05	**0.31**	0.09	0.06	0.07	0.05	0.07	0.03	0.08	-0.01	0.10	0.01	0.08	0.65
	AUTH4	**0.59**	0.04	**0.46**	0.09	-0.06	0.13	-0.01	0.07	0.00	0.10	0.05	0.11	-0.05	0.08	0.43
	AUTH5	**0.58**	0.05	**0.41**	0.10	0.17	0.09	0.06	0.08	-0.11	0.08	-0.13	0.14	0.02	0.08	0.44
	AUTH6	**0.65**	0.04	0.10	0.14	-0.13	0.08	0.16	0.09	0.04	0.10	0.00	0.09	-0.01	0.10	0.53
	AUTH7	**0.63**	0.04	-0.02	0.22	-0.13	0.21	0.02	0.08	-0.10	0.09	-0.16	0.10	-0.26	0.09	0.48
	AUTH8	**0.53**	0.05	0.04	0.21	-0.14	0.14	0.16	0.09	0.12	0.13	-0.12	0.10	0.05	0.11	0.64
	AUTH9	**0.58**	0.04	0.17	0.13	-0.13	0.07	0.22	0.07	0.03	0.09	-0.01	0.09	-0.08	0.09	0.56
2 - Info Creation	INFOCR1	**0.61**	0.04	-0.02	0.08	0.13	0.09	-0.06	0.08	-0.06	0.10	0.01	0.08	-0.02	0.08	0.61
	INFOCR2	**0.54**	0.04	0.18	0.15	-0.05	0.16	-0.05	0.08	-0.05	0.09	-0.19	0.10	-0.15	0.09	0.61
	INFOCR3	**0.58**	0.04	-0.03	0.10	-0.05	0.12	-0.10	0.09	-0.03	0.14	0.08	0.11	-0.07	0.11	0.64
	INFOCR4	**0.58**	0.04	0.04	0.09	**0.17**	0.07	-0.09	0.07	-0.11	0.09	0.17	0.09	-0.04	0.07	0.59
	INFOCR5	**0.67**	0.03	-0.16	0.07	0.05	0.15	-0.04	0.07	-0.19	0.10	-0.05	0.10	0.07	0.08	0.48
	INFOCR6	**0.55**	0.04	-0.10	0.08	**-0.11**	0.15	-0.12	0.07	-0.15	0.09	0.00	0.13	-0.18	0.08	0.61
	INFOCR7	**0.56**	0.05	0.07	0.10	**0.35**	0.07	0.06	0.10	-0.01	0.10	0.14	0.17	-0.03	0.09	0.53
	INFOCR8	**0.59**	0.04	-0.09	0.10	-0.02	0.08	-0.05	0.09	-0.08	0.12	0.17	0.12	-0.12	0.10	0.59
3 - Info Value	INFOVAL1	**0.59**	0.04	-0.05	0.09	0.20	0.09	**0.27**	0.11	-0.10	0.10	-0.05	0.13	0.06	0.07	0.53
	INFOVAL2	**0.48**	0.05	0.02	0.09	0.20	0.08	**0.31**	0.12	0.03	0.10	-0.15	0.12	-0.03	0.08	0.61
	INFOVAL3	**0.49**	0.05	0.16	0.08	0.19	0.10	**0.34**	0.10	0.12	0.08	-0.03	0.14	-0.03	0.07	0.57
	INFOVAL4	**0.51**	0.05	-0.06	0.07	0.14	0.09	**0.21**	0.10	-0.02	0.08	-0.14	0.09	0.09	0.07	0.65
	INFOVAL5	**0.56**	0.04	0.05	0.11	-0.05	0.11	0.11	0.11	-0.10	0.13	0.18	0.10	-0.10	0.08	0.62
	INFOVAL6	**0.28**	0.05	0.13	0.09	0.15	0.07	**0.15**	0.09	0.03	0.10	-0.05	0.11	0.08	0.07	0.85
	INFOVAL7	**0.41**	0.04	0.05	0.11	-0.15	0.13	**0.47**	0.08	0.05	0.08	0.06	0.09	-0.01	0.07	0.58
	INFOVAL8	**0.44**	0.05	0.06	0.10	-0.31	0.11	**0.51**	0.17	-0.11	0.11	0.09	0.14	0.02	0.07	0.43
4 - Inquiry	RESINQ1	**0.56**	0.04	-0.01	0.07	-0.16	0.05	-0.02	0.08	**0.28**	0.13	0.00	0.10	0.05	0.08	0.58
	RESINQ2	**0.57**	0.03	-0.01	0.07	0.01	0.08	-0.11	0.07	**0.32**	0.12	0.03	0.08	0.18	0.08	0.52
	RESINQ3	**0.56**	0.05	-0.06	0.10	-0.17	0.08	-0.05	0.10	0.14	0.14	0.15	0.12	0.02	0.09	0.61
	RESINQ4	**0.56**	0.04	-0.06	0.08	-0.09	0.08	-0.12	0.09	0.17	0.13	0.06	0.11	0.06	0.08	0.63

		G factor		Authority		InfoCreation		InfoValue		Inquiry		Conversation		Search		
Factor	**Item**	**λ**	**S.E.**	**λ**	**S.E.**	**λ**	**S.E.**	**λ**	**S.E.**	**λ**	**S.E.**	**λ**	**S.E.**	**λ**	**S.E.**	**δ**
4 – Inquiry (cont.)	RESINQ5	0.45	0.05	0.07	0.08	0.18	0.09	0.13	0.09	**0.29**	0.10	-0.09	0.13	0.09	0.08	0.65
	RESINQ6	**0.52**	0.04	0.09	0.14	0.00	0.11	-0.01	0.09	0.19	0.12	0.11	0.09	0.18	0.08	0.64
	RESINQ7	**0.47**	0.04	0.00	0.08	-0.10	0.06	0.04	0.07	0.17	0.10	-0.03	0.08	0.02	0.07	0.74
	RESINQ8	**0.52**	0.04	0.11	0.10	-0.01	0.13	0.04	0.08	0.19	0.11	0.12	0.08	0.11	0.08	0.65
	RESINQ9	**0.60**	0.04	0.04	0.09	-0.10	0.10	-0.09	0.09	0.10	0.15	0.10	0.12	0.17	0.09	0.57
	RESINQ10	**0.54**	0.05	-0.04	0.09	0.05	0.10	0.03	0.10	**0.46**	0.18	-0.08	0.10	-0.01	0.09	0.49
	RESINQ11	**0.60**	0.04	-0.09	0.08	0.17	0.08	-0.03	0.09	**0.20**	0.09	-0.03	0.10	0.18	0.07	0.53
5 - Conversation	SCHOL1	**0.54**	0.03	-0.04	0.07	0.05	0.07	-0.05	0.06	0.13	0.07	0.15	0.09	-0.05	0.06	0.66
	SCHOL2	**0.61**	0.04	-0.05	0.07	-0.03	0.08	-0.03	0.06	-0.01	0.07	**0.30**	0.08	0.02	0.06	0.53
	SCHOL3	**0.62**	0.05	-0.02	0.08	-0.25	0.06	-0.14	0.09	-0.02	0.11	0.25	0.17	-0.07	0.07	0.47
	SCHOL4	**0.55**	0.04	-0.03	0.08	-0.11	0.10	-0.03	0.08	-0.08	0.09	**0.32**	0.10	0.04	0.06	0.57
	SCHOL5	**0.52**	0.04	-0.03	0.11	-0.06	0.07	0.03	0.08	0.06	0.09	0.16	0.08	0.03	0.08	0.70
	SCHOL6	**0.54**	0.04	-0.05	0.09	0.09	0.15	0.06	0.06	0.04	0.08	**0.43**	0.10	0.01	0.07	0.51
	SCHOL7	**0.53**	0.05	-0.08	0.08	-0.09	0.09	-0.02	0.06	-0.01	0.07	**0.32**	0.11	-0.07	0.06	0.60
	SCHOL8	**0.43**	0.06	0.24	0.12	0.40	0.13	0.05	0.09	0.03	0.11	0.25	0.24	0.03	0.09	0.53
6 - Exploration	SEARCH1	**0.60**	0.04	0.00	0.10	0.17	0.10	0.05	0.11	0.03	0.13	-0.06	0.10	**0.19**	0.10	0.57
	SEARCH2	**0.50**	0.04	0.10	0.08	-0.01	0.08	0.03	0.07	0.04	0.09	0.09	0.08	**0.22**	0.07	0.68
	SEARCH3	**0.56**	0.04	0.05	0.06	-0.12	0.08	-0.13	0.07	-0.03	0.09	0.00	0.10	**0.30**	0.07	0.57
	SEARCH4	**0.57**	0.04	-0.13	0.07	0.13	0.06	0.10	0.07	-0.02	0.07	-0.08	0.08	**0.47**	0.07	0.40
	SEARCH5	**0.42**	0.04	-0.01	0.07	0.04	0.07	0.09	0.08	0.10	0.10	-0.03	0.07	**0.23**	0.08	0.75
	SEARCH6	**0.52**	0.04	0.00	0.08	-0.03	0.08	-0.01	0.07	0.07	0.10	0.07	0.07	**0.34**	0.08	0.61
	SEARCH7	**0.51**	0.04	0.04	0.07	-0.09	0.06	-0.05	0.08	0.07	0.10	-0.05	0.08	**0.39**	0.08	0.58
	SEARCH8	**0.63**	0.03	0.02	0.07	-0.06	0.07	-0.01	0.08	0.14	0.11	-0.05	0.08	**0.37**	0.09	0.44
	SEARCH9	**0.59**	0.03	0.01	0.06	0.06	0.07	-0.08	0.05	0.12	0.08	0.10	0.07	**0.29**	0.06	0.53
	SEARCH10	**0.45**	0.04	0.04	0.08	-0.08	0.08	0.01	0.07	0.03	0.09	0.07	0.09	**0.20**	0.07	0.74
	SEARCH11	**0.58**	0.04	-0.06	0.07	0.07	0.07	0.01	0.07	0.12	0.09	-0.11	0.07	**0.27**	0.06	0.56
	SEARCH12	**0.57**	0.04	-0.12	0.11	-0.07	0.08	0.09	0.07	0.08	0.10	0.02	0.08	**0.23**	0.09	0.59
	SEARCH13	**0.52**	0.04	0.08	0.06	-0.02	0.07	0.01	0.07	0.01	0.07	0.08	0.07	**0.30**	0.07	0.63

Note: **Bold items**, significant target loadings ($p < 0.05$); Underlined items indicate significant cross-loadings; λ, standardized factor loadings; S.E., standard error; δ, item uniqueness