



Strengthening Librarian Professional Development: An Assessment of Collection Development Training Evaluation Using Item Analysis

Willian S.A. Frias^a and Jonas T. Sahagun^b

^a De La Salle University, Philippines

^b University of Santo Tomas, Philippines

This work is licensed under Creative Commons Attribution 4.0 International. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>. Authors retain copyright for their article content, while the Association of Special Libraries of the Philippines and Information Generation & Management holds copyright for the publication's design and layout.



Received 17 March 2026; Received in revised form 14 April 2026; Accepted 13 May 2026

Published online

Frias, W.S.A., & Sahagun, J.T. (2026). Strengthening librarian professional development: An assessment of collection development training evaluation using item analysis. *Information Generation & Management*, 5(1), 55-76. <https://dx.doi.org/10.5281/zenodo.20905870>.

Strengthening Librarian Professional Development: An Assessment of Collection Development Training Evaluation Using Item Analysis

Willian S.A. Frias^a and Jonas T. Sahagun^b

^a De La Salle University, Philippines

^b University of Santo Tomas, Philippines

Corresponding author: willian.frias@dlsu.edu.ph

Abstract

Statement of the Problem / Objectives. As the Network of Academic Law Librarians, Inc. (NALL) prepares for subsequent iterations of the Academic Law Librarians Certificate Program (ALLCP), ensuring the pedagogical integrity of its assessment tools is paramount. This study aims to evaluate the diagnostic quality and reliability of the ALLCP post-training quiz on Collection Development. The objective is to identify specific test items requiring revision, specifically those with poor discrimination or inadequate cognitive demand, to ensure the instrument accurately measures professional competency.

Methods. The study utilized a quantitative descriptive design, applying Classical Test Theory (CTT) to evaluate a 10-item post-training assessment. The setting was a national training certificate program for Filipino law librarians, with a subject pool of 52 participants. Data analysis involved calculating the Difficulty Index (p), Discrimination Index (D), item variance, and internal consistency using Cronbach's Alpha. Participants were categorized into top (27%), middle (46%), and bottom (27%) groups to assess the instrument's ability to differentiate between performance levels.

Results. The assessment yielded a Cronbach's Alpha of 0.6766, indicating moderate internal consistency. Item analysis revealed a diverse range of psychometric performance: Items 2 and 7 were identified as problematic, exhibiting high difficulty ($p < 0.36$) and low discrimination ($D < 0.45$), suggesting structural ambiguity. Conversely, Items 3 and 5 were categorized as "Very Easy" ($p > 0.86$), functioning as foundational recognition tasks rather than measures of analytical depth. While some items showed excellent discrimination ($D = 0.80$), the findings indicate that 40% of the instrument requires strategic revision to better align with higher-order cognitive objectives.

Conclusion. The study concludes that while the ALLCP assessment provides a functional measure of learning, its diagnostic rigor is currently hampered by items that are either structurally ambiguous or cognitively simplistic. To improve the tool, "Difficult" items must be refined to eliminate trick distractors, and "Very Easy" items must be transitioned from simple recall to scenario-based application. Implementing these psychometric improvements will ensure the ALLCP remains a robust benchmark for professional excellence in law librarianship.

Originality. This research provides a rare data-driven evaluation of professional development assessments within the specific context of Philippine law librarianship. It demonstrates a practical application of item analysis that is often reserved for high-stakes academic testing to specialized professional certificate programs, offering a replicable model for other library associations to institutionalize quality assurance in their training curricula.

Keywords: Item analysis; Collection development; Library training assessment; Professional development; Learning assessment design

Introduction

The profession of academic law librarianship in the Philippines is slowly but surely being given notice in the country's legal education landscape. Since 2018, when the Legal Education Board (LEB) started to put things in order for academic law libraries, academic law librarianship has flourished. With the founding of the Network of Academic Law Librarians, Inc. (NALL), a Securities and Exchange Commission-recognized professional organization that pushes for the upskilling and development of academic law librarians in the country, this profession found its sustainability.

In November 2022, NALL, in partnership with LEB's Legal Education Advancement Program (LEAP), through the leadership of Prof. Ma. Tanya Karina Lat started offering the Basic Training Module of the Academic Law Librarians Certificate Program (ALLCP), where basic courses (1) Collection Development, (2) Acquisition, (3) Cataloging and Classifying, (4) Indexing and (5) Library Services were offered. The Module was conducted for 44 hours: 16 hours for Creating a Collection Development Program for Law Collection and Building a Law Collection for an Academic Community, 8 hours for Acquisition of Law Resources, 8 hours for Cataloging and Classifying Law Resources, 4 hours for Indexing Law Resources, and 8 hours for Legal Education-Centered Library Services (LEAP, 2022). All courses were evaluated using 10 to 20-item quizzes. Moreover, as NALL plans to offer another round of the ALLCP Basic Module, it is necessary to assess the quality of the assessment tool used to ensure that it fits the course being assessed.

Assessment is vital in professional development programs, especially in library and information science, where lifelong learning is essential. Well-designed assessments not only measure what participants have learned but also serve as tools for evaluating the effectiveness of training content and instructional strategies (Brown & Knight, 2012; Chevalier, 2004; Turnbow & Zeidman-Karpinski, 2016). They provide critical feedback to learners and educators, ensuring that learning objectives are met and professional competencies are developed effectively.

However, the tools used, such as quizzes and tests, must be carefully evaluated for assessments to serve this purpose. A poorly constructed test may include questions that are too easy or too difficult, items that do not align with the learning goals, or items that fail to distinguish between more knowledgeable and less knowledgeable participants (Nedeau-Cayo et al., 2013; Xie et al., 2019). Failure to address these issues may mislead assessment results, which may compromise the effectiveness of the training (Haladyna & Rodriguez, 2013).

This situation is where item analysis becomes important. Item analysis is a method of evaluating individual test items based on how participants respond to them. By looking at data such as the number of correct and incorrect answers from multiple choice questions, the difficulty level of each item, and how well the items differentiate between high- and low-performing learners, educators can gain insights into the quality of the assessment (Ashraf & Jaseem, 2020; Rezigalla, 2022). These insights can then be

used to improve future versions of the test, making the assessment tool more accurate, reliable, and helpful in guiding learning.

Saponaro and Evans (2019) applied item analysis to a quiz during a professional training session on collection development, a core function in librarianship that involves selecting, evaluating, and managing library resources. The study aimed to assess each quiz item's performance and demonstrate how item analysis can improve the quality of assessments in library training programs. This case study supports the broader aim of enhancing teaching and learning practices by continuously improving assessment tools. By using item analysis as a reflective and data-driven process, library educators can better ensure that their assessments truly support the development of knowledge and skills among information professionals (Maulina & Novirianthy, 2020). This study will serve as a guide in assessing the quality of the assessment tools used to measure the effectiveness of the different courses offered in the ALLCP.

The first course offered in ALLCP was Collection Development, which also happened to be the longest and most intensive course in the program. To evaluate the effectiveness of this course, participants completed a 10-item multiple-choice quiz administered after an 8-hour lecture, workshop sessions, and asynchronous activities. This study aims to examine the quality and effectiveness of the assessment tool used in the Collection Development course by conducting an item analysis of the post-quiz results.

Specifically, the study seeks to answer the following research questions:

1. How did the participants perform in the quiz following the lecture and workshops on collection development?
2. How effective is the ALLCP assessment tool for the Collection Development course?
3. What recommendations can be made to improve the design of the assessment tool for future training sessions?

Review of Related Literature

A practical training and continuing professional development (CPD) program rely on well-designed content and sound assessment practices to ensure that learning outcomes are achieved (Filipe et al., 2014; Marceau, 2023). One often overlooked yet critical component of professional development is evaluating the assessment tools. Evaluation materials, such as quizzes, tests, and performance-based tasks, must be regularly reviewed and refined to ensure validity, clarity, and relevance (Dorado, 2024). It is particularly important in training programs for librarians, such as those focusing on collection development, where the rapid evolution of information resources and user needs necessitates current, high-quality training.

Evaluating evaluation tools' effectiveness has long been recognized as a key strategy in instructional improvement. Guskey (2000) asserts that high-quality assessments help instructors determine whether participants gain the intended knowledge and skills. Without reliable evaluation tools, trainers cannot make evidence-based decisions about modifying content or teaching strategies. Assessment tools that are vague, overly difficult, or misaligned with course objectives can obscure actual learning gains, leading to misleading results and missed opportunities for improvement. Regularly assessing evaluation instruments is essential for professional development programs aiming to maintain instructional excellence.

A common method for evaluating the quality of an assessment tool is item analysis, which involves examining individual test items to determine how well they function to distinguish high and low performers (Rezigalla, 2022). This process includes calculating item difficulty, item discrimination, and internal consistency to identify which questions are practical and which need revision. According to Talebi et al. (2013), item analysis helps eliminate questions that may confuse learners or fail to measure the intended competencies. This approach not only strengthens the overall reliability of the tool but also improves the learning experience for participants.

The field of librarianship, primarily academic and law librarianship, has undergone profound changes over the past few decades (Balleste et al., 2006; McAllister et al., 2022). With the rise of digital content, e-resources, open access, and hybrid library models, librarians are expected to adapt quickly to emerging trends and technologies. Continuous training in core areas like collection development has become essential. Collection development, a foundational aspect of librarianship, encompasses resource selection, acquisition, evaluation, and maintenance in various formats. Training in this area enables librarians to make informed decisions that align with institutional goals, user needs, and budget constraints (Saponaro & Evans, 2019).

To ensure these training efforts are practical, assessment tools used in such programs must accurately reflect the skills and knowledge being taught. Poorly designed quizzes or evaluation forms may fail to measure learning outcomes and discourage participants, particularly adult learners, from fully engaging with the material. According to Knowles, Holton, and Swanson (2014), adult learners need assessments that are clearly relevant, respectful of their prior experience, and directly tied to their practical responsibilities. Assessments are especially relevant for mid-career professionals such as librarians, who often participate in CPD programs while juggling institutional responsibilities.

Merriam and Bierema (2014) further emphasize that adult learning is most effective when it is self-directed, experience-based, and problem-centered. In this context, evaluation tools must measure what participants have learned and allow for reflection, application, and feedback. Assessments that fail to meet these criteria risk undermining the learning experience and may not provide trainers with valuable data

for improving future iterations of the course. These principles are likewise reflected in international frameworks such as ISO 29993:2017, which mandates the evaluation of learning through valid and reliable methods, particularly in non-formal education and training environments (ISO 29993:2017-Learning Services Outside Formal Education, 2017).

Programs such as the ALLCP have emerged as important vehicles for professional development for law librarians in the Philippines and similar contexts. With its focus on specialized areas like collection development, ALLCP plays a vital role in strengthening institutional capacity and improving the quality of legal information services. However, the effectiveness of such programs hinges not only on the delivery of content but also on the strength of their evaluation mechanisms. Systematic assessment of tools, particularly quizzes and post-training evaluations, can reveal critical insights about content clarity, learning retention, and question relevance.

Varlejs et al. (2016) note that continuing professional development is central to sustaining a competent and future-ready library workforce. As libraries shift from being repositories of books to becoming dynamic knowledge hubs, the competencies required of librarians also evolve. Ensuring that training programs are responsive to these shifts requires rigorous evaluation of participant performance and the tools used to gauge that performance. By improving these tools through item analysis and other evaluation strategies, programs like ALLCP can enhance the learning outcomes of their participants and contribute to long-term institutional development. This aligns with the competency framework of the American Association of Law Libraries, that is strengthened by the organization's Body of Knowledge (BoK). The said framework emphasizes continuous professional growth and active engagement in continuing education as a core competency of law librarians (American Association of Law Libraries, 2010, 2018).

Moreover, these efforts align with the United Nations' Sustainable Development Goal 4 (Quality Education), which advocates for inclusive and equitable quality education and lifelong learning for all. Professional development programs incorporating robust assessment strategies contribute to this goal by ensuring the learning process is meaningful, measurable, and adaptable to changing societal needs.

Assessing evaluation materials, particularly through item analysis, is crucial in enhancing the effectiveness of training programs for librarians. It ensures alignment between instructional objectives and learning outcomes, supports adult learning principles, and contributes to the broader goals of professional excellence and institutional resilience. In collection development training, especially under initiatives like the ALLCP, such evaluation practices are beneficial and essential.

Finally, the consistent use and assessment of quizzes across training modules are not merely academic exercises but function as mechanisms for fulfilling a professional and regulatory mandate. The standards established under Republic Act No. 9246 (The Philippine Librarianship Act of 2003, 2024) and Republic Act No. 10912 (Continuing

Professional Development Act of 2016, 2016) require that librarians engage in competency-based professional development with demonstrable learning outcomes, while LEBMO 26, s.2022 (Technical Working Group, 2022) supports the continuing education, training, and development of all academic law library personnel, including their support staff. Within this framework, assessment, such as the use of Item Analysis, is not optional but integral to professional accountability. Item analysis, therefore, operates as an evidentiary tool that ensures evaluation instruments are valid, reliable, and aligned with intended competencies, thereby confirming that training initiatives produce measurable and defensible outcomes consistent with the expectations of a regulated profession.

Conceptual Framework

This study is grounded in the belief that effective professional development hinges on the quality of assessment tools used to measure learning outcomes. In the context of the ALLCP, where the Collection Development course plays a foundational role, the conceptual framework seeks to understand how evaluating the quality of assessment materials (e.g., quizzes) can support the broader goal of improving training sessions and fostering professional growth among librarians.

Anchored on an integrated framework that draws upon Classical Test Theory (CTT), this study incorporates three educational and quality improvement models: (1) Kirkpatrick's Four-Level Training Evaluation Model; (2) Constructive Alignment Theory; and (3) Continuous Quality Improvement (CQI) principles. Together, these perspectives provide a comprehensive lens for assessing and refining the effectiveness of the assessment tool used in the Collection Development course for librarians undergoing professional development.

Classical Test Theory (CTT) is at the center of the framework, which serves as the primary psychometric approach to analyzing the post-quiz results for small-scale assessments (Ohiri & Okoye, 2023). CTT posits that an observed score comprises a true score and an error component. The study determines how well each item differentiates among high- and low-performing participants by evaluating item difficulty and item discrimination indices. This analysis helps identify poorly performing items that may require revision, thereby enhancing the overall reliability and validity of the assessment tool.

Three theoretical models that provide pedagogical and developmental context complement the quantitative component. First, Kirkpatrick's Four-Level Training Evaluation Model, which emphasizes how assessment of knowledge contributes to performance improvement (Kirkpatrick & Kirkpatrick, 2006; Praslova, 2010), guides the interpretation of the participants' performance data. The model evaluates training at four levels:

1. Reaction – how participants perceive the learning experience;
2. Learning – the extent of knowledge or skills acquired;
3. Behavior – the degree to which learning is applied on the job; and
4. Results – the broader impact of training outcomes.

This study aligns particularly with the first two levels: (1) Reaction through anecdotal participant feedback on quiz items, and (2) Learning as measured by quiz scores and item analysis.

Second, Constructive Alignment Theory underscores the necessity of aligning learning outcomes, teaching strategies, and assessment tasks (Biggs, 1996). The theory presents that assessments should measure what was taught and should be deliberately aligned with intended learning outcomes. This study reviews item clarity and content relevance to ensure that the post-quiz accurately reflects the competencies targeted during the Collection Development sessions. Where misalignment is found, e.g., if quiz items do not reflect workshop objectives, recommendations for revision are made.

Finally, the framework incorporates principles from Continuous Quality Improvement (CQI), a cyclical process that emphasizes data-driven decision-making and ongoing refinement (Radawski, 1999). The item analysis acts as a feedback mechanism, enabling the training program to improve its evaluation tools iteratively. Through this lens, the study views the assessment tool not as a static product but as a dynamic instrument subject to review, improvement, and re-validation.

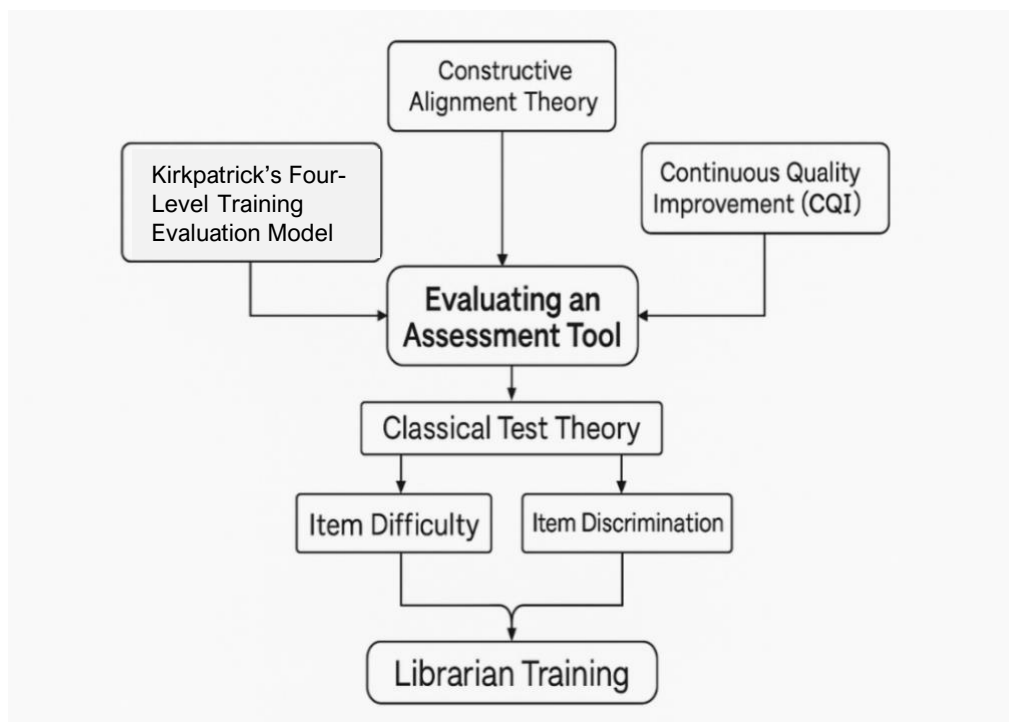


Fig. 1 The Conceptual Framework of this study (Authors' own work)

In synthesis, this conceptual framework integrates CTT's statistical rigor with the theoretical grounding provided by Kirkpatrick's evaluation model, Constructive Alignment Theory, and CQI. The resulting framework enables a multifaceted investigation into the design, effectiveness, and future improvement of the assessment

tool used in librarian training programs. Fig. 1 presents the diagram for the conceptual framework of this study.

Methodology

Research Design

This study employed a quantitative descriptive research design utilizing classical test theory (CTT) to perform item analysis on a multiple-choice quiz administered during a training session on collection development. The goal was to evaluate the quality of individual test items in terms of difficulty, discrimination, and reliability.

Participants

A total of 52 participants completed the assessment. They consisted of librarians and library staff who attended a professional development session on collection development. Based on their overall quiz scores, participants were grouped into three performance tiers using tertile ranking: top 27%, middle 46%, and bottom 27%. This stratification followed standard item analysis procedures to identify the discriminative power of each question.

Instrument

The instrument was a 10-item multiple-choice quiz designed to assess participants' knowledge of collection development concepts, policies, and practices. Each item had one correct answer and distractors based on common misconceptions or related ideas.

Data Collection and Processing

Participant responses were encoded in a spreadsheet. Each item's data was tabulated to include the number of correct and incorrect responses overall and broken down by top, middle, and bottom student groups.

Statistical Measures

The following metrics were calculated for each item:

1. Difficulty Index (p) – The proportion of participants who correctly answered the item.
2. Discrimination Index (D) – Determined by comparing the proportion of correct responses between the top and bottom student groups.
3. Cronbach's Alpha (α) – Calculated to assess the internal consistency reliability of the overall quiz instrument. The Cronbach's Alpha was extracted using *Jamovi* computer software version 2.5.

The statistical computations for the Difficulty Index and Discrimination Index were performed using Microsoft Excel with built-in functions and formula-based analysis below:

Difficulty Index (p)

$$p = \frac{\text{Total number of participants who answered the item}}{\text{Number of participants who answered correctly}}$$

Discrimination Index (D)

$$D = \frac{(\text{High group correct count} - \text{Low group correct count})}{\text{Number of participants in each group}}$$

Results and Discussion

The post-training quiz for the training session on Collection Development was designed to evaluate how well participants grasped core concepts after attending lectures and workshops. With 52 participants submitting quiz responses, the data presents a clear picture of strengths and gaps in comprehension.

Participants' Performance Overview

The analysis of participants' performance across the ten test items on collection development provides meaningful insights into their level of understanding and the relative difficulty of the questions. Overall, the internal consistency of the test, as indicated by Cronbach's Alpha ($\alpha = 0.677$). This result suggests an acceptable reliability level for a 10-item quiz, as test lengths has a direct impact to alpha, where shorter item-tests produce lower alphas (Edelsbrunner et al., 2025; Uyanah & Nsikhe, 2023).

Items with higher correct student ratios demonstrated greater accessibility. For instance, Item 5 ("The collection development criteria change as user needs and the information environment change") recorded the highest correct student count (50, ratio = 0.96) with minimal variance (0.037). Even among top and bottom performers, the majority answered correctly (20 top participants, 4 bottom participants), underscoring the broad consensus around this foundational principle. Similarly, Item 3 ("Which among these processes is an activity in completing the collection development process?") yielded a high correct ratio (0.87), with 20 top participants and 4 bottom participants answering correctly. These items, therefore, can be considered "easy" but still helpful in reinforcing essential knowledge.

In contrast, items with lower correct ratios revealed areas of conceptual difficulty. Item 2 ("The following are the basic components of a collection development policy, except:") had the lowest correct ratio (0.31), with only 16 participants overall answering correctly. Even among top performers, only 9 answered correctly, while just two from the bottom group succeeded. This result suggests that the item may have been ambiguously worded or that misconceptions persist regarding the structural components of collection development policy. Item 7 ("The following are reasons why

we conduct collection assessment except:") showed a similar pattern, with only 18 correct responses overall and no correct responses from bottom participants. The absence of correct responses from weaker performers indicates that the item did not discriminate effectively, suggesting the need to refine distractors.

Items 1, 6, and 10 fell within the moderate range of difficulty, with correct student ratios around 0.50. These items demonstrated strong discrimination, as evident in the higher proportion of correct responses among top participants compared to bottom participants (e.g., in Item 6, 15 top participants answered correctly, while none from the bottom group did). This stark contrast suggests that these items effectively distinguish between high- and low-performing participants.

Interestingly, certain items demonstrated high accessibility yet retained discriminative value. Item 4 ("The most important factor in building a collection development program is the ____") had a correct ratio of 0.69, with 19 top participants answering correctly compared to just 1 bottom student. Similarly, Item 8 ("This type of collection assessment focuses on the degree to which the users find the collection useful") achieved a 0.72 correct ratio with 17 top participants versus 3 bottom participants answering correctly. These results indicate that while these items are relatively easy, they still effectively distinguish stronger from weaker learners.

Including the Correct Top Student Count and the Correct Bottom Student Count highlights nuances not captured by the correct ratios alone. For example, while Item 9 had a modest correct ratio (0.40), it showed strong discrimination: 14 top participants answered correctly, compared to none in the bottom group. This result demonstrates that even moderately difficult items can serve as excellent discriminators, identifying differences in mastery levels.

Table 1. Participants' Quiz Performance and the Internal Consistency of the Quiz Items (Authors' own work)

Item No.	Items	Correct Student Count	Correct Top Student Count	Correct Bottom Student Count	Variance
1	Which of the following statements is true about collection development policy?	26	15	1	0.25
2	The following are the basic components of a collection development policy, except: Which among these processes is an activity in completing the collection development process?	16	9	2	0.213018
3	The most important factor in building a collection development program is the ____.	36	19	1	0.213018
4	The collection development criteria change as user needs and the information environment change.	50	20	4	0.036982
5	Minimum, Basic Information, Instructional Support, and Research are assigned ratings in ____.	26	15	0	0.249904

7	The following are reasons why we conduct collection assessment except: This type of collection assessment focuses on the degree to which the users find the	18	9	0	0.2304
8	collection useful. In conducting user surveys, using a	36	17	3	0.2016
9	questionnaire is better than interviewing respondents	20	14	0	0.24
10	There is a perfect assessment measure for every situation.	26	14	1	0.2496

These findings (Table 1) suggest that participants possess strong familiarity with overarching principles and procedural aspects of collection development (Items 3, 4, and 5), but continue to face challenges in nuanced or technical areas, such as policy components (Item 2), assessment purposes (Item 7), and contextual application of rating systems (Item 6). These insights emphasize the importance of balancing item difficulty while ensuring clarity and alignment with learning outcomes.

Item-Level Analysis

The results of the item analysis reveal a nuanced distribution of difficulty and discrimination indices across the ten items included in the Academic Law Librarian Certificate Program (ALLCP). Item difficulty, expressed as the proportion of participants answering each question correctly, ranged from 0.31 to 0.96, reflecting a broad spectrum of item complexity. Items such as Item 2 ("The following are the basic components of a collection development policy, except") and Item 7 ("The following are reasons why we conduct collection assessment except") demonstrated low difficulty indices (0.31 and 0.36, respectively), indicating that these were among the most challenging for participants. Conversely, Item 5 ("The collection development criteria change as user needs and the information environment change") had the highest difficulty index at 0.96, suggesting that most participants found this concept straightforward and unambiguous.

The discrimination index, operationalized by comparing the performance of the top 27% of participants with the rest of the group, provides further insight into item quality. Items such as Item 1, Item 6, and Item 10 displayed strong discrimination values (~0.25), reflecting their effectiveness in differentiating high-performing participants from their peers. By contrast, Item 3, with a lower variance value (0.116), exhibited limited discriminative capacity, suggesting that both high- and low-achieving participants performed similarly on this question. These findings may indicate that the item was either too easy or insufficiently aligned with higher-order cognitive skills required to challenge advanced learners.

The integration of top student performance in the computation of discrimination indices is particularly noteworthy. By selecting the top 27% of examinees, based on their overall test performance, the analysis isolates the subgroup most likely to possess mastery of the content. Items that these participants answer correctly at a disproportionately higher rate compared to the lower-performing subgroup are regarded as effective discriminators. For example, Item 1 ("Which of the following

statements is true about collection development policy?") recorded 15 correct responses among top participants, underscoring its utility in distinguishing levels of understanding. This methodological approach underscores the importance of considering subgroup performance, as it ensures that assessment tools remain sensitive to differences in cognitive proficiency across the examinee population.

Effectiveness of the ALLCP Assessment Tool

The ALLCP demonstrated an acceptable level of internal consistency, with Cronbach's alpha values consistently hovering around 0.677 across items. While slightly below the conventional threshold of 0.70, this value nonetheless indicates a reasonably reliable instrument for assessing knowledge in collection development. Notably, the instrument's mix of easy, moderate, and difficult items aligns with established best practices in test construction, ensuring that the tool can capture foundational knowledge and a more nuanced understanding.

Table 2. Diagnostic Matrix of Quiz Items (Authors' own work)

Item No.	Item Description	Difficulty Index (p)	Discrimination Index (D)	Difficulty Level	Discrimination Level	Diagnostic Comment
1	Which of the following statements is true about collection development policy?	0.50	0.70	Moderate	Good	Balanced difficulty; discriminates well.
2	The following are the basic components of a collection development policy, except:	0.308	0.35	Difficult	Fair	Too challenging; low discrimination; may need revision.
3	Which among these processes is an activity in completing the collection development process?	0.865	0.80	Very Easy	Excellent	Very easy but discriminates well; may consider adjusting wording.
4	The most important factor in building a collection development program is the ____	0.692	0.90	Easy	Excellent	Slightly easy; excellent discrimination.
5	The collection development criteria change as user needs and the information environment change.	0.962	0.80	Very Easy	Excellent	Very easy; excellent discrimination.
6	Minimum, Basic Information, Instructional Support, and Research are assigned ratings in ____.	0.510	0.75	Moderate	Good	Well-balanced item; discriminates well.
7	The following are reasons why we conduct collection assessment except:	0.36	0.45	Difficult	Moderate	Challenging; acceptable discrimination.
8	This type of collection assessment focuses on the degree to which the users find the collection useful.	0.72	0.70	Easy	Good	Slightly easy; discriminates well.
9	In conducting user surveys, using a questionnaire is better	0.40	0.70	Moderate	Good	Borderline moderate; good discrimination.

Item No.	Item Description	Difficulty Index (p)	Discrimination Index (D)	Difficulty Level	Discrimination Level	Diagnostic Comment
10	than interviewing respondents There is a perfect assessment measure for every situation.	0.52	0.65	Moderate	Good	Balanced difficulty and discrimination.

The balance of items suggests that the ALLCP effectively meets its dual purposes: (1) measuring core competencies in collection development among participants, and (2) identifying specific areas of weakness for instructional improvement. The presence of highly discriminating items strengthens its diagnostic value, enabling instructors to pinpoint knowledge gaps among lower-performing participants while confirming the mastery of advanced learners.

However, the tool's limitations must also be acknowledged. Items with very high difficulty indices (e.g., Item 5, with 0.96) may contribute little to distinguishing between participants, as nearly all respondents answer them correctly. Similarly, items with low discrimination indices may need to be revised or replaced to improve the overall robustness of the assessment. A refined instrument that balances item difficulty more evenly and maximizes discrimination would further enhance the reliability and validity of the ALLCP. Table 2 presents the matrix of the quiz items.

Recommendations for the Quiz Items

The item analysis of the ten multiple-choice questions yielded important insights into the overall effectiveness of the instrument in assessing knowledge of collection development concepts. Both the difficulty index (p) and discrimination index (D) were considered in determining whether each item should be retained, revised, or replaced.

Table 3. Decisions on Retention, Improvement, or Revision of Items (Authors' own work)

Item No.	Item Description	Decision	Rationale
1	Which of the following statements is true about collection development policy?	Retain	Balanced in difficulty and shows good discrimination, making it an effective item.
2	The following are the basic components of a collection development policy, except:	Revise	Item is too difficult with weak discrimination, suggesting ambiguity or misleading choices.
3	Which among these processes is an activity in completing the collection development process?	Improve	Item is very easy yet discriminates well; minor revision in wording can enhance challenge.
4	The most important factor in building a collection development program is the ____	Retain	Slightly easy but with excellent discrimination, showing it effectively distinguishes learners.
5	The collection development criteria change as user needs and the information environment change.	Improve	Extremely easy, though with strong discrimination; revision can reduce predictability.
6	Minimum, Basic Information, Instructional Support, and Research are assigned ratings in ____.	Retain	Well-balanced in both difficulty and discrimination; functions effectively as a test item.
7	The following are reasons why we conduct collection assessment except:	Revise	Difficult with only moderate discrimination, indicating possible wording clarity issues.
8	This type of collection assessment focuses on the degree to which the users find the collection useful.	Retain	Slightly easy but with good discrimination, confirming reliability.

Item No.	Item Description	Decision	Rationale
9	In conducting user surveys, using a questionnaire is better than interviewing respondents	Retain	Borderline moderate difficulty with good discrimination; valid for measuring understanding.
10	There is a perfect assessment measure for every situation.	Retain	Balanced in both difficulty and discrimination, making it a sound item.

As shown in Table 3, six items (Items 1, 4, 6, 8, 9, and 10) are recommended for retention, since they have moderate to easy difficulty levels with strong discrimination power. These items reflect well-constructed questions that balance accessibility with the ability to distinguish between high- and low-performing examinees.

These results suggest that most items in the instrument are functioning well, but selective revisions will enhance the overall reliability and validity. These revisions should focus on refining clarity, balancing the difficulty level, and ensuring that items remain aligned with the intended learning outcomes.

Analysis of Items Targeted for Revision

The results show that four items (Items 2, 3, 5, and 7) are recommended for revision.

Item 2: Ambiguity and Structural Flaws

This paper identified specific questions requiring immediate refinement to improve the diagnostic utility of the assessment tool. Most prominent among these was Item 2, which addressed the components of a collection development policy. Despite being classified as "Difficult" with a difficulty index (p) of 0.308, the item yielded a discrimination index (D) of only 0.35, categorized as "Fair." In an optimized assessment, high-difficulty items are expected to be mastered predominantly by top-tier performers; however, the data revealed that only 64% of the high-achieving group (9 out of 14) selected the correct response. The low overall success rate—with only 16 out of 52 participants answering correctly—suggests the presence of structural flaws. Specifically, the "except" qualifier may have introduced an unnecessary cognitive load, or the distractors may have functioned as "trick" options rather than valid measures of professional knowledge. Consequently, it is recommended that this item be revised to ensure the "exception" is distinct from core policy components, or that distractors are adjusted to be less plausible to those possessing foundational knowledge.

Item 7: Distractor Ineffectiveness and Zero-Discrimination Risk

Similarly, Item 7, regarding the rationales for collection assessment, presented significant psychometric challenges. While its difficulty index ($p = 0.36$) and discrimination index ($D = 0.45$) appear mathematically acceptable, a more granular analysis of the raw data reveals a critical diagnostic gap: zero participants from the bottom-performing group answered the item correctly. Such a total failure among low-performing students often indicates that the distractors are too similar to the correct choice, reducing the item to a matter of chance rather than logical deduction.

Furthermore, the 50% failure rate among the top-performing group suggests a potential deficiency in content validity, implying that the item may be testing nuances not adequately addressed during the instructional module. To rectify this, the item requires revision to clarify its phrasing and ensure that the correct response is explicitly grounded in the provided training materials.

Item 3: Increasing Procedural Complexity

Similarly, Item 3 demonstrated a high ease of response ($p = 0.865$), suggesting that while it successfully discriminated between performance groups ($D = 0.80$), its distractors were likely too distinct to challenge the learners effectively. The original prompt required only the identification of a single activity within the collection development lifecycle. To enhance the procedural complexity of this item, the proposed revision focuses on the relational link between "Selection" and "Evaluation." This modification requires participants to understand the sequential and strategic relationships between library functions, thereby elevating the cognitive demand from "Remembering" to "Understanding." By implementing these strategic adjustments, the assessment moves beyond surface-level recognition and toward a model of constructive alignment, ensuring that the ALLCP modules provide a more nuanced and rigorous measure of professional competency in law librarianship.

Item 5: From Common Sense to Professional Application

Meanwhile, items with high facility rates can bolster participant confidence, those possessing a difficulty index (p) as high as 0.962, such as Item 5, lose their functional utility in measuring substantive learning depth. In its original form, Item 5 ("The collection development criteria change as user needs and the information environment change") functioned as a foundational "truism," allowing nearly all participants to arrive at the correct response through general intuition rather than specialized training. To increase the diagnostic rigor of this item, a shift from simple recognition to professional application is proposed. By transitioning to a scenario-based prompt, such as analyzing the impact of a digital transformation on specific collection criteria, the assessment forces participants to exercise professional judgment regarding budget allocation or technical infrastructure. This adjustment aligns with the upper tiers of Bloom's Taxonomy, moving the evaluator's focus from mere fact acknowledgment to the analysis of complex institutional shifts.

As shown in Appendix A, the revision strategy for the ALLCP assessment focuses on two fronts: (1) clarifying 'Difficult' items with low discrimination (Items 2 and 7) to ensure they accurately measure mastery, and (2) increasing the cognitive depth of 'Very Easy' items (Items 3 and 5) by moving from basic recognition to situational analysis. This alignment ensures the tool remains a rigorous measure of professional competency in law librarianship.

Implications for Instructional Assessment Design

The findings have several implications for designing and using instructional assessments in library and information science education. First, the wide range of item

difficulties suggests that assessment instruments should be carefully calibrated to ensure that they challenge participants across the performance spectrum. Items that are too easy or too difficult may either fail to engage advanced learners or discourage less proficient participants, undermining the assessment's diagnostic utility.

Second, the role of discrimination indices highlights the importance of constructing items that tap into higher-order cognitive processes. Questions that require mere recall may achieve high difficulty indices but often lack discriminatory power. Conversely, items designed to test application, analysis, and synthesis skills tend to produce stronger discrimination, enhancing the instrument's ability to distinguish between levels of student understanding.

Third, the consistency of Cronbach's Alpha across items underscores the value of reliability testing as a central component of assessment validation. Educators designing similar tools should prioritize item-level analysis and psychometric evaluation of internal consistency, ensuring that instruments measure intended constructs cohesively.

Finally, the results reinforce the need for a dynamic approach to assessment design. As the information environment continues to evolve, so too must the assessment tools used to measure competencies in collection development. Continuous refinement of the ALLCP, through item revision, pilot testing, and psychometric validation, will ensure that it remains pedagogically relevant and empirically sound.

Conclusion and Recommendations

The item-level analysis of the Academic Law Librarian Certificate Program (ALLCP) instrument provides compelling evidence of its instructional value and diagnostic potential. The findings demonstrate that while the tool adequately covers a spectrum of difficulty levels, from items that most participants answered correctly to those that challenged even the top performers, it also reveals specific areas where refinements are necessary to ensure balanced discrimination across student ability levels.

The consistently moderate reliability (Cronbach's Alpha = 0.677) signals that the assessment achieves a degree of internal consistency but can be further strengthened by revising ambiguous or overly difficult items. Notably, this study identified a clear "revision roadmap" for the ALLCP: Items 2 and 7 require structural clarification to improve their low discrimination indices, while Items 3 and 5 should be adjusted to increase cognitive rigor, shifting from basic recall to situational application.

By pinpointing which items discriminate effectively between high- and low-performing participants, educators gain a more transparent window into patterns of student misunderstanding, curricular gaps, and areas for targeted instructional intervention. Ultimately, this study affirms the potential of item analysis as a powerful practice in instructional assessment design. When paired with an iterative approach to teaching, the ALLCP quizzes do more than evaluate learning, they actively shape a more

rigorous and effective learning environment for the next generation of Filipino law librarians.

Based on the findings and analyses, this study recommends the following:

1. **Revision Strategy**
Prioritize the revision of Items 2, 3, 5, and 7 using the "Appendix A" roadmap to ensure the quiz captures higher-order analytical skills rather than just foundational recognition
2. **Routine Item Analysis**
Trainors are encouraged to conduct periodic item analysis of their own assessments. This practice can help identify persistently weak areas in item design and student comprehension.
3. **Alignment with Learning Outcomes**
Future revisions of the ALLCP should explicitly map each item to the intended learning outcomes of collection development courses, ensuring more substantial alignment between instruction and evaluation.
4. **Integration into Teaching Practice**
Beyond grading, results from the ALLCP should be used to inform lesson planning, classroom interventions, and student feedback sessions, positioning assessment as a tool for continuous learning improvement.
5. **Further Reliability Testing**
Expanding the sample size and re-administering the ALLCP across diverse contexts will help establish stronger psychometric reliability and generalizability of the tool.
6. **Professional Development for Faculty**
Training sessions on test construction, item writing, and psychometric analysis should be provided to instructors to maximize the effectiveness of assessment in supporting student learning and program evaluation.

References

- American Association of Law Libraries. (2010, April). *AALL competencies of law librarianship*. <https://www.aallnet.org/wp-content/uploads/2022/08/AALL-Competencies-of-Law-Librarianship.pdf>
- American Association of Law Libraries. (2018). *AALL body of knowledge [BoK]*. <https://www.aallnet.org/wp-content/uploads/2023/02/BoK-FINAL-02.8.2023-Hi-Res.pdf>

- Ashraf, Z. A., & Jaseem, K. (2020). Classical and modern methods in item analysis of test tools. *International Journal of Research and Review*, 7(5), 397–403. [IJRR005820200611-48713-y31tfs-libre.pdf](https://doi.org/10.5771/9781461669128)
- Balleste, R., Smith-Butler, L., & Luna-Lamas, S. (2006). *Law librarianship in the twenty-first century*. Bloomsbury Publishing PLC. <https://doi.org/10.5771/9781461669128>
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347–364. <https://doi.org/10.1007/BF00138871>
- Brown, S., & Knight, P. (2012). *Assessing learners in higher education*. Routledge.
- Chevalier, R. (2004). Evaluation: The link between learning and performance. *Performance Improvement*, 43(4), 40. <https://www.proquest.com/scholarly-journals/evaluation-link-between-learning-performance/docview/237232242/se-2>
- Continuing professional development act of 2016*. (2016, July 21). Senate of the Philippines. <https://elibrary.judiciary.gov.ph/thebookshelf/showdocs/2/70325>
- Dorado, F. A. A. (2024). Assessing the craft: Use of specific descriptors as guide for evaluating instructional materials. *English Journal Antartika*, 2(1), 16–23. <https://doi.org/10.70052/eja.v2i1.450>
- Edelsbrunner, P. A., Simonsmeier, B. A., & Schneider, M. (2025). The Cronbach's Alpha of domain-specific knowledge tests before and after learning: A meta-analysis of published studies. *Educational Psychology Review*, 37(1), 4-. <https://doi.org/10.1007/S10648-024-09982-Y>
- Filipe, H. P., Silva, E. D., Stulting, A. A., & Golnik, K. C. (2014). Continuing professional development: best practices. *Middle East African journal of ophthalmology*, 21(2), 134–141. <https://doi.org/10.4103/0974-9233.129760>
- Guskey, T. R. (2000). *Evaluating professional development*. Corwin Press.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.
- ISO 29993:2017 - Learning services outside formal education*. (2017). International Organization for Standardization (ISO). <https://www.iso.org/standard/70357.html>
- Kirkpatrick, D., & Kirkpatrick, J. (2006). *Evaluating training programs: The four levels*. Berrett-Koehler Publishers.
- Knowles, M. S., Holton III, E. F., & Swanson, R. A. (2014). *The adult learner: The definitive classic in adult education and human resource development*. Routledge.
- Legal Education Advancement Program (LEAP). (2022, November 2). *LEAP off the page: Academic Law Librarian Certification Program (ALLCP)*. Retrieved from

<https://www.facebook.com/photo/?fbid=164644712877877&set=a.157124570296558>

- Marceau, M., Lachiver, É. V., Lambert, D., Daoust, J., Dion, V., Langlois, M. F., ... & St-Onge, C. (2024). Assessment practices in continuing professional development activities in health professions: A scoping review. *Journal of Continuing Education in the Health Professions*, 44(2), 81–89. <https://doi.org/10.1097/CEH.0000000000000507>
- Maulina, N., & Novirianthy, R. (2020). Item analysis and peer-review evaluation of specific health problems and applied research block examination. *Jurnal Pendidikan Kedokteran Indonesia: The Indonesian Journal of Medical Education*, 9(2), 131-137. <https://doi.org/10.22146/jpki.49006>
- McAllister, A. D., Flierl, M., Caswell, T. R., Costello, L., Hall, A. R., Li, C., & Walker, G. (2022). Top trends in academic libraries: A review of the trends and issues. *College & Research Libraries News*, 83(6), 243–253. <https://doi.org/10.5860/crln.83.6.243>
- Merriam, S. B., & Bierema, L. L. (2014). *Adult Learning: Linking theory and practice*. Jossey-Bass.
- Nedeau-Cayoa, R., Laughlin, D., Rus, L., & Hall, J. (2013). Assessment of item-writing flaws in multiple-choice questions. *Journal for Nurses in Professional Development*, 29(2), 52–57. <https://doi.org/10.1097/NND.0b013e318286c2f1>
- Ohiri, S. C., & Okoye, R. O. (2023). Application of classical test theory as linear modeling to test item development and analysis. *International Research Journal of Modernization in Engineering Technology and Science*, 5(1), 2152–2159. <https://www.doi.org/10.56726/IRJMETS45379>
- Praslova, L. (2010). Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in higher education. *Educational assessment, evaluation and accountability*, 22(3), 215-225. <https://doi.org/10.1007/s11092-010-9098-7>
- Radawski, D. (1999). Continuous quality improvement: origins, concepts, problems, and applications. *The Journal of Physician Assistant Education*, 10(1), 12-16. <https://doi.org/10.1097/01367895-199910010-00002>
- Rezigalla, A. A. (2022). Item analysis: Concept and application. *Medical education for the 21st century*, 1-16.
- Saponaro, M. Z., & Evans, G. E. (2019). *Collection management basics*. Bloomsbury Publishing USA.
- Talebi, G. A., Ghaffari, R., Eskandarzadeh, E., & Oskouei, M. A. E. (2013). Item analysis an effective tool for assessing exam quality, designing appropriate exam and determining weakness in teaching. *Research and Development in Medical Education*, 2(2), 69–72. <https://doi.org/10.5681/RDME.2013.016>

Technical Working Group, L.E.B. (2022). *The academic law library standards and guidelines. 1.*

The Philippine librarianship act of 2003. (2024, February 19). Senate of the Philippines Legislative Reference Bureau. <https://issuances-library.senate.gov.ph/subject/the-philippine-librarianship-act-of-2003>

Turnbow, D., & Zeidman-Karpinski, A. (2016). Don't use a hammer when you need a screwdriver: How to use the right tools to create assessment that matters. *Communications in information literacy*, 10(2). <https://doi.org/10.15760/comminfolit.2016.10.2.30>

Varlejs, J., Lewis, V., Schnuer, S., & Jara de Sumar, J. (2016). *IFLA guidelines for continuing professional development: Principles and best practices.*

Uyanah, D. A., & Nsikhe, U. I. (2023). The theoretical and empirical equivalence of Cronbach Alpha and Kuder-Richardson formular-20 reliability coefficients. *International Research Journal of Innovations in Engineering and Technology (IRJIET)*, 7(5), 17–23. <https://doi.org/10.47001/IRJIET/2023.705003>

Xie, B., Davidson, M. J., Li, M., & Ko, A. J. (2019). An item response theory evaluation of a language-independent CS1 knowledge assessment. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 699–705). <https://doi.org/10.1145/3287324.3287370>

APPENDIX A
Proposed Item Revisions Based on Item Analysis

ITEM #	ORIGINAL METRIC	EVALUATION	PROPOSED REVISION / STRATEGY	RATIONALE FOR CHANGE
2	<i>p</i> =0.308 / <i>D</i> =0.35	Difficult / Fair	<p>Current: "The following are the basic components... except:"</p> <p>Proposed: "A law librarian is drafting a Collection Development Policy for the institution. Upon review, the director notices that one of the proposed sections does not align with standard CDP components. Which of the following should the director recommend removing?"</p>	Apply the scenario-based questioning rather than negative stems for items with lower discrimination scores.
3	<i>p</i> =0.865 / <i>D</i> =0.80	Very Easy / Good	<p>Current: "Which among these processes is an activity..."</p> <p>Proposed: "Which activity serves as the primary link between 'Selection' and 'Evaluation' in the CDP lifecycle?"</p>	Increase cognitive demand from simple identification to understanding the relationship between library processes.
5	<i>p</i> =0.962 / <i>D</i> =0.55	Very Easy / Fair	<p>Current: "CDP criteria change as user needs and the environment change."</p> <p>Proposed: "When a law library transitions from print-heavy to digital-first, which CDP criterion requires the most significant reassessment?"</p>	Shift from a general "truism" to a scenario-based application of professional judgment.
7	<i>p</i> =0.360 / <i>D</i> =0.45	Difficult / Moderate	<p>Current: "The following are reasons why we conduct collection assessment except:"</p> <p>Proposed: " A law library is conducting a formal collection</p>	Apply the scenario-based questioning rather than negative stems for items with lower discrimination scores.

assessment to support its strategic planning and budget justification. The team is identifying the most appropriate purposes for conducting the assessment. Which of the following best reflects a purpose that would justify a formal collection assessment in this context?"