

**INTERNATIONAL JOURNAL OF
MODERN EDUCATION
(IJMOE)**www.gaexcellence.com/ijmoe**STUDENTS AS INSTRUCTIONAL DESIGNERS: A
PRACTICE-BASED EVALUATION OF A STUDENT-LED
INTERNATIONAL SERVICE-LEARNING LIBRARY
OUTREACH PROGRAMME**

Nordibradini Selamat¹, Anaqi Ahmad Saleh², Nor Azlina Azmi³, Salasiah M Said⁴,
Sharunizam Shari^{5*}

¹ Faculty of Information Science, Universiti Teknologi MARA Kedah Branch, Merbok, Kedah, Malaysia

✉ nordi314@uitm.edu.my

iD <https://orcid.org/0009-0008-8039-6979>

² Faculty of Information Science, Universiti Teknologi MARA Kedah Branch, Merbok, Kedah, Malaysia

✉ anaqisaleh@uitm.edu.my

iD <https://orcid.org/0000-0001-9528-1404>

³ Faculty of Information Science, Universiti Teknologi MARA Kedah Branch, Merbok, Kedah, Malaysia

✉ nazlina@uitm.edu.my

iD <https://orcid.org/0009-0000-5603-7839>

⁴ Faculty of Information Science, Universiti Teknologi MARA Kedah Branch, Merbok, Kedah, Malaysia

✉ salasiah@uitm.edu.my

iD <https://orcid.org/0009-0004-9429-0949>

⁵ Faculty of Information Science, Universiti Teknologi MARA Kedah Branch, Merbok, Kedah, Malaysia

✉ nizam057@uitm.edu.my

iD <https://orcid.org/0000-0003-0052-8174>

*Corresponding Author

Article Info:**Article history:**

Received date: 03.08.2026

Revised date: 30.08.2026

Accepted date: 15.09.2026

Published date: 23.09.2026

To cite this document:

Selamat, N., Saleh, A. A., Azmi, N. A., M Said, S., & Shari, S. (2026). Students As Instructional Designers: A Practice-Based Evaluation Of A Student-Led International Service-Learning Library Outreach Programme.

Abstract:

In academic library outreach, satisfaction is often mistaken for evidence of learning. This study evaluates a one-day international service-learning programme in which undergraduate diploma students at a Malaysian branch campus designed, facilitated and delivered library instruction for twenty-eight visiting undergraduates from an Indonesian partner institution. It assesses both what participants reported learning and what they were observed doing during the instruction itself. A single-site case study combined documentary analysis of the programme archive with analysis of a questionnaire completed at the start of the programme and again at its close. Perceived information-search competence rose from $M = 3.47$ to $M = 4.72$ on a five-point scale, a within-participant gain of 1.26 points, 95% CI [0.99, 1.52], $p < .001$, and every participant whose score changed moved upward. Satisfaction was high ($M = 4.70$) but showed no association with the size of that gain ($r = .08$, $p = .71$), suggesting that satisfaction alone did not explain the reported improvement. Observation told a different story. During the single activity designed to exercise discovery-system searching, several

International Journal of Modern Education, 8(31), 1306-1327.

groups bypassed the system and browsed the subject shelves directly, completing the task without practising the skill. This gap between reported confidence and observed behaviour, rather than the measured gain, is the study's central finding. The design is single-group and pre-experimental, the outcome is self-reported, and no causal claim is made. The study contributes a replicable programme blueprint and a caution against reading self-reported competence as competence.

DOI: 10.35631/IJMOE.831076

Keyword:

Discovery Systems; Information Literacy; Library Outreach; Malaysia; Peer-Led Instruction; Self-Efficacy; Service-Learning



© The authors (2026). This is an Open Access article distributed under the terms of the Creative Commons Attribution (CC BY NC) (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact ijmoe@gaexcellence.com.

Introduction

Academic library outreach occupies an awkward position in the professional literature. It is widely practised, institutionally valued and frequently reported, and yet the reporting is dominated by a genre that might be called the appreciative programme narrative: a description of what was done, a photograph of participants smiling, and a satisfaction score in the high fours. Such accounts are not worthless. They circulate practical ideas and they document institutional effort (Baba & Abrizah, 2018). But they rarely permit a reader to judge whether anything was learned. Reviews of gamified and activity-based library instruction repeatedly conclude that engagement is well documented while learning outcomes are not, and that stronger evaluation is needed before educational value can be established (Adewojo & Adefila, 2025; Zhang & Hasim, 2023). Compounding this, self-reported confidence is commonly treated as though it were demonstrated capability, although students who report adequate search skills are still observed struggling with catalogue and discovery interfaces in practice (Mwiiyale et al., 2025). To these documented weaknesses we would add a third, less often remarked: individual-level responses are routinely summarised into frequency counts and then discarded, as they were in the present case. The consequence is a literature rich in programme description and thin in evidence, in which the most obvious objection to any reported gain, that participants were simply pleased, is rarely tested because the data needed to test it no longer exist.

This paper attempts something different with a programme of exactly that type. On 25 May 2025, thirty final-semester students of the Diploma in Library Informatics at Universiti Teknologi MARA (UiTM) Cawangan Kedah, Malaysia, delivered a one-day library outreach programme for twenty-eight visiting undergraduates from the Faculty of Vocational Studies, Universitas Airlangga (UNAIR), Indonesia, hosted at Perpustakaan Sultan Badlishah (PSB) and the campus e-Library facility. The programme was a course requirement (IML301 Library

Outreach) delivered under Malaysia's national Service-Learning Malaysia: University for Society (SULAM) policy framework.

Two features distinguish it from the standard case, and together they are the basis of the paper's claim to originality. First, the students were not the audience. They were the instructional designers, facilitators, evaluators and, in an important sense, the researchers: they wrote the activity specifications, ran the sessions and administered the questionnaire. The library staff hosted; the lecturer supervised; the students taught. Second, the audience was foreign, a Malaysia–Indonesia exchange conducted in a shared regional lingua franca between two ASEAN institutions. Cross-border library outreach of this South–South kind is barely present in the English-language literature, which remains dominated by North American and Western European practice (Smith & Smith, 2020).

We set out to write neither a promotional account nor an inflated one. The evaluation evidence available from this programme is weak in specific, nameable ways, and we say so. It is also stronger in one respect than most outreach evaluations, because recovering the individual-level responses permitted a direct test of whether enthusiasm for the programme predicted the reported learning. What the evidence finally supports is more interesting than a competence gain: it supports a documented mismatch between what participants believed they had learned and what the facilitators observed them doing.

The paper addresses four questions:

- RQ1. How can a student-led international library outreach programme be specified in sufficient detail to be replicated?
- RQ2. How did participants evaluate the programme's acceptability?
- RQ3. What change in perceived information-search competence did participants report between a pre-instruction baseline and the close of the programme?
- RQ4. How far can that reported change be distinguished from response bias, and what does participants' observed behaviour during the intervention add to the assessment?

In answering them the paper makes three contributions. First, it supplies a fully specified, costed and timed programme blueprint, including group sizes, durations and materials, so that the intervention can be reproduced rather than merely admired. Second, it demonstrates that two inexpensive analytic steps, recovering individual-level responses from the original forms and correlating satisfaction against measured gain, allow a routine outreach evaluation to test its own most obvious objection; both steps are available to the many programmes that already collect such data and discard it. Third, it reports an unplanned but instructive mismatch between what participants believed they had learned and what they were observed doing, which offers a concrete illustration of why self-reported competence should not be read as competence.

Literature Review

Service-Learning as Pedagogy, Not Charity

Service-learning is distinguished from volunteering by reciprocity: the service must meet a genuine community need and the academic learning of the student, with structured reflection binding the two (Bringle & Hatcher, 1995; Furco, 1996). It is classified among the high-impact educational practices associated with deep learning and persistence (Kuh, 2008).

Malaysia has institutionalised the approach nationally. Service-Learning Malaysia: University for Society (SULAM) has been embedded in the curriculum as a high-impact educational practice since 2015 (Jamhari et al., 2024) and requires participating courses to integrate community engagement into assessed coursework. Jamhari et al. (2024) specify that compliant implementation must address personal growth, civic learning and academic enhancement together, and that qualifying community partners are those identified through consultation with non-profit organisations, government or communities. Hamzah et al. (2023) identify six elements of an effective community-based learning framework in Malaysian institutions, emphasising alignment between course learning outcomes, andragogical activities and assessment. Govindaraju (2025) traces the policy's absorption into institutional frameworks and identifies uneven implementation and variable faculty preparedness as the principal remaining barriers, a diagnosis this case illustrates from the inside.

Empirical accounts of Malaysian SULAM remain concentrated in environmental, community-development and stakeholder-relations settings rather than in library or information-service contexts. Hanafiah (2024) reports a service-learning intervention on environmental consciousness in Yan, Kedah, in the same state as the present programme, and Wahi et al. (2024) map SULAM initiatives against quadruple-helix stakeholder relations, finding recurring themes of university-industry collaboration, community empowerment and human capital development. Neither addresses academic library practice, and we located no published SULAM study situated in an academic library.

The evaluation instruments accompanying SULAM activities have converged on a recognisable form. Osman et al. (2026), reporting a SULAM employability module, describe a questionnaire structured as programme effectiveness, knowledge evaluation administered before and after, and overall evaluation: the same three-part architecture as the instrument used here. That convergence matters for this paper, because it means the methodological argument developed in the Discussion applies well beyond a single programme.

The reciprocity criterion also governs how this case should be read. If the visiting UNAIR students were merely an audience for a Malaysian student assignment, the exercise is pedagogically thin whatever its outcome data show. The evidence presented below suggests the exchange was genuinely two-directional, though asymmetrically documented: the participants' gains were measured, the organisers' gains only narrated.

Peer-Led and Student-Designed Information Literacy Instruction

Peer involvement in academic library services is well established, and the evidence for it is favourable. Graewingholt et al. (2023) report that embedding students in research services both nurtures experiential learning for the student staff and contributes to a more approachable service environment for users, and they position the arrangement as a high-impact practice in its own right.

The distinction that matters for the present case is between peer reference and peer teaching. Lowe et al. (2023) observe that although peer teaching is widely implemented across other disciplines, its adoption in academic libraries has concentrated on reference provision rather than classroom instruction, and that establishing a genuine near-peer teaching programme requires deliberate curriculum design combining information literacy content with teaching pedagogy. Their account of programme creation is candid about the difficulty: securing

stakeholder confidence and reaching the point at which a student can credibly teach alone, are both non-trivial.

What occurred in the present case sits a further step along that continuum. The students here were not delivering librarian-authored content under supervision. They were authoring the instructional design: selecting learning objectives, sequencing activities, writing assessment items and setting facilitation protocols. We located no published account of student-authored instructional design in an academic library outreach setting, which is the basis of this paper's originality claim rather than an incidental observation.

The theoretical warrant is Vygotskian. Learners working with a near-peer operate within a narrower zone of proximal development than expert-novice pairs, and the near-peer's own recent struggle with the material is an instructional asset rather than a deficiency (Vygotsky, 1978). Experiential learning theory supplies the complementary claim that learning consolidates as the learner moves from conceptual exposure to active experimentation (Kolb, 1984), while contemporary information literacy frameworks situate searching within a wider repertoire of evaluation and knowledge creation rather than treating it as a procedural skill (Association of College and Research Libraries, 2015).

The counter-risk follows from the same premise. Near-peers may transmit their own misconceptions and may lack the diagnostic frame that would tell them a task has been completed by the wrong route (Lowe et al., 2023). Precisely that failure is documented in the Findings.

Gamification and Activity-Based Orientation

Gamification, defined as the use of game design elements in non-game contexts (Deterding et al., 2011), has been widely adopted in library orientation as a corrective to the passive walkthrough (Broussard, 2012; Walsh, 2014). The activity-based orientation literature is now reasonably substantial and encompasses several formats directly comparable to the present programme. Martin and Martin (2015) describe adapting an information literacy game specifically for outreach events, where time is short and participants are not a captive instructional audience. Johnson and Westbrook (2021) report augmented reality applications, escape kits and scavenger hunts as alternatives to conventional instruction, and Walsh and Spence (2018) document an escape-room orientation for first-year engineering students.

Recent work sustains the enthusiasm while sharpening the caveats. Adewojo and Adefila (2025) find consistent gains in motivation and engagement but identify overemphasis on rewards, design complexity and sustainability as recurring failure modes. Durodolu et al. (2025), surveying information science students and professionals, report broadly favourable perceptions of gamified approaches to library resource exploration alongside limited familiarity with how to design them well. Reviewing gamification in language instruction, Zhang and Hasim (2023) note that reported benefits must be weighed against technical difficulties, short-lived positive effects and the negative influence of gamified competition, a set of caveats that transfers directly to single-session library formats.

Nicholson (2015) draws the distinction that matters most here: between reward-based gamification, which attaches extrinsic incentives to compliance, and meaningful gamification, which builds play around the underlying activity's actual purpose. Four of this programme's

eight activities were gamified (quiz, crossword, memory game, and the timed Info Hunt), and three of the four sit closer to the reward-based pole. This is not a criticism of the design so much as an observation that bears directly on the interpretation of the outcome data.

Cross-Border and South–South Academic Collaboration

Universities across ASEAN increasingly pursue cross-border engagement, and the wider role of universities in such collaboration has been examined (Smith & Smith, 2020). The library and information science literature documenting library-to-library instructional collaboration between Southeast Asian institutions, however, is sparse. Comparative regional LIS scholarship does exist: Sung et al. (2023) analyse caregivers’ perceptions of the public library’s role across Taiwan, Malaysia and Vietnam, and Baba and Abrizah (2018) examine transformation strategies in Malaysian library community engagement. Both are valuable, and neither document instructional collaboration between academic libraries across a national border, which is the configuration at issue here.

Where such engagement is reported, the dominant paradigm is North-South. Smith and Smith (2020) characterise the long-term role of universities in cross-border development largely in terms of institutional capacity-building, a framing that presumes asymmetry between partners in resources, funding and expertise.

The present case departs from that pattern. It is a South-South exchange between two institutions with comparable resource profiles and mutually intelligible languages, a configuration that removes several barriers common in the literature (translation cost, resource asymmetry, unequal partnership status) and introduces others. Without the funding, memorandum-level obligations and institutional oversight structures that typically accompany North-South partnerships, documentation, evaluation and ethical governance must be constructed ad hoc by the participating unit. The methodological consequences of that absence are visible throughout the Method section of this paper and are not incidental to it.

The Measurement Problem in Outreach Evaluation

The methodological weakness of outreach and gamified-instruction evaluation is well documented and rarely remedied. Systematic and scoping reviews in this area converge on the same diagnosis: engagement and satisfaction are reported in abundance, while reliability evidence, effect sizes and demonstrated learning outcomes are not, and reviewers repeatedly call for more rigorous evaluation before educational value can be claimed (Adewojo & Adefila, 2025; Zhang & Hasim, 2023). Most consequentially, self-reported confidence is treated as though it were demonstrated capability. Learners routinely rate their information-search ability above what they can demonstrate, and the discrepancy is largest among the least proficient; recent work continues to document students who report adequate search confidence while struggling with catalogue and discovery interfaces in practice (Mwiiyale et al., 2025). Any evaluation resting solely on self-report is therefore measuring something real but measuring it indirectly, and the distance between the measure and the capability is not fixed.

The activity-based orientation literature is not uniformly guilty of this. Walsh and Spence (2018) are explicit that although voluntary participation in their escape-room orientation exceeded all expectations, there is no expectation that participants emerged as expert or even intermediate researchers; they left with foundational knowledge on which further instruction

could be built. That is the appropriate register for a single-session intervention, and it is the register this paper attempts to sustain.

Two specific issues bear on the present study. First, the single-group pretest-posttest design is pre-experimental, not quasi-experimental, in Campbell and Stanley's (1963) taxonomy; it lacks any comparison against which history, maturation or testing effects can be discounted. Whatever the quality of the measurement, this design cannot isolate the intervention as the cause of any observed change.

Second, self-report competence measures administered before and after a short intervention are subject to response shift: the intervention alters the participant's internal standard of measurement, so that the two ratings are not made on the same scale (Drennan & Hyde, 2008; Howard, 1980; Howard & Dailey, 1979; Sibthorp et al., 2007). The direction of the resulting distortion depends on the learner's starting position. Novices asked to rate competence in a domain whose boundaries they cannot yet see tend to over-rate themselves at baseline; once instructed, they rate that same prior state lower (Howard, 1980; Howard & Dailey, 1979). The underlying discrepancy between reported and demonstrated ability is itself well attested in information literacy research (Mwiiyale et al., 2025). Where this occurs, a conventional pretest-posttest comparison understates the gain. This is the argument for retrospective, post-then-pre measurement, and it cuts in favour of the present study's estimate being conservative rather than inflated.

Working against that are the ordinary threats to any immediate posttest: social desirability, demand characteristics, and, in this instance, the fact that baseline and outcome ratings shared a single physical instrument, so that participants could see their own earlier responses when making the later ones. The Discussion weighs these against each other.

Methodology

Design and Researcher Position

This is a single-site descriptive case study drawing on two data sources: documentary analysis of the programme archive, and secondary analysis of the programme's community-impact questionnaire.

The programme archive comprises the organisers' final report and its appendices: the organisational chart and role specifications, minutes of the planning meetings, activity-by-activity descriptions with timings and materials lists, a benefits narrative, a challenge log, a recommendations section, and the completed evaluation forms. The archive was produced by the student organisers as an assessed course deliverable, not as a research instrument. This is a limitation and also, in a specific respect, a strength: it was written for a supervising lecturer rather than for peer reviewers, and it is correspondingly frank.

The author team includes one lecturer who supervised the course and monitored the programme's implementation, and who compiled the programme report. The team member contributed the documentary record and contextual detail but took no part in the statistical analysis, which was conducted independently by co-authors who had no role in the programme's design or delivery. The analysis itself is retrospective and archival. This division is stated because insider access and analytic independence pull in opposite directions here: the

former made the archive and its candour available, while the latter is intended to limit the risk that those closest to the programme also judged its success.

Setting and Partners

The host was Perpustakaan Sultan Badlishah (PSB), the main library of UiTM Cawangan Kedah at the Sungai Petani campus, a three-floor academic library serving a branch-campus population, together with the campus Experimental Library (ELIB) facility located in a separate building. The library's discovery layer is PERMATA, the UiTM library system's unified search interface, which offers differentiated resource access to members and non-members, a distinction the organisers made explicit in the orientation, since the visiting cohort held no borrowing rights.

The visiting partner was the Faculty of Vocational Studies, Universitas Airlangga, Surabaya. Participants were drawn from three diploma-level study programmes, Teknik Informatika (n = 8), Teknologi Rekayasa Instrumentasi dan Kontrol (n = 10) and Kearsipan dan Informasi Digital (n = 10).

Participants

Twenty-eight UNAIR undergraduates participated and all twenty-eight returned the evaluation form. Thirty UiTM diploma students served as organisers and facilitators, giving a facilitator-to-participant ratio of approximately 1:1, an unusually rich staffing level that should be borne in mind when assessing transferability.

Participants were predominantly in their first two years of study and were characterised by the organisers as requiring guidance in information search and retrieval. Age and gender data were collected on the evaluation forms but were not tabulated in the programme report.

The Intervention

The programme ran from morning to late afternoon on a single day. Participants were split into two cohorts which rotated between two venues at midday; within each venue, participants worked in three facilitated small groups. Table 1 specifies each component.

Table 1: Programme Structure and Instructional Function

No.	Activity	Venue	Duration	Grouping	Intended instructional function
1	Discovery-system briefing (PERMATA)	PSB	+5 min	Whole cohort	Interface orientation; member vs. non-member access rights
2	Orientation tour (floors 3->2->1)	PSB	+20 min	3 groups	Spatial mapping of zones, collections, service points
3	Info Hunt (search-to-shelf)	PSB, floor 2	10–15 min	3 groups, 8–10 items per group	Procedural enactment: query -> call number -> shelf retrieval

4	Info Sharing	PSB, floor 3	+10 min reading + discussion	3 groups	Peer explanation; consolidation; cross- cultural talk
5	Interactive quiz (10 items)	PSB	+15 min	Whole group, hands-up	Retrieval practice on tour and PERMATA content
6	Handcraft bookmark	ELIB	40 min	Individual	Affective anchor; memento
7	Crossword	ELIB	30 min	3 groups	Vocabulary and service-terminology recognition
8	Book Talk	ELIB	30 min	Whole group	Reading advocacy; oral sharing
9	Memory card game	ELIB	30 min	3 groups, 12 cards / 6 pairs	Recognition; collaborative recall

Note. Component 1 is nested within the PSB track; the eight activities named in the programme documentation are items 2–9. Each ELIB activity ran twice (morning and afternoon session) to accommodate the cohort rotation.

Materials were minimal and locally sourced: blank wooden bookmark blanks (12 x 3.2 cm), tassels, marker pens and institutional logo stickers for component 6; printed crossword sets, whiteboard markers and tissues for component 7; twelve-card printed sets for component 9; and confectionery as quiz incentives. The programme was funded by a committee levy of RM10 per organiser (approximately RM300 total) supplemented by institutional academic development funds.

Two design decisions deserve comment. The Info Hunt was the only component requiring participants to enact a search process rather than observe or recall one; it is, on the programme's own logic, the load-bearing activity. And the rotation structure meant that no participant experienced the activities in a single coherent sequence, half encountered the reinforcement games before the instruction they were meant to reinforce.

Instrument

A single two-page questionnaire, Survey for Evaluation of Impact on Community, was used. It comprised:

- Section A: Programme evaluation (6 items). Five-point satisfaction scale (5 = Very Satisfied ... 1 = Very Dissatisfied). Items covered objective suitability, content, speaker proficiency, facilitator efficiency, programme effectiveness and overall evaluation.
- Section B: Competency towards knowledge (6 items × 2 columns). Five-point agreement scale (5 = Strongly Agree ... 1 = Strongly Disagree), with adjacent "Before Program" and "After Program" rating columns for each item. Items covered interest, knowledge held, application in daily life, sharing with others, teaching others, and continuous dissemination to the community.
- A free-text comments box.

The "Before Program" column of Section B was completed at the start of the day, prior to any instructional component. The form was collected after the baseline rating and redistributed at the close. The "After Program" column and the whole of Section A were completed at the close

of the programme, on the same physical instrument. Section B is therefore a genuine pretest–posttest measure in timing, though not a blind one: participants could see their morning ratings when making their afternoon ones.

The Section B items are not an information-literacy scale in any technical sense. They are a generic community-engagement impact template: interest, knowledge, application, sharing, teaching, and dissemination. Item 1 measures interest, an affective disposition, not competence. No content validity evidence, expert review, pilot testing, translation procedure or dimensionality analysis was undertaken. The instrument’s language of administration is not recorded.

Transcription and Analysis

The programme report summarised the questionnaire results as aggregate frequency counts only. For the present study all 28 original response forms were located in the archive and re-transcribed at item level, yielding a participant-by-item dataset. Transcription was verified against the frequency tables published in the programme documentation.

That verification identified nine cells across three tables where the original hand-tally diverged from the forms. Five of the six baseline competence items differed by one or two responses in adjacent categories, and the tabulated distribution for baseline item 1 summed to 25 rather than 26; re-inspection confirmed that all 26 participants with baseline data had responded to that item and the original tally had undercounted by one. Three of six satisfaction items differed by a single response, as did one outcome competence item. The programme narrative additionally reported one “disagree” response on baseline item 6 where its own table recorded two. Five of six outcome competence items reproduced exactly. All analyses reported here use values read directly from the original forms. The discrepancy pattern is consistent with hand-tallying error in the original documentation rather than with transcription error: the outcome column, comprising almost entirely ratings of 4 and 5, was tallied nearly perfectly, whereas the baseline column, spread across four response categories, was not.

Twenty-eight participants completed the satisfaction section and the outcome competence ratings; 26 also provided baseline ratings. Satisfaction descriptives and reliability use $N = 28$. Paired competence analyses and the satisfaction–gain correlation use the 26 complete cases. The two participants without baseline data are retained in outcome-only descriptives and are compared against the analytic sample in Findings section.

Internal consistency was assessed by Cronbach’s alpha with corrected item–total correlations and alpha-if-item-deleted. Composite indices are the unweighted mean of the six items (range 1–5). The primary outcome is the within-participant change score; normality of the difference distribution was checked by Shapiro–Wilk before applying a paired-samples t-test, with a Wilcoxon signed-rank test reported alongside as an ordinal robustness check, since Likert responses are ordinal and the appropriateness of parametric treatment is contested (Norman, 2010). The mean difference and its 95 per cent confidence interval are treated as the primary effect estimate; d_z and d_{av} are reported for comparability but interpreted cautiously, since compressed outcome variance inflates standardised within-subject effect sizes. Item-level tests are reported with Holm correction as a secondary analysis. Analyses were conducted in Python 3 (NumPy, SciPy). $\alpha = .05$, two-tailed.

Findings

Reliability of the Measures

All three scales showed strong internal consistency (Table 2). No item deletion improved alpha on any scale, and corrected item–total correlations were uniformly acceptable, supporting treatment of each six-item set as a single composite.

Table 2: Internal Consistency

Scale	<i>n</i>	Cronbach's α	Corrected item–total <i>r</i>
Satisfaction	28	.901	.62 – .83
Competence, baseline	26	.916	.74 – .83
Competence, outcome	26	.865	.57 – .71

The baseline competence scale is the strongest of the three ($\alpha = .916$). This matters, because the instrument was an inherited engagement template with no prior validity evidence; the reliability evidence does not establish that it measures information-search competence, but it does establish that the six items cohere.

Programme Acceptability

All 28 participants completed the satisfaction section. Ratings were uniformly high and tightly clustered (Table 3), composite $M = 4.70$ (SD 0.42, range 3.50–5.00).

Table 3: Programme Evaluation Ratings (N = 28)

Item	5	4	3	M	SD	% at 4–5
S1 Suitability of objectives	20	7	1	4.68	0.55	96.4
S2 Content of programme	19	9	0	4.68	0.48	100.0
S3 Speaker proficiency	22	5	1	4.75	0.52	96.4
S4 Facilitator efficiency	21	5	2	4.68	0.61	92.9
S5 Programme effectiveness	20	8	0	4.71	0.46	100.0
S6 Overall evaluation	19	9	0	4.68	0.48	100.0
Composite				4.70	0.42	

Note. No participant selected “Dissatisfied” or “Very Dissatisfied” on any item.

The item means span 0.07 scale points. Participants did not discriminate between objectives, content, facilitation and overall quality, a flatness more consistent with global affect than with differentiated judgement.

Change in Perceived Competence

Baseline ratings were collected before any instruction; outcome ratings at the close of the day. Composite competence rose from $M = 3.47$ (SD 0.68) to $M = 4.72$ (SD 0.35), a mean within-participant gain of 1.26 points on a five-point scale, 95% CI [0.99, 1.52].

Difference scores were approximately normally distributed (Shapiro–Wilk $W = .961$, $p = .42$), supporting the paired *t*-test as the primary analysis: $t(25) = 9.83$, $p < .001$. The Wilcoxon signed-rank test converged, $W = 0.0$, $p < .001$. A value of zero means that every participant whose score changed, changed upward. Not one decline was recorded.

Standardised within-subject effect sizes were $d_z = 1.93$ and $d_{av} = 2.44$. Both are inflated by the compression of outcome variance (SD 0.35) combined with a modest pre–post correlation ($r = .34$), and neither should be compared with between-group benchmarks. The confidence interval on the raw mean difference is the more interpretable estimate. Item-level results are reported in Table 4.

Table 4: Item-Level Pre–Post Comparison (Paired, N = 26)

Item	Pre (SD)	M	Post (SD)	M	Δ	$t(25)$	p_{Holm}
C1 Interested in the knowledge/skill shared	3.54 (0.76)		4.69 (0.47)		+1.15	8.72	< .001
C2 Have knowledge/skill regarding information shared	3.31 (0.84)		4.81 (0.40)		+1.50	8.89	< .001
C3 Can apply the knowledge/skill in daily life	3.38 (0.85)		4.65 (0.49)		+1.27	7.82	< .001
C4 Can share the knowledge/skill with others	3.54 (0.81)		4.73 (0.45)		+1.19	6.79	< .001
C5 Can teach the knowledge/skill to others	3.58 (0.81)		4.85 (0.37)		+1.27	8.32	< .001
C6 Can disseminate to the community continuously	3.46 (0.81)		4.62 (0.50)		+1.15	6.68	< .001
Composite	3.47 (0.68)		4.72 (0.35)		+1.26	9.83	< .001

Note. Wilcoxon $W = 0$ for every item. All p values Holm-corrected across the six item-level tests.

Four features of the distribution are worth naming.

The baseline was low and honest. Only 41.0 per cent of baseline item-responses reached “Agree” or above; every item except C5 attracted at least two responses at “Disagree”. Participants were willing, at the start of a day hosted for their benefit, to record that they did not yet have the skills on offer. That is a credible novice baseline, and it is not what an instrument eliciting compliant positivity produces.

Movement was near-universal. Individual gains ranged from 0.00 to +2.50. Twenty-four of 26 participants improved; the two who did not had already recorded 5.00 on all six baseline items and had no room to move. Excluding those two ceiling cases, the mean gain rises to 1.36.

Dispersion halved. Baseline SD 0.68 fell to 0.35, and outcome endorsement reached 100 per cent of item responses. Convergence of this kind is consistent with genuine cohort-wide learning; it is also consistent with a shared response norm, and the Discussion weighs the alternatives.

The item gradient is coherent. Largest movement on C2 (having the knowledge, +1.50); smallest on C1 (interest, +1.15) and C6 (continuous dissemination, +1.15). Participants were most willing to claim acquisition and least willing to claim sustained onward transmission. Respondents were not simply marking fives down the column.

Satisfaction Did Not Predict Gain

Because individual-level data were recovered for both instruments, it is possible to test directly whether the reported competence gain was an artefact of generalised goodwill toward the programme. If participants who enjoyed the day most also reported the largest gains, a halo-response account would be difficult to dismiss.

They did not. Satisfaction was uncorrelated with competence gain: $r = .077$, $p = .71$, 95% CI $[-.32, .45]$; Spearman $\rho = .151$, $p = .46$.

The wider correlation pattern sharpens the interpretation (Table 5).

Table 5: Satisfaction and Competence (N = 26)

	<i>r</i>	<i>p</i>
Satisfaction × baseline competence	.201	.326
Satisfaction × outcome competence	.540	.004
Satisfaction × gain	.077	.710

Satisfaction predicted where participants finished but not how far they travelled. This is what one would expect if satisfaction and end-state confidence share a common positivity component while the change is driven by something else. It does not establish that the data are bias-free, a uniform inflation affecting everyone equally would be invisible to this test, but it rules out the most common and most damaging alternative explanation, namely that differential enthusiasm generated differential reported learning.

Given the design's other weaknesses, this is the study's strongest single piece of internal evidence.

Participants Without Baseline Data

Two participants provided outcome but no baseline ratings; the most plausible explanation is arrival after the morning administration, and this should be verified against the attendance record. They did not differ substantively from the analytic sample on either available measure: satisfaction $M = 4.67$ versus 4.70, outcome competence $M = 4.67$ versus 4.72. Their exclusion from paired analyses is therefore unlikely to have biased the estimates.

The Fidelity Failure

The most consequential finding in the programme archive is not in the evaluation data at all. It is in the organisers' own description of the Info Hunt:

Every group had to use the PERMATA system to find books on the given topic, but not every group used the PERMATA search; some UNAIR students simply went to the computer science and information science shelves. (SULAM programme report, Section 6.2, "Info Hunt", paraphrased)

This single sentence, recorded without apparent recognition of its significance, undermines the most important inference the programme might otherwise support. The Info Hunt was the only component requiring enactment of a search process. Its instructional purpose, stated by the organisers themselves, was “to make sure they know how to search for the materials correctly, from searching through PERMATA to finding them on the shelves.” Where groups skipped the discovery system and went directly to the classified shelves, the task was completed, the books were retrieved, the time limit was met, and the target skill was not practised.

Two conditions made this substitution easy and rational. The task specified only two broad subject areas, computer science and information science, whose holdings sit contiguously under a classification scheme; and it was timed at 10–15 minutes for 8–10 items across a group, creating pressure to choose the fastest retrieval route. Browsing was simply the dominant strategy given the task design. The participants were not cheating; they were optimising, and the task permitted it.

The programme thus produced, within a single afternoon, a compact demonstration of the perceived-versus-actual competence gap that information literacy research continues to document (Mwiiyale et al., 2025). Participants rated their ability to search, apply and teach information retrieval far higher after the programme than before. Some of them had, in the one activity that would have exercised that ability, not used the system at all.

Organiser-Side Outcomes

The archive narrates substantial gains for the thirty organisers, communication and public speaking, programme management, problem-solving under time pressure, budgeting and financial accountability, digital production skills (Canva, CapCut, video editing, social media management), and negotiation with institutional units. These claims are plausible and consistent with the service-learning literature (Eyler & Giles, 1999).

They are also entirely unmeasured. The programme instrumented its beneficiaries and left its designers to self-narrate. Given that SULAM’s pedagogical rationale rests principally on what the students learn, this is the evaluation design’s largest structural omission, larger, arguably, than the absence of a control group.

Discussion

What This Study Can and Cannot Claim

It is worth stating plainly what the design permits.

The programme was highly acceptable to participants: this is well supported, though acceptability was measured under conditions that would have made a negative result nearly impossible to obtain. Participants reported large increases in perceived competence between a pre-instruction baseline and the close of the day: this is what the data show, subject to the qualifications below. The programme caused competence gains: this is not supported, and the design cannot support it. Participants’ actual information-search capability improved: this is not supported, was not measured, and is placed in active doubt by the fidelity failure reported below.

The distance between the second and fourth of those statements is the substance of this paper.

How Much Weight Can the Change Bear?

Three considerations pull in different directions and are best set out separately.

In favour of the estimate. The baseline was collected before any instruction, which rules out the reconstruction problem that afflicts retrospective (post-then-pre) measures, participants were not asked to remember a prior self, they recorded it. The baseline is also unflattering, with the majority of responses below “Agree”, which is not what an instrument eliciting compliant positivity produces. And response shift, where present, works against the observed effect: novices who over-rate themselves at baseline on a scale they do not yet understand generate a pretest that is too high, compressing the apparent gain (Howard, 1980; Sibthorp et al., 2007). On this argument the reported movement is a floor rather than a ceiling.

Against it. The posttest was completed at the close of a hospitable day, on a named form, by international guests, distributed by the people who had hosted them, every structural condition for social desirability. The baseline was visible on the same sheet, so participants could see the direction of change they were producing. Post-programme dispersion approximately halved and 99.4 per cent of responses landed at “Agree” or above; genuine cohort-wide learning produces convergence, but so does a shared reading of what the form wants. And the single-group design admits no way to separate the intervention from a day spent in a novel environment among new acquaintances.

What the halo test adds. The satisfaction–gain correlation reported in the Findings settles one of these disputes empirically. Had enthusiasm been driving the reported gains, the participants who rated the day most highly would have reported the largest movement. They did not ($r = .08$, $p = .71$), even though satisfaction did predict outcome scores ($r = .54$, $p = .004$). Differential social desirability is therefore an unlikely explanation for the observed change. A uniform inflation affecting every participant equally would remain undetectable by this test, and no self-report design can exclude it, but the specific mechanism most often invoked against evaluations of this kind does not fit these data.

What survives. The construct actually measured is perceived competence, not competence. That distinction is not a technicality here, because the fidelity failure reported above records participants completing the one search task without searching. The defensible claim is that the programme substantially raised participants’ self-efficacy with respect to academic information retrieval, that is, their belief that they can locate, apply and pass on what they encountered. Self-efficacy is a consequential outcome in its own right: it predicts persistence, willingness to attempt difficult searches, and future engagement with unfamiliar systems (Bandura, 1977, 1997). A participant who leaves believing she can navigate an academic discovery layer is more likely to try one.

That claim is worth making and this design supports it. It is not the claim that she can.

What The Fidelity Failure Teaches

The Info Hunt substitution reported in the Findings deserves to be the finding practitioners take away, and it generalises well beyond this programme.

The design error is specific and correctable. A retrieval task that can be completed faster by browsing than by searching will be completed by browsing, and no amount of instruction beforehand changes that, because the task itself rewards the wrong behaviour. The fix is not more explanation but better task construction: require items whose classification numbers are non-contiguous; require a bibliographic detail obtainable only from the record (publication year, edition, availability status, a specific location code); require participants to log the search string used; distribute a partial citation rather than a subject. Any of these makes the discovery system the shortest path rather than an obstacle to it.

There is also a facilitation lesson, and it has a name in the literature. Adeyemi et al. (2025) argue that the effective delivery of gamified library services depends on facilitators' own game literacy, meaning their capacity to read what a game-structured activity is doing to participants' behaviour. Near-peer facilitators at approximately 1:1 ratio observed the substitution and recorded it faithfully but did not intervene during the activity or treat it as a problem afterwards. They saw it; they did not read it as failure. This is precisely the diagnostic gap anticipated in the literature review: near-peer facilitators possess the content but not yet the assessment frame that would tell them a completed task can be a failed one. A short facilitator briefing on what success and failure look like behaviourally, not on the content but on the observable process, would likely have caught it.

Students as Instructional Designers

Notwithstanding the above, the delegation of instructional design to students produced things a librarian-designed session probably would not have.

The programme was unusually well-differentiated in modality: spatial, procedural, discursive, competitive, creative and recall-based components in a single day, with an affective anchor (the handcraft bookmark) that participants took home. The register was near peer throughout, and the Info Sharing component was designed to be genuinely bidirectional, Malaysian and Indonesian students discussed books together rather than one group instructing the other. The archive records a moment of real reciprocity, when UNAIR participants drew Malaysian and Indonesian cultural motifs on their bookmarks, that no instructional objective specified and no rubric could have produced.

Against this, the sequencing was compromised by the rotation structure (half the cohort met the reinforcement games before the instruction), the load-bearing activity was under-designed, and the assessment instrument was inherited rather than constructed.

The pattern is consistent: students designed engagement extremely well and designed measurement poorly. That is a predictable division, and it suggests where supervisory input adds most value. It is not the activity design, which was strong, but task construction and assessment.

Gamification as Scaffolding

Three of the four gamified components (quiz, crossword, memory game) operated on vocabulary and recall rather than on search behaviour, and used extrinsic rewards and penalties, sweets for correct answers, a singing forfeit for the slowest group. In Nicholson's (2015) terms this is reward-based rather than meaningful gamification: play attached to the activity rather

than built from it, precisely the overemphasis on extrinsic reward that Adewojo and Adefila (2025) identify as the most common design failure in library gamification.

That is not a fatal criticism of a one-day orientation, where sustaining attention across seven hours is a real constraint and these components plainly did so. But it does bound the interpretation. The large, reported gains cannot reasonably be attributed to the gamified components, which did not exercise the target skill. If anything drove real learning it was components 1–4; and component 3, as shown, was only partially executed.

Methodological Note

Three features of the present analysis respond directly to the critiques of outreach evaluation set out in the literature review and are cheap enough to be adopted generally. Individual-level responses were recovered rather than accepting the aggregate tally, which made reliability estimation and paired inference possible at the cost of a few hours' transcription. The transcription was validated against the published marginals, which surfaced nine tallying errors in the original documentation that would otherwise have propagated into print. And the satisfaction–gain correlation converted an ordinarily untestable objection, that positive affect manufactured the result, into an empirical question with an answer.

None of this required a better design. It required only that the data already collected be used fully.

The wider point is that the instrument used here is not idiosyncratic. Osman et al. (2026) report a 2026 SULAM evaluation built on the same three-part architecture, and the template is in routine use across Malaysian higher education. Every programme using it generates individual-level responses that are typically summarised into frequency counts and then discarded. The analytic gain demonstrated in this paper is therefore available, at essentially no cost, to a large number of existing evaluations.

Limitations

Beyond those already stated: the study is a single site with a single cohort and no comparison group; the sample is small ($N = 28$); the outcome instrument is a generic engagement template whose items cohere reliably ($\alpha = .87-.92$) but for which no content-validity or dimensionality evidence exists, and whose language of administration is unrecorded; baseline and outcome ratings shared one physical form, so the baseline was visible at posttest; measurement was immediate, so nothing is known about durability; participants were guests of the evaluators; and organiser-side outcomes, central to the SULAM rationale, were not measured at all.

Perhaps the most important limitation is one of standpoint. The archive was written by the organisers, for assessment, in a genre that rewards demonstrating success. Its candour is notable precisely because the genre does not require it, but that candour is uneven, and it is unlikely that everything which went wrong was recorded.

Implications for Practice

1. Design retrieval tasks that make the system the shortest path. Non-contiguous classifications, record-only data points, logged search strings.
2. Brief facilitators on observable failure, not just content. Tell near-peer facilitators what a wrongly-completed task looks like and instruct them to intervene.
3. Instrument the organisers. Where service-learning's rationale rests on student learning, measure the students.
4. Separate acceptability from learning in reporting. Report satisfaction as satisfaction; do not let it stand in for competence.
5. Retain individual-level evaluation data in analysable form. Aggregate counts foreclose most useful analysis at negligible saving.

Future Research

The immediate priority is measurement of behaviour rather than belief: timed retrieval tasks, search-log capture, or a scored search-strategy artefact, administered alongside self-report so that the perceived–actual gap can be quantified rather than inferred. A comparison condition, an equivalent cohort receiving a conventional tour without the enactment component, would allow attribution. Organiser-side gains should be measured with a validated instrument. And delayed follow-up, ideally with the partner institution's cooperation at three and six months, would establish whether anything persisted.

A replication with the design improvements set out under Implications for practice would be more valuable than further analysis of this dataset, which has now been exhausted.

Conclusion

A student-led international library outreach programme achieved near-universal acceptability among its twenty-eight visiting participants and produced a large, statistically robust gain in perceived information-search competence between a pre-instruction baseline and the close of the day (+1.26 points, 95% CI [0.99, 1.52]), on scales with strong internal consistency, with every participant who moved moving upward. Satisfaction did not predict that gain, which counts against the most obvious alternative explanation. Read at face value, this is a success story, and the measurement is better than the genre usually manages.

Read more carefully, it is something more useful. The measure remains self-reported confidence, collected from guests on a named form at the end of a hospitable day; it indexes self-efficacy, not capability. And the programme's own records show that in the single activity designed to exercise information search, some participants bypassed the discovery system and browsed the shelves, completing the task, satisfying the facilitators, and not practising the skill.

The value of this case therefore lies less in the outcome data than in three things it makes visible: a fully specified, cheaply resourced, replicable programme blueprint; a demonstration that two inexpensive analytic steps allow a routine evaluation to test its own most obvious objection; and a compact, unplanned account of the gap between what learners believe they have acquired and what they have actually done.

Academic libraries wishing to move beyond descriptive outreach reporting need not begin with sophisticated designs. They can begin by constructing tasks that cannot be completed the wrong way, by briefing facilitators to notice when they are, and by declining to let a satisfaction score stand in for evidence of learning.

-
- Acknowledgements:** The authors thank the final-semester students of the Diploma in Library Informatics (IML301), UiTM Kedah Branch, who designed, facilitated and evaluated the outreach programme reported here, and whose programme archive made this analysis possible. We are grateful to the Research Management Unit (RMU), UiTM Cawangan Kedah, for its institutional support of this work.
- Funding Statement:** This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. The outreach programme itself was financed by a levy of RM10 per organising committee member, supplemented by institutional academic development funds; no funding was received for the analysis reported in this article.
- Conflict of Interest Statement:** Nordibradini Selamat was the lecturer in charge of the SULAM programme evaluated in this article. She supervised the student organisers, monitored implementation and compiled the programme report. She contributed the documentary record and contextual detail to this manuscript but took no part in the statistical analysis, which was conducted independently by Anaqi Ahmad Saleh and Sharunizam Shari, neither of whom had any role in the programme's design or delivery. The authors declare no other conflict of interest.
- Ethics Statement:** The programme was an assessed course activity, and its evaluation was routine institutional programme assessment rather than research at the point of collection. Participants were informed at the time of collection that the data would be used for publication purposes, and completion of the questionnaire was voluntary. Consent was obtained verbally and no signed consent forms were retained, this being standard practice for programme evaluation of this kind. The participants were adult undergraduates; none was a minor, and none belonged to a category conventionally regarded as vulnerable in research ethics terms. Personal identifiers recorded on the original forms served only to link each respondent's pre- and post-programme responses and were not transcribed into the analytic dataset; no individual is identifiable in this article.
- Author Contribution Statement:** Sharunizam Shari conceptualised the study, designed the methodology, conducted the formal analysis, prepared the original draft and served as corresponding author. Anaqi Ahmad Saleh carried out the formal analysis, developed the analysis software and validated the results. Nurdibradini Selamat supervised the outreach
-

programme, conducted the investigation, provided the institutional resources on which the study draws, and compiled the programme documentation; Nurdibradini Selamat additionally handled project administration. Nor Azlina Azmi and Salasiah M Said undertook data curation and contributed to the original draft. All authors contributed to reviewing and editing the manuscript, and all have read and approved the final version. All authors have read and approved the final manuscript.

References

- Adewojo, A., & Adefila, E. (2025). Game-changing libraries: Enhancing information literacy through gamification. *Business Information Review*, 42(2), 106–111. <https://doi.org/10.1177/02663821251328838>
- Adeyemi, I. O., Akanbi, M. L., & Issa, A. O. (2025). Influence of game literacy on gamified library services: A systematic review of literature. *Alexandria: The Journal of National and International Library and Information Issues*. Advance online publication. <https://doi.org/10.1177/09557490251335953>
- Association of College and Research Libraries. (2015). *Framework for information literacy for higher education*. American Library Association. <https://www.ala.org/acrl/standards/ilframework>
- Baba, Z., & Abrizah, A. (2018). Transformation strategies in community engagement: Selected initiatives by Malaysian libraries. *IFLA Journal*, 44(2), 90–105. <https://doi.org/10.1177/0340035218778435>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
- Bringle, R. G., & Hatcher, J. A. (1995). A service-learning curriculum for faculty. *Michigan Journal of Community Service Learning*, 2(1), 112–122.
- Broussard, M. J. S. (2012). Digital games in academic libraries: A review of games and suggested best practices. *Reference Services Review*, 40(1), 75–89. <https://doi.org/10.1108/00907321211203649>
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining “gamification.” In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning future media environments* (pp. 9–15). Association for Computing Machinery. <https://doi.org/10.1145/2181037.2181040>
- Drennan, J., & Hyde, A. (2008). Controlling response shift bias: The use of the retrospective pre-test design in the evaluation of a master’s programme. *Assessment & Evaluation in Higher Education*, 33(6), 699–709. <https://doi.org/10.1080/02602930701773026>
- Durodolu, O. O., Ojo, J. O., & Ajayi, T. B. (2025). Examining the characteristics and perceptions of gamification strategies for enhancing library resource exploration among information science students and professionals. *Library Management*, 46(3/4), 286–300. <https://doi.org/10.1108/LM-10-2024-0108>
- Eyler, J., & Giles, D. E., Jr. (1999). *Where’s the learning in service-learning?* Jossey-Bass.

- Furco, A. (1996). Service-learning: A balanced approach to experiential education. In B. Taylor and Corporation for National Service (Eds.), *Expanding boundaries: Serving and learning* (pp. 2–6). Corporation for National Service.
- Govindaraju, V. (2025). Bridging academia and society: Institutionalizing Service-Learning Malaysia (SULAM) within Malaysian higher education frameworks. *International Journal of Research and Innovation in Social Science*, 9(8), 7461–7479. <https://doi.org/10.47772/IJRISS.2025.908000619>
- Graewingholt, M., Cornforth, J., & Parramore, S. (2023). Better together: Experiential learning in peer-led research services. *Reference Services Review*, 51(2), 151–170. <https://doi.org/10.1108/RSR-09-2022-0043>
- Hamzah, N. F. A., Tajuddin, A. J. A., Romly, R., Wahi, W., & Aris, S. R. S. (2023). From classroom to community: Understanding community-based learning practices in Malaysian higher education institutions. *International Journal of Learning, Teaching and Educational Research*, 22(3), 33–50. <https://doi.org/10.26803/ijlter.22.3.3>
- Hanafiah, N. A. H. (2024). Integrating service learning in higher education for enhanced environmental consciousness in Yan, Kedah. *Journal of Governance and Development*, 20(2), 195–216. <https://doi.org/10.32890/jgd2024.20.2.2>
- Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review*, 4(1), 93–106. <https://doi.org/10.1177/0193841X8000400105>
- Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, 64(2), 144–150. <https://doi.org/10.1037/0021-9010.64.2.144>
- Jamhari, M. R., Nordin, M. S., Aripin, M. A., & Ismail, I. H. (2024). Implementation of service learning in Malaysian higher-educational institutions: A literature review. *Journal of Workforce Education & Research*, 1(1), Article a0000518. <https://doi.org/10.36877/jwer.a0000518>
- Johnson, K. L., & Westbrook, K. (2021). Quelling the boredom with alternative instruction: Augmented reality, escape kits, and scavenger hunts. *International Journal on Social and Education Sciences*, 3(1), 180–196. <https://doi.org/10.46328/ijonses.65>
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall.
- Kuh, G. D. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter*. Association of American Colleges and Universities.
- Lowe, M. S., Stone, S. M., & Macy, K. (2023). Peer teachers taking the lead in classroom instruction: Program creation and challenges faced. *Reference Services Review*, 51(2), 171–189. <https://doi.org/10.1108/RSR-09-2022-0041>
- Martin, L., & Martin, W. (2015). Modifying an information literacy game for outreach events. *Reference Services Review*, 43(4), 643–655. <https://doi.org/10.1108/RSR-02-2015-0009>
- Mwiiyale, J. M., Hamutumwa, N. U., & Shatona, M. N. (2025). Exploring students’ views on information literacy skills training in higher education. *The Reference Librarian*, 66(1–2), 30–47. <https://doi.org/10.1080/02763877.2025.2481088>
- Nicholson, S. (2015). A RECIPE for meaningful gamification. In T. Reiners & L. C. Wood (Eds.), *Gamification in education and business* (pp. 1–20). Springer. https://doi.org/10.1007/978-3-319-10208-5_1
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15(5), 625–632. <https://doi.org/10.1007/s10459-010-9222-y>

- Osman, I., & Jahya, A. (2026). An educational approach to employability competencies through SULAM: Evaluating talent acquisition module for job searching skills among madrasah students. *International Journal of Research and Innovation in Social Science*, 10(26), 2946–2960. <https://doi.org/10.47772/IJRISS.2026.1026EDU0238>
- Sibthorp, J., Paisley, K., Gookin, J., & Ward, P. (2007). Addressing response-shift bias: Retrospective pretests in recreation research and evaluation. *Journal of Leisure Research*, 39(2), 295–315. <https://doi.org/10.1080/00222216.2007.11950109>
- Smith, R. B., & Smith, N. N. (2020). Cross-border development: A long-term role for universities. *ASEAN Journal of Community Engagement*, 4(1), 235–258. <https://doi.org/10.7454/ajce.v4i1.1077>
- Sung, H.-Y., Abdullah, N., & Ngo, H. T. (2023). Early literacy and family engagement: A cross-country analysis of caregivers' perceptions of the public library's role in the digital age in Taiwan, Malaysia, and Vietnam. *Information Development*, 39(3), 480–499. <https://doi.org/10.1177/02666669221143746>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wahi, R., Syed Aris, S. R., Mohamed, A. M. D., Zulkipli, Z. A., Awang Hashim, R., & Babiera, R. I. M. (2024). Exploring the nexus of service learning initiatives and quadruple helix stakeholders in Malaysia. *Environment-Behaviour Proceedings Journal*, 9(29), 21–27. <https://doi.org/10.21834/e-bpj.v9i29.6018>
- Walsh, A. (2014). The potential for using gamification in academic libraries in order to increase student engagement and achievement. *Nordic Journal of Information Literacy in Higher Education*, 6(1), 39–51. <https://doi.org/10.15845/noril.v6i1.214>
- Walsh, B., & Spence, M. (2018). Leveraging escape room popularity to provide first-year students with an introduction to engineering information. *Proceedings of the Canadian Engineering Education Association*. <https://doi.org/10.24908/pceea.v0i0.13054>
- Zhang, S., & Hasim, Z. (2023). Gamification in EFL/ESL instruction: A systematic review of empirical research. *Frontiers in Psychology*, 13, Article 1030790. <https://doi.org/10.3389/fpsyg.2022.1030790>