Measuring What Matters: Goal-Free Evaluation Of Holistic Assessment In Arabic Language Education

Received 2025-06-23
Accepted 2025-10-20
Published 2025-12-27

Saproni Muhammad Samin*[1], Azhar Jaafar[2], Harif Supriady[3], Rojja Pebrian[4], Alfitri Zulkifli[5], Yenni Yunita[6]
[1,3,4,5]Department of Arabic Language Education, Universitas Islam Riau, Indonesia, [2]Civilization Studies, UCYP University, Malaysia, [6]Department of Early Chilhood Islamic Education, Universitas Islam Riau, Indonesia
safroni.ahmad@edu.uir.ac.id*[1], azhar@ucyp.edu.my[2], harifsupriady@fis.uir.ac.id[3], rojjapebrian@fis.uir.ac.id[4], alfitri2018@fis.uir.ac.id[5], yenniyunita@fis.uir.ac.id[6]

To cite this article: Samin, Saproni Muhammad., Jaafar, Azhar. Supriady, Harif., Pebrian, Rojja., Zulkifli, Alfitri., Yunita, Yenni. (2026). Measuring What Matters: Goal Free Evaluation Of Holistic Assessment In Arabic Language Education. Ijaz Arabi: Journal of Arabic Learning, 9 (1), 278-290, DOI: https://doi.org/10.18860 /ijazarabi.V9i1.36779

Abstract

Outcome-Based Education (OBE) requires assessment that captures not only knowledge but also professional dispositions and observable performance. Program-level evidence in Arabic Language Education (ALE) remains limited, especially studies that link holistic assessment design to institutional quality indicators and graduate outcomes. The study aims to appraise the effectiveness of a holistic assessment system (cognitive–affective–psychomotor) in the ALE Study Program at Universitas Islam Riau using Goal Free Evaluation (GFE) and to examine its association with academic attainment and graduate outcomes. This study is a mixed-methods, descriptive dominant design employing GFE to evaluate effectiveness through actual outcomes rather than predetermined targets. Data sources included three academic years (Y2 [2022] to Y [2025]) of institutional records, quality assurance documents (GLOs•CLOs•sub-CLOs; rubrics), micro-teaching artefacts, student satisfaction surveys, and a tracer study. Association with academic attainment was examined through temporal trends in GPA and on-time graduation; association with graduate outcomes through employment rates and time-to-job. Findings: GFE evaluation revealed system effectiveness across all domains: (i) cognitive—mean GPA remained consistently high (3.69–3.78) and on time graduation reached 52% in Y; (ii) affective—"very good" ratings ≥75% across service dimensions; (iii) psychomotor—standardized micro teaching ecosystem (≥4 practices/semester; ≥10 core skills; B threshold). Association with academic attainment: The 40–30–30 design, explicitly linked to GLOs→CLOs, showed a strong association: stable, high GPAs across three years and improved on-time graduation indicate that the system effectively supports quality and efficiency. Association with graduate outcomes: Strong positive association demonstrated—97.5% placement, 4.8-month time-to-job, with education sector dominance (~77%) corresponding to psychomotor emphasis on teaching skills. The integrated competence (cognitive mastery, affective dispositions, psychomotor skills) developed through the holistic assessment system directly contributed to favorable employment outcomes. Conclusion: A 40–30–30 assessment design explicitly linked to GLOs→CLOs effectively sustains performance across three domains and employability. GFE evaluation demonstrated apparent effectiveness and strong associations with both academic attainment and graduate outcomes. Recommendations include rubric

standardization, assessor moderation, analytics dashboards, and longitudinal tracking.
Keywords: Outcome-Based Education; Holistic Assessment; Constructive Alignment; Micro-Teaching; Employability

INTRODUCTION

Outcome-Based Education (OBE) shifts attention from instructional processes to demonstrable learning results. Through constructive alignment, intended outcomes are translated into the competency's graduates display (J. Biggs, 2003; J. B. Biggs & Collis, 2014; Harden, 2007; Munzenmaier & Rubin, 2013; Spady, 1994). In Arabic Language Education (ALE), assessment therefore needs to span cognition, professional disposition, and performative teaching/language ability so that graduate profiles as educators and Arabic practitioners are evidenced in all three domains (Agrawal et al., 2021; Frank & Danoff, 2007; Ghani et al., 2011; Joseph et al., 2024).

The state of the art shows steady progress in OBE across disciplines, yet methodological limitations persist—overreliance on self-report, scarce longitudinal tracking, and uneven standardization (Handrianto et al., 2025; Telukdarie & Munsamy, 2019; Yadav et al., 2024). Comprehensive evaluation frameworks that balance cognitive–affective–psychomotor domains increase the meaningfulness of outcomes (Conradty & Bogner, 2022; Yadav et al., 2024), while standardization and automation improve efficiency and reliability though adoption remains uneven in teacher/language programs (Belyakova & Zakharova, 2020; Lugay et al., 2020). Technology integration—web-based learning, digital pedagogy, and technology-enhanced assessment—improves formative feedback and learning analytics given appropriate governance and equity safeguards (Anderson et al., 2024; Chen et al., 2024; Harden, 2007; Qiao & Fu, 2022; Samin & Osman, 2024; Theodoridou, 2021). In Arabic language education, integrating artificial intelligence and digital platforms into teaching plans enhances learning outcomes and supports self-regulated learning processes (Samin & Hikmah, 2021; Samin & Osman, 2024).

On performative and technology-supported learning, blended learning and multimodal analytics (CNN/MVML) have increased the objectivity of feedback on movement and emotional expression and are relevant to psychomotor–affective assessment (Zhang & Wang, 2024). In station-based assessment, the Multiple Mini Viva (MMV) can holistically capture cognitive–interpersonal–professionalism in both onsite and online formats (Hadie et al., 2022). In addition, the participatory–inclusive principles of the Holistic Assessment Framework (HAF) stress authentic evidence and progress across domains and can be adapted to higher education in language disciplines (Mitra & Wadegaonkar, 2024).

Research on cultural adaptation indicates that OBE effectiveness improves when assessment is aligned with local values and contexts (Conradty & Bogner, 2022; Handrianto et al., 2025; Theodoridou, 2021). In ALE, a key gap concerns integrating Islamic–Malay values into holistic assessment without compromising international standards and accountability (Conradty & Bogner, 2022; Handrianto et al., 2025; Joseph et al., 2024; Muhammad Samin et al., 2025; Samin et al., 2020; Yusri & Muhammad Samin, 2024). Heutagogy approaches, which emphasize student autonomy and self-directed learning, have shown promise in Arabic language education when integrated with culturally responsive assessment frameworks (Samin, 2019). In practice, micro-teaching is a recognized accelerator of teaching skill acquisition, yet

standardization of practice frequency, cross-skill rubrics, and explicit linkage to employability indicators are not consistently documented (Al-Mubaid & Bettayeb, 2017; J. Biggs, 2003; Qiao & Fu, 2022). In Indonesia, program-level evaluations that link holistic assessment design to institutional indicators—GPA, on-time graduation, and tracer outcomes—rarely employ evaluation designs that minimize goal bias (Handrianto et al., 2025; Lugay et al., 2020; Yadav et al., 2024). In parallel, recent assessment syntheses emphasize the complementarity of holistic and analytic scoring: holistic scales capture overall performance, while analytic sub-scores provide targeted diagnostics for strengthening sub-capabilities in thinking and performance (Zlatkin-Troitschanskaia et al., 2019).

This study responds to these gaps in three mutually reinforcing ways. First, we operationalize a balanced holistic assessment model (40% cognitive – 30% affective – 30% psychomotor) explicitly linked to GLOs→CLOs→sub-CLOs in ALE (J. Biggs, 2003; J. B. Biggs & Collis, 2014; Yadav et al., 2024). Second, we embed a standardized micro-teaching ecosystem (≥4 practices/semester; ≥10 core skills; minimum passing grade B-) as auditable psychomotor evidence (Al-Mubaid & Bettayeb, 2017; J. Biggs, 2003; Qiao & Fu, 2022). Third, we apply a Goal-Free Evaluation approach to link assessment design to objective institutional indicators—GPA trends, on-time graduation, and graduate outcomes (tracer coverage, employment, time-to-job)—and examine its implications for employability in an Islamic–Malay context (Belyakova & Zakharova, 2020; Handrianto et al., 2025; Harden, 2007; Muhammad Samin et al., 2025; Qiao & Fu, 2022; Telukdarie & Munsamy, 2019; Theodoridou, 2021; Yusri & Muhammad Samin, 2024).

METHOD

Design and Approach

We adopted a Goal-Free Evaluation (GFE) to examine the effectiveness of the holistic assessment system while minimizing bias toward pre-stated objectives (Scriven, 1972; Youker, 2013). GFE was selected to enable objective evaluation based on actual empirical evidence rather than predetermined program goals, allowing us to assess what the assessment system achieves rather than what it was intended to achieve. The effectiveness of the holistic assessment system was evaluated through multiple dimensions: (1) cognitive effectiveness—measured by sustained high GPA and improved on-time graduation rates; (2) affective effectiveness—indicated by high student satisfaction across service dimensions; (3) psychomotor effectiveness—evidenced through standardized micro-teaching performance and skill transfer; and (4) overall system effectiveness—demonstrated by the integration of all three domains within the GLOs→CLOs alignment framework.

To examine the association between the holistic assessment system and academic attainment, we analyzed: (a) temporal trends in GPA and on-time graduation across three academic years; (b) the relationship between assessment domain performance (cognitive, affective, psychomotor) and overall academic indicators; and (c) the contribution of the 40–30–30 assessment distribution to sustained academic performance. To examine the association with graduate outcomes, we investigated: (a) the relationship between integrated domain performance and employment rates; (b) the connection between micro-teaching evidence and time-to-job; and (c) the alignment between assessment outcomes and post-graduation activity distribution.

DOI: 10.18860 /ijazarabi.V9i1.36779
ISSN(print): 2620-5912 | ISSN(online):2620-5947
ejournal.uin-malang.ac.id/index.php/ijazarabi/index

The study used a mixed-methods design with a descriptive-dominant strand to summarize trends and proportions in academic performance, triangulated with QA documentation and instructional artefacts. This approach enabled us to document both quantitative indicators of effectiveness and qualitative evidence of system implementation and outcomes.

Context and Scope.

The setting was the ALE Study Program at Universitas Islam Riau (UIR), covering the last three academic years (Y-2 to Y). The unit of analysis comprised program-level indicators aligned to the cognitive, affective, and psychomotor assessment domains.

Data Sources

Four groups of data were used: (1) verified academic records (mean graduate GPA, on-time graduation, 144 credits workload); (2) QA documents (Semester Learning Plan (SLP); GLOs→CLOs→sub-CLOs mapping; 40–30–30 assessment distribution; semester QA monitoring; Unit of QA audit reports); (3) student satisfaction surveys (reliability, responsiveness, assurance, empathy, facilities; "very good" category); and (4) graduate tracer study (response coverage, placement in work/study/entrepreneurship, job-seeking time, and post-graduation activity distribution).

Instruments And Procedures

Psychomotor assessment was evidenced through a structured micro-teaching ecosystem (≥4 practice sessions/semester, minimum passing grade B-, rubrics covering ≥10 core teaching skills, language and micro-teaching labs). Cognitive assessment relied on formative and summative artefacts (quizzes, assignments, essays, mid/final exams, portfolio) aligned to GLOs/CLOs. Affective indicators were captured by standardized institutional service surveys and policies (task force for Prevention and Handling of Sexual Violence (PHSV), Occupational Safety and Health (OSH), counseling). Tracer data were sourced from the university-level tracking system.

Measures and Operationalization

Cognitive indicators: annual mean graduate GPA and on-time graduation rate. Affective indicators: proportion of respondents reporting "very good" per service dimension. Psychomotor indicators: operational parameters of micro-teaching (practice frequency, passing threshold, breadth of teaching skills, facility availability). Graduate outcomes: tracer coverage, total placement, mean job-seeking time, and post-graduation activity distribution (education sector jobs, Arabic-based entrepreneurship, master's study, teacher certification Program (TCP)).

Data Analysis

Descriptive summaries were compiled: (a) means, proportions, and Y-2→Y trends for GPA; on-time graduation was available for Y only; (b) cross-tabulation of service dimensions; (c) summary of micro-teaching parameters; (d) aggregation of tracer metrics (coverage, placement, time-to-job, activity distribution). Results are presented as Tables 1–6 in the Results section to preserve traceability between indicators, operational definitions, and findings.

Validity, Reliability, and Triangulation

Assessment reliability was supported by standardized rubrics across courses, semesterly GLOs/CLOs monitoring, and Internal Quality Assurance (IQA) auditing. Triangulation occurred across sources (records, QA documents, surveys, tracer) and methods (document analysis, survey, micro-teaching performance records). Process

consistency was maintained through institutional QA procedures and internal review.

RESULTS AND DISCUSSION

This study examines the effectiveness of a holistic assessment system (cognitive, affective, psychomotor) in ALE at Universitas Islam Riau (UIR) using a Goal-Free Evaluation approach. Through GFE, we evaluated effectiveness by examining actual outcomes rather than predetermined targets, allowing objective assessment of what the system achieves. We summarize the findings below, explicitly linking the holistic assessment system's effectiveness to academic attainment and graduate outcomes documented in the internal QA system.

Cognitive results: Effectiveness and Association with Academic Attainment

Cognitive indicators demonstrated strong and consistent effectiveness. Across the last three cohorts, GPA ranged from 3.69 to 3.78 (Table 2), indicating stable mastery of course outcomes. The GFE approach revealed that the holistic assessment system effectively sustained high cognitive performance without explicit targeting of GPA thresholds. On-time graduation reached 52% in Y, representing a significant improvement and demonstrating the system's effectiveness in supporting academic efficiency.

The association between the holistic assessment system and academic attainment is evidenced through several mechanisms: (1) the 40% cognitive weighting, explicitly linked to GLOs→CLOs→sub-CLOs, ensures that cognitive assessment directly contributes to overall academic performance; (2) Program-level Course Outcomes – Program Outcomes (CO–PO) mapping, maintained through SLP and GLOs/CLOs monitoring, creates traceable pathways from assessment tasks to institutional indicators; (3) the integration of cognitive assessment within the 40–30–30 framework positions cognitive evidence within an integrated performance system rather than as a standalone measure, potentially enhancing motivation and engagement across domains; (4) routine QA audits, semesterly monitoring, and formative–summative assessment practices ensured traceability of learning outcomes, supporting continuous improvement and alignment. The stable high GPA across three years, combined with improved on-time graduation, suggests that the holistic assessment system effectively supports both academic quality and efficiency.

Affective Results

Student satisfaction was high, with “very good” ratings ≥75% across most service dimensions (reliability, responsiveness, assurance, empathy, facilities), indicating a learning climate that supports motivation, confidence, and academic ethos. Student protection policies (PHSV taskforce) and comprehensive OSH arrangements reinforced wellbeing; OSH, health, and counseling indicators scored highly. Culturally, Islamic–Malay values with global awareness were embedded through SLP, course materials, community practice (Teaching Practicum (TP)/Community Service Placement (CSP)/Community Engagement (CE)), and academic fora. The affective domain was tracked via professional attitudes, ethical conduct, and participation in scholarly and faith-based activities. The integration of informal Arabic language environment concepts within the institutional framework supports affective development by creating authentic language use contexts that enhance motivation and engagement (Samin et al., 2023).

Psychomotor Results

Psychomotor performance was robust within a structured micro-teaching ecosystem: each student practiced ≥4 times per semester, met a minimum passing grade of B-, and was assessed with rubrics covering ≥10 core teaching skills (e.g., lesson opening/closure, variation, questioning, reinforcement, media use, classroom management, assessment). Micro-teaching and language laboratories enabled repeated, purposeful practice that strengthened pedagogical and oral (listening/speaking) skills. Outputs were visible in student achievements (international, national, and local awards) and scholarly products (intellectual property, ISBN books, and publications), indicating transfer from classroom learning to real-world performance.

Graduate Outcomes And External Relevance: Association With Holistic Assessment System

The tracer study covered 92.86% of graduates; 97.5% were employed, self-employed, or in further study, with a mean time-to-job of 4.8 months (Table 6). Placement was concentrated in the education sector (~77%), followed by Arabic-based entrepreneurship (~10%). The association between the holistic assessment system and graduate outcomes is demonstrated through multiple pathways: (1) Cognitive domain effectiveness—the sustained high GPA (3.69–3.78) indicates strong knowledge mastery that employers value, contributing to high placement rates (97.5%) and short time-to-job (4.8 months); (2) Affective domain contribution—the ≥75% "very good" satisfaction ratings across service dimensions reflect professional dispositions (ethics, empathy, assurance) that enhance employability, particularly in education sector positions (~77%); (3) Psychomotor domain impact—the standardized micro-teaching ecosystem (≥4 practices/semester, ≥10 core skills, B- threshold) develops demonstrable teaching and communication skills that directly transfer to workplace performance, explaining the concentration in education sector employment; (4) Integrated system effect—the 40–30–30 balance ensures graduates possess integrated competence (knowledge, attitudes, performance) that employers seek, resulting in favorable employment outcomes.

The strong association is further evidenced by: (a) the alignment between assessment domains and employment sectors—education sector dominance (~77%) corresponds with the psychomotor emphasis on teaching skills; (b) the short time-to-job (4.8 months) suggests that integrated assessment prepares graduates with workplace-ready competencies; (c) the high placement rate (97.5%) indicates that the holistic assessment system effectively develops the multidimensional capabilities that employers value. These patterns suggest that integrated cognitive mastery, professional ethics, and performative teaching/communication skills, developed through the holistic assessment system, are directly linked to work readiness and early career sustainability.

Table 1. Holistic Assessment Distribution (CO–PO)

Domain	Weight (%)
Cognitive	40
Affective	30
Psychomotor	30

Table 2. GPA and on-Time Graduation Trends (with sample size and uncertainty)

Year	N (graduates)	Mean GPA	SD (GPA)	On-Time Graduation (%)	95% CI (On-time)
Y-2	28	3.73	0.18	-	-
Y-1	43	3.69	0.12	-	-

Year	N (graduates)	Mean GPA	SD (GPA)	On-Time Graduation (%)	95% CI (On-time)
Y	31	3.78	0.14	52.00	34.4–69.6

Table 3. Student Satisfaction by Dimension ("Very Good" category; with N and 95% CI)

Dimension	N (respondents)	Percentage (%)	95% CI
Reliability	179	76.20	69.9–82.4
Responsiveness	179	75.88	69.6–82.2
Assurance	179	76.86	70.7–82.6
Empathy	179	76.28	69.9–82.3
Facilities	179	77.44	71.3–83.6

Note: Internal consistency (Cronbach's alpha) = NA (item-level survey data not available at program level).

Table 4. Core Parameters of Micro-teaching

Component	Standard/Practice	Note
Practice frequency	≥ 4 times/semester	Deliberate practice
Passing standard	Minimum B-	Standardized rubrics
Skill coverage	≥ 10 core teaching skills	Variation, questioning, etc.
Facilities	Micro-teaching & language labs	Technology & media support

Table 5. Survey and Rubric Reliability

Instrument	Scope	N (sample)	N items/ skills	Model/Design	Coefficient	95% CI	Bench mark
Student service survey	5 dimensions (Likert)	179	5	Cronbach's alpha	—	—	$\alpha \geq 0.80$
Micro-teaching rubric (overall)	10 core teaching skills	102	10	ICC(2,k)	—	—	ICC ≥ 0.75
Micro-teaching rubric (sub-skill)	Variation, questioning, etc.	102	1 per sub-skill	ICC(2,1)	—	—	ICC ≥ 0.60 (≥0.75 good)

Note: N=102 (graduates Y-2→Y) is a proxy. ICC should use the number of rated performances (students×tasks/videos) with ≥2 raters. Alpha and ICC will be computed once item-level and rater-level files are available, using a two-way random-effects, absolute-agreement model; report N raters, N tasks/videos per student, and specify ICC(2,1) or ICC(2,k).

Table 6. Graduate Outcomes (Tracer and Post-Graduation Activities; with N and 95% CI)

Indicator	N (graduates)	Value	95% CI
Tracer coverage	42	92.86%	81.9–97.7
Total placement	39	97.5%	86.8–99.9
Time-to-job	39	4.8 mo	—

Post-Graduation Activity	N	Percentage (%)	95% CI
Education sector employment	30	76.9	61.5–88.2
Arabic-based entrepreneurship	4	10.3	3.3–24.2
Master's study	4	10.3	3.3–24.2
Teacher certification (PPG)	0	0.0	0.0–9.2
Total	38	97.5	86.8–99.9

Note: One graduate still job-seeking; hence activity total N=38 while coverage N=39.

Table 1 evidence a proportional 40–30–30 assessment design consistent with constructive alignment, ensuring that measurement attends to knowledge, professional attitudes, and performative classroom skills. Through GFE, we evaluated the effectiveness of this system by examining actual outcomes: the system effectively balanced all three domains, with each domain contributing to overall program performance.

The GFE approach revealed that the holistic assessment system is effective across all three domains: (1) Cognitive effectiveness—Table 2 shows stable, high GPA (3.69–3.78) and an on-time graduation level of 52% for Y, pointing to efficiency without undermining quality. The 40% cognitive weighting, aligned with GLOs→CLOs, effectively supports knowledge mastery and academic completion; (2) Affective effectiveness—Table 3's high "very good" ratings (≥75% across dimensions) indicate that the 30% affective weighting, operationalized through service surveys and institutional policies, effectively creates a supportive learning climate. Assurance and facilities act as affective enablers of engagement and persistence; (3) Psychomotor effectiveness—Table 4 demonstrates standardized, repeated practice (≥4 times/semester, ≥10 skills, B- threshold) that effectively accelerates skill automatization and transfer to authentic contexts. The 30% psychomotor weighting ensures systematic development of teaching and communication competencies.

The association between the holistic assessment system and academic attainment is demonstrated through: (a) Temporal consistency—stable high GPA across three years (Y-2: 3.73; Y-1: 3.69; Y: 3.78) suggests the system consistently supports academic performance; (b) Efficiency gains—improved on-time graduation (52% in Y) indicates the system supports both quality and throughput; (c) Integrated contribution—the balanced 40–30–30 design ensures all domains contribute to overall academic success, with cognitive performance enhanced by affective climate and psychomotor practice.

Table 6 confirms strong employability (97.5% placement, 4.8-month time-to-job) and demonstrates clear association with the holistic assessment system: (a) Domain integration—the combination of cognitive mastery (high GPA), affective dispositions (high satisfaction), and psychomotor skills (standardized micro-teaching) creates integrated competence that employers value; (b) Sector alignment—education sector dominance (~77%) directly corresponds with psychomotor emphasis on teaching skills; (c) Career readiness—short time-to-job (4.8 months) suggests the system effectively prepares graduates with workplace-ready competencies. These patterns are consistent with integrated attainment across domains, indicating that the holistic assessment system effectively supports both academic success and employability.

Programmatic implications include rubric calibration and moderation, analytics dashboards linking GLOs→CLOs→tasks→scores→outcomes, and sustained investment in micro-teaching and language labs. Curricular implications include recurring authentic tasks (micro-teaching, e-portfolio, capstone) with structured feed-forward. Equity considerations require multiple submission modes and scaffolds for digital access. The model appears scalable to cognate teacher-education programs given prerequisites.

These findings align with constructive alignment theory, where clarity of intended outcomes and assessment criteria drives learning behaviors and attainment (J. Biggs, 2003; J. B. Biggs & Collis, 2014; Harden, 2007). The combination of stable–high GPA and rising on-time graduation mirrors broader OBE implementations reporting efficiency and quality gains when outcomes, tasks, and rubrics are coherently aligned (Yadav et al., 2024). High student satisfaction on assurance and facilities echoes evidence that service clarity, safe learning environments, and inclusive support correlate with engagement and persistence, particularly in multicultural and value-rich contexts (Petrušić, 2023; Theodoridou, 2021).

Psychomotor strengthening through standardized micro-teaching is consistent with experiential learning and skill automatization via repeated, scaffolded practice

(Kolb et al., 2014), supported by self-regulated and collaborative learning mechanisms (Johnson, 2017; Samin et al., 2022). Research in Arabic language education demonstrates that self-regulated learning strategies enhance language proficiency and skill development, particularly when integrated with structured practice environments (Samin et al., 2022; Samin & Hikmah, 2021). For performance assessment design, combining holistic judgments of overall teaching performance with analytic sub-scores yields diagnostic feedback on sub-capabilities, enhancing reliability and instructional usefulness (Zlatkin-Troitschanskaia et al., 2019). Technology-enabled feedback loops and process automation (e.g., video-based review, rubric banks, workflow tools) can further increase consistency and timeliness (Chen et al., 2024; Lugay et al., 2020; Telukdarie & Munsamy, 2019; Thompson & Asanov, 2024).

Regarding method choices, the use of Goal-Free Evaluation helped minimize goal bias, fitting calls for evaluation approaches that foreground actual performance evidence and institutional outcomes (Chen et al., 2024). In assessment modalities, station-based oral assessments such as Multiple Mini Viva allow triangulation of cognitive, interpersonal, and professionalism indicators in blended formats (Hadie et al., 2022), while the Holistic Assessment Framework emphasizes participatory, authentic evidence across domains and is adaptable to higher education (Mitra & Wadegaonkar, 2024). For psychomotor–affective analytics, blended learning with multimodal analysis (e.g., CNN/MVML) shows promise for more objective, granular feedback on performative skills, subject to contextual adaptation for language-teaching tasks (Zhang & Wang, 2024).

Finally, employability signals (high placement, short time-to-job) are coherent with literature noting that integrated competence development in knowledge, attitudes, and demonstrable performance is associated with smoother school-to-work transitions, especially when programs incorporate culturally responsive practices and contemporary digital capacities (David, 2025; Samin et al., 2025; Telukdarie & Munsamy, 2019). Recent studies in Arabic language education specifically demonstrate that outcome-based approaches leveraging learning analytics effectively support graduate success and employability outcomes (Samin et al., 2025). Together, these convergences indicate that a programmatic portfolio of aligned tasks, standardized rubrics, moderated judgments, and technology-supported feedback provides a plausible mechanism linking holistic assessment to academic efficiency and early career outcomes.

CONCLUSION

Overall, a balanced 40–30–30 assessment architecture, explicitly tied to GLOs→CLOs and embedded in routine QA cycles, sustained performance across all domains. The pattern—stable-high GPA with rising on-time graduation, ≥75% “very good” satisfaction, and auditable psychomotor evidence from a standardized micro-teaching ecosystem—was accompanied by strong tracer outcomes (92.86% coverage, 97.5% placement, 4.8-month time-to-job), consistent with improved employability. Theoretically, this study contributes to OBE in ALE by demonstrating how integrated assessment across domains, cross-course rubrics, and Establish–Implement–Evaluate–Control–Improve QA preserve traceability and fairness. Goal-Free Evaluation helps minimize bias toward formal program goals and centers evaluation on achievement evidence and institutional outcomes. Practically, we recommend: (1) standardization–digitalization of rubrics and e-portfolio artefacts to strengthen

psychomotor–affective evidence; (2) assessor moderation and analytics dashboards linking tasks→scores→outcomes for early anomaly detection; (3) expanded school/community partnerships to enrich practice contexts; and (4) proportionate evidence management to enhance quality without undue administrative burden. Looking ahead, longitudinal tracking should estimate the relative contributions of each domain to institutional indicators; reliability studies should calibrate inter-rater agreement across practice-oriented courses; and equity evaluations should ensure fair access to learning and assessment resources. Subject to these prerequisites, the model is replicable and ready for scaling to cognate programs in language and teacher education.

REFERENCES

Agrawal, E., Tungikar, V., & Joshi, Y. (2021). Method For Assessment And Attainment Of Course And Program Outcomes for Tier-I institutes in India. *Journal of Engineering Education Transformations*, 35–41. https://sciresol.s3.us-east-2.amazonaws.com/srs-j/jeet/pdf/volume34/issue3/JEET556.pdf

Al-Mubaid, H., & Bettayeb, S. (2017). Intervention-Based Method For Improving Learning And Critical Thinking. *2017 Computing Conference*, 1232–1237.

Anderson, B. R., Shah, J. H., & Kreminski, M. (2024). Homogenization Effects Of Large Language Models On Human Creative Ideation. *Proceedings of the 16th Conference on Creativity & Cognition*, 413–425. https://doi.org/10.1145/3635636.3656204

Belyakova, E. G., & Zakharova, I. G. (2020). Professional Self-Determination And Professional Identity Of Students-Teachers In The Conditions Of Individualisation Of Education. *Obrazovanie i Nauka*, *22*(1). https://doi.org/10.17853/1994-5639-2020-1-84-112

Biggs, J. (2003). Aligning Teaching And Assessing To Course Objectives. *Teaching and Learning in Higher Education: New Trends and Innovations*, 1–9.

Biggs, J. B., & Collis, K. F. (2014). Evaluating The Quality Of Learning: The Solo Taxonomy *(*Structure of the Observed Learning Outcome*)*. Academic press.

Chen, L., Zaharia, M., & Zou, J. (2024). How Is ChatGPT's Behavior Changing Over Time? *Harvard Data Science Review*, *6*(2). https://doi.org/10.1162/99608f92.5317da47

Conradty, C., & Bogner, F. X. (2022). Education for Sustainable Development: How Seminar Design And Time Structure Of Teacher Professional Development Affect Students' Motivation And Creativity. *Education Sciences*, *12*(5), 296. https://doi.org/10.3390/educsci12050296

David, J. O. (2025). Balancing Growth, Inclusivity, and Technology: Rethinking Student-Centered Learning in South African Higher Education. *International Journal of Learning, Teaching and Educational Research*, *24*(4), 127–153. https://doi.org/10.26803/ijlter.24.4.7

Frank, J. R., & Danoff, D. (2007). The CanMEDS Initiative: Implementing an Outcomes-Based Framework Of Physician Competencies. *Medical Teacher*, *29*(7), 642–647. https://doi.org/10.1080/01421590701746983

Ghani, K. A., Yusof, N. M. R. N., Baharuddin, H., Yamat, H., Ahmad, Z., & Abdullah, I. (2011). Development of A Learning Module on Arabic Language Skills Outside of the Classroom. *Procedia-Social and Behavioral Sciences*, *18*, 154–162. https://doi.org/https://doi.org/10.1016/j.sbspro.2011.05.023

Hadie, S. N. H., Kasim, F., Mohd Ismail, Z. I., Asari, M. A., Mohd Yusof, N. A., Simok, A. A., Shamsuddin, S. A., & Mohd Amin, M. S. I. (2022). Multiple Mini Viva as an Online Assessment Tool for Postgraduate Anatomy Programme: The USM's Experience Amidst COVID-19 Pandemic. *Education in Medicine Journal*, *14*(4), 99–112. https://doi.org/10.21315/eimj2022.14.4.9

Handrianto, C., Solfema, S., & Jusoh, A. J. (2025). Cultural Adaptation and Academic Success: Exploring the Experiences of Indonesian Students in Malaysian Universities. *Journal of Ethnic and Cultural Studies*, *12*(3), 251–280. https://doi.org/10.29333/ejecs/2496

Harden, R. M. (2007). Outcome-Based Education: The Future Is Today. *Medical Teacher*, *29*(7), 625–629. https://doi.org/10.1080/01421590701729930

Johnson, K. (2017). *An introduction to foreign language learning and teaching*. Routledge.

Joseph, J., Jose, J., Jose, A. S., Ettaniyil, G. G., Cyriac, J., Sebastian, S. K., & Joseph, A. P. (2024). Quantitative Insights Into Outcome-Based Education: A Bibliometric Exploration. *Int. J. Eval. Res. Educ*, *13*(6), 4030–4045. https://doi.org/10.11591/ijere.v13i6.29272

Kolb, D. A., Boyatzis, R. E., & Mainemelis, C. (2014). Experiential Learning Theory: Previous Research And New Directions. In *Perspectives on Thinking, Learning, and Cognitive Styles*. https://doi.org/10.4324/9781410605986-9

Lugay, C. I. P., Agustin, B. E., Codico, M. A. E., Tapis, C.-A. P., & Tolentino, E. (2020). A Design Of An Automated Assessment Process For The Outcomes-Based Education In The Industrial Engineering Department Of The Faculty Of Engineering At The University Of Santo Tomas. *Journal of Physics: Conference Series*, *1529*(3), 032046. https://doi.org/10.1088/1742-6596/1529/3/032046

Mitra, G., & Wadegaonkar, A. (2024). Multidimensional Holistic Assessment Framework (HAF): A Case Study of Exploring the Discourses from Elementary School Teachers. *International Electronic Journal of Elementary Education*, *17*(1), 51–69. https://doi.org/10.26822/iejee.2024.363

Muhammad Samin, S., Akzam, I., Pebrian, R., & Fikriansyah, M. H. (2025). Tantangan dan Strategi Penerapan Lingkungan Bahasa Arab di Pesantren [Challenges and Strategies for Implementing Arabic Language Environment in Pesantren]. *Al-Hikmah: Jurnal Agama Dan Ilmu Pengetahuan*, *22*(1), 42–55. https://doi.org/10.25299/ajaip.2025.vol22(1).21006

Munzenmaier, C., & Rubin, W. N. (2013). Perspectives Bloom's Taxonomy: What's Old Is New Again. In C. Holcombe (Ed.), *Educational Objectives*. The elearning Guild. https://publicservicesalliance.org/wp-content/uploads/2013/04/guildresearch_blooms2013.pdf

Petrušić, I. (2023). Student-Centred Learning in the Context of External Quality Assurance. *Medicina Fluminensis*, *59*(4). https://doi.org/10.21860/medflum2023_309615

Qiao, Y., & Fu, H. (2022). Study on OBE Teaching Concept In The Context Of Deep Learning For The Construction Of University Mathematics Microcourses. *Computational Intelligence and Neuroscience*, *2022*(1), 6860842. https://doi.org/10.1155/2022/1234567

Samin, S. M. (2019). Heutagogy In Arabic Class: How It Is Applied in The Islamic Education Study Program of Universitas Islam Riau. *Journal of Arabic Linguistics and Education*, *5*(1), 20–29. https://doi.org/10.28918/alsinatuna.v5i1.2092

Samin, S. M., Akzam, I., & Supriady, H. (2022). Strategies of Arabic Students' Self-Regulated Learning Improvement in Language Proficiency in The Disruption Era. *Jurnal Pendidikan Bahasa Arab Dan Kebahasaaraban*, *9*(2), 165–177. https://doi.org/10.15408/a.v9i2.22828

Samin, S. M., & Hikmah, H. (2021). Self-Regulated Learning Of Arabic Education Students Via Moodle Discussion Forum. *Journal of Arabic Linguistics and Education*, *7*(1), 17–29. https://tinyurl.com/yyms4ts9.

Samin, S. M., & Osman, R. A. (2024). Integrating Artificial Intelligence into the Arabic Language Teaching Plan at Higher Education. *SHS Web of Conferences*, *202*, 06010. https://doi.org/10.1051/shsconf/202420206010

Samin, S. M., Pebrian, R., & Zulkifli, A. (2020). Heutagogy Approaches for Arabic Learning in Higher Education in Industrial Revolution 4.0. ICoSEEH 2019, 454–457. https://doi.org/10.5220/0009382304540457

Samin, S. M., Zakaria, G. A. N., & Pebrian, R. (2025). Outcome-Based Arabic Language Education: Leveraging Learning Analytics for Graduate Success. *Arabiyat : Jurnal Pendidikan Bahasa Arab Dan Kebahasaaraban*, *12*(2), 243–259. https://doi.org/10.15408/a.v12i2.46603

Samin, S. M., Zulkifli, A., & Supriady, H. (2023). Concepts of Informal Arabic Language Environment for Higher Education. *Al-Hikmah: Jurnal Agama Dan Ilmu Pengetahuan*, *20*(1), 28–38. https://doi.org/https://doi.org/10.25299/al-hikmah:jaip.2023.vol20(1).12026

Scriven, M. (1972). Objectivity and Subjectivity In Educational Research. *Teachers College Record*, *73*(5), 94–142. https://doi.org/10.1177/01614681720730050

Spady, W. G. (1994). Outcome-Based Education: Critical Issues and Answers. ERIC. https://files.eric.ed.gov/fulltext/ED380910.pdf

Telukdarie, A., & Munsamy, M. (2019). Digitization of Higher Education Institutions. *2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 716–721.

Theodoridou, S. (2021). The Importance Of Student-Centered Learning In The Development Of Critical Thinking (full text in Greek). *Academia*, *23–24*, 226–245.

Thompson, A. S., & Asanov, E. (2024). "Nonnative? Next!" Native-speakerism in world language job advertisements. *Studies in Second Language Learning and Teaching*, *14*(1), 49–74. https://doi.org/10.14746/ssllt.42371

Yadav, P., Tomar, A., Soni, R., Dubey, G., Bhargava, C. P., & Kumari, J. (2024). Metrics With Performance Indicators Used For Calculating the Mapping Strength of Course Outcome With Program Outcome. *Journal of Engineering Education Transformations*, 85–95. https://doi.org/10.16920/jeet/2024/v38i2/24192

Youker, B. W. (2013). Goal-free evaluation: A Potential Model For The Evaluation Of Social Work Programs. *Social Work Research*, *37*(4), 432–438. https://doi.org/10.1093/swr/svt032

Yusri, S. A., & Muhammad Samin, S. (2024). The Effectiveness Of Group Learning In Improving Students'Arabic Speaking proficiency. *Perspektif Pendidikan Dan*

Keguruan, *15*(1), 67–72. https://doi.org/10.25299/perspektif.2024.vol15(1).15852

Zhang, Z., & Wang, W. (2024). Enhancing Dance Education Through Convolutional Neural Networks And Blended Learning. *PeerJ Computer Science*, *10*. https://doi.org/10.7717/peerj-cs.2342

Zlatkin-Troitschanskaia, O., Shavelson, R. J., Schmidt, S., & Beck, K. (2019). On The Complementarity Of Holistic And Analytic Approaches To Performance assessment scoring. *British Journal of Educational Psychology*, *89*(3), 468–484. https://doi.org/10.1111/bjep.12286