Research Paper

Analysis of the Effectiveness of the Assessment System in Enhancing Communicative Competence in Teaching Arabic as a Foreign Language: A Field Study of Students at the Faculty of Usuluddin, Sultan Sharif Ali Islamic University, Brunei Darussalam

Achmad Yani[1], Siti Sara Binti Haji Ahmad[2], Rafidah Binti Abdullah, Siti Shalihah[4] and Anis Malik Thoha[5].

[1,2,3&5] *Universiti Islam Sultan Sharif Ali (UNISSA), Brunei Darussalam.*
[4]*Universitas Islam Negeri Sultan Maulana Hasanuddin Banten, Indonesia.*

Abstract
This study aims to evaluate the effectiveness of the assessment system employed in teaching Communicative Arabic at the Faculty of Usuluddin, Sultan Sharif Ali Islamic University, Brunei Darussalam. The research specifically examines the system's alignment with instructional objectives, its ability to measure students' communicative competencies, and its comprehensiveness in covering the four essential language skills. Adopting a descriptive-analytical methodology with a quantitative evaluative dimension, the study utilized a closed-ended questionnaire designed to assess ten evaluative dimensions, including content validity, construct validity, result validity, comprehensiveness, balance in score distribution, alignment with course content, and the integration of theoretical and practical components. The sample comprised 12 students, representing 40% of the total study population. Data were analyzed using descriptive statistical techniques, including frequency distributions, percentages, and mean scores. Findings revealed a high degree of alignment between the assessment system and curriculum goals (88.3%), satisfactory effectiveness in measuring oral (85%) and written (90%) competencies, and a relative balance in the distribution of scores across skills. Furthermore, the system demonstrated acceptable construct validity (81.7%), high result validity (93.3%), and notable comprehensiveness (88.3%). However, discrepancies were observed in the integration of practical components (80%), indicating a need for enhancement of authentic performance-based assessment tools. The study recommends redesigning oral assessment tools, expanding authentic assessment tasks, adopting global standards such as the Common European Framework of Reference for Languages (CEFR), and promoting a culture of self-assessment. This research constitutes a significant contribution to advancing assessment strategies in teaching Arabic as a foreign language, ensuring the development of communicative competence and reinforcing the pedagogical-assessment synergy within Islamic higher education contexts.

Keywords: *Assessment System, Communicative Arabic, Sultan Sharif Ali Islamic University.*

I. Introduction

In recent decades, the field of Teaching Arabic as a Foreign Language (TAFL) has undergone a significant transformation in both its objectives and methodologies. Linguistic competence is no longer confined to mastery of grammatical structures or isolated vocabulary; it has evolved into what is now recognized as communicative competence—the ability to use language effectively and appropriately in authentic social contexts. This paradigm shift is a direct reflection of communicative language theories, which place interaction and meaning at the heart of the learning process (Canale & Swain, 1980).

In line with this evolution, language assessment tools are increasingly required to adapt in ways that reflect the communicative dimension of language use, rather than focusing solely on linguistic knowledge. The success of any language program is not measured merely by the quality of its content, but by the capacity of its

assessment instruments to track students' progress in listening, speaking, reading, and writing—within the framework of clearly defined and standardized learning outcomes (Brown, 2004, p. 49).

Accordingly, the need to evaluate the testing systems employed in communicative Arabic instruction within academic institutions becomes both urgent and essential. While many institutions continue to rely on traditional assessment models, there is a growing demand for empirical field studies that objectively examine the validity, comprehensiveness, balance, and authenticity of these evaluations (Weir, 2005, pp. 21–24; Hughes, 2003, pp. 26–28).

Sultan Sharif Ali Islamic University (UNISSA) in Brunei Darussalam stands as one of the leading institutions committed to Arabic language education, viewing Arabic as both a medium for academic engagement and religious communication. The university has integrated a course titled Communicative Arabic across all its faculties, including the Faculty of Usuluddin. This course is designed to equip non-native speakers with genuine communicative skills, enabling them to participate in academic and religious discourse using Arabic.

Despite the recognized importance of this course, the effectiveness of its current assessment system remains to be rigorously evaluated. This study seeks to determine the extent to which the administered assessments align with curricular objectives, comprehensively cover the core language skills, and accurately reflect the learners' actual language proficiency. Previous studies have indicated that many assessment frameworks suffer from an overemphasis on cognitive elements, while neglecting practical and contextual dimensions (Luoma, 2004, pp. 12–14; Fulcher & Davidson, 2007, pp. 91–92).

Alderson, Clapham, and Wall (1995, p. 107) emphasized that any imbalance in score distribution or the neglect of specific skills within assessments can result in biased outcomes, which in turn undermine pedagogical decisions concerning curriculum and program development. Furthermore, Gronlund and Linn (2000, pp. 42–44) underscored that alignment between assessments and classroom content is a fundamental prerequisite for achieving what is referred to as instructional validity.

This study derives its significance from its focus on an authentic and representative sample: first-year students at the Faculty of Usuluddin who have completed the Communicative Arabic course. As such, its findings are not only generalizable to the broader university context in Brunei but also relevant to comparable educational environments across Southeast Asia.

The urgency of this research lies in the pressing need to evaluate assessment systems in Arabic language instruction as a foreign language, particularly in light of modern principles of language assessment—such as validity, authenticity, comprehensiveness, and the integration of theoretical and practical components (Bachman & Palmer, 1996, pp. 18–34). This study also aims to fill an evident gap in the educational literature regarding the evaluation of communicative assessment tools within Islamic, non-Arab university contexts.

II.   Theoretical Framework and Previous Studies

2.1. Theoretical Framework

2.1.1. Assessment in Foreign Language Education

Assessment constitutes one of the fundamental pillars of the educational process and holds particular significance in the context of foreign language teaching due to its central role in evaluating linguistic competence, determining the achievement of instructional objectives, and enhancing teaching practices (Bachman & Palmer, 1996, pp. 18–34). Modern language assessment transcends traditional written examinations, embracing comprehensive strategies that incorporate cognitive, skill-based, and affective domains to reflect learners' authentic language performance (Luoma, 2004, pp. 12–14).

2.1.2. Communicative Competence and Its Impact on Assessment

The theory of Communicative Competence emerged as a paradigmatic shift from a focus on purely linguistic knowledge to an emphasis on real-life language use in communicative situations. This competence encompasses multiple dimensions: grammatical competence, sociolinguistic competence, strategic competence, and discourse competence (Canale & Swain, 1980). Language assessment tools must therefore be grounded in these dimensions to authentically measure communicative performance (Weir, 2005, pp. 21–24).

2.1.3. Types of Validity in Assessment Tools

Among the key characteristics of sound assessment instruments is validity. The most prominent types in educational research include content validity, construct validity, and criterion-related validity. Content validity refers to the extent to which a test represents the intended domain of behavior or knowledge. Construct validity indicates the tool's ability to measure the abstract concept it purports to assess. Criterion-related validity pertains to the correlation of test results with independent external benchmarks (Hughes, 2003, pp. 26–28; Brown, 2004, p. 49).

2.1.4. Oral and Written Assessment

Evaluating oral skills requires performance-based tools grounded in authentic interaction, such as presentations, interviews, and discussions—an area that presents significant pedagogical challenges. Written assessment, on the other hand, involves tests that accurately gauge writing skills, including textual organization, coherence, and

appropriate language usage. Assessment instruments should therefore be constructed within an integrated framework that balances both dimensions (Fulcher & Davidson, 2007, pp. 91–92).

2.1.5. Fairness and Score Distribution

Fairness is a foundational principle in educational measurement and is achieved through balanced score distribution across various skills, ensuring that the instrument does not favor one domain over another. Alderson et al. (1995, p. 107) note that imbalance in this distribution may lead to inaccurate results, thereby compromising the integrity of the assessment system.

2.2. Previous Studies

2.2.1. Bachman & Palmer (1996)

In their seminal work, Bachman and Palmer emphasized the necessity of designing language assessments that demonstrate validity, reliability, authenticity, and practicality. They argued that effective language tests must mirror real-life communicative tasks rather than merely theoretical knowledge. Their conceptual framework has significantly influenced the development of communicative assessment instruments.

2.2.2. Luoma (2004)

Published under the auspices of Cambridge University, this study focused on the assessment of speaking skills. It concluded that oral competencies can only be measured using contextually authentic tools such as interviews, group discussions, and dialogue tasks. The research also highlighted the need for precise, standardized criteria for oral assessment.

2.2.3. Weir (2005)

Weir presented a comprehensive framework for evaluating the quality of language assessment tools based on four dimensions: validity, reliability, comprehensiveness, and authenticity. He advocated for aligning written assessments with instructional contexts, emphasizing that they should evaluate both linguistic and textual competencies, rather than merely grammatical knowledge.

2.2.4. Hughes (2003)

Hughes' book is considered a foundational reference in language testing. He meticulously explored concepts such as test validity, objectivity, and alignment with learning objectives. The work also includes detailed models of both oral and written tests, with rigorous scoring criteria.

2.2.5. Fulcher & Davidson (2007)

This study emphasized the importance of comprehensiveness in language assessment, advocating for the inclusion of all components of language competence in testing instruments. The authors also examined the interplay between test construction and its practical applications in academic and professional settings.

2.2.6. Alderson, Clapham & Wall (1995)

Focusing on test design and the processes of validating and ensuring test reliability, this study underscored the necessity of balanced skill representation in scoring systems. It warned that the absence of such balance could result in distorted or inequitable assessment outcomes.

III. Research Methodology

3.1. Type and Methodological Approach

This study adopts a descriptive-analytical methodology with a quantitative evaluative dimension, a widely recognized approach in educational research. It is designed to collect descriptive data related to educational phenomena and analyze them statistically to interpret relevant patterns and trends (Creswell, 2012, pp. 295–298). The method was deemed appropriate for investigating linguistic and pedagogical issues—particularly the effectiveness of assessment systems. Specifically, this research aims to evaluate the effectiveness of the current assessment system employed in the teaching of Communicative Arabic at the Faculty of Usuluddin, Sultan Sharif Ali Islamic University, Brunei Darussalam. It assesses the alignment of this system with the intended learning objectives, its capacity to measure key communicative competencies (oral and written), and the extent to which it demonstrates comprehensiveness and validity in reflecting students' true language proficiency.

3.2. Research Instrument

A structured closed-ended questionnaire served as the primary data collection tool. This instrument is a reliable and widely accepted method in educational research, allowing for standardized data collection from multiple respondents within a short timeframe (Dörnyei, 2007, pp. 101–104). The questionnaire was meticulously constructed to address key evaluative dimensions of the assessment system, including alignment with curricular goals, assessment of oral and written skills, validity, comprehensiveness, balance in scoring, and integration of both theoretical and practical elements.

3.3. Instrument Validity and Reliability

To ensure content and face validity, the questionnaire was reviewed by a panel of academic experts in Arabic language education and educational program evaluation (Messick, 1995, pp. 742–743). Furthermore, Cronbach's

Alpha was employed to test the internal consistency of the instrument, confirming that all components exceeded the generally accepted threshold of 0.70 in educational research, thereby reinforcing the credibility of the findings (Muijs, 2004, p. 71).

3.4. Population and Sample

The research population comprised all first-year students at the Faculty of Usuluddin, Sultan Sharif Ali Islamic University, who enrolled in the Communicative Arabic course during the second semester of the 2024 academic year, totaling 30 students. The sample, selected through simple random sampling, consisted of 12 students—representing 40% of the total population. Despite its relatively modest size, the sample constitutes a significant proportion suitable for quantitative analysis within the framework of micro-case studies (Fraenkel, Wallen, & Hyun, 2012, pp. 91–92).

3.5. Data Collection Procedures

The researchers distributed the questionnaires during the final two weeks of the semester, following a detailed briefing on the study's objectives and strict assurances of data confidentiality. The data were collected using printed forms, which students completed under the supervision of the researchers to ensure accuracy and reliability.

3.6. Data Analysis Technique

The data were analyzed using descriptive statistical methods, including frequencies, percentages, and mean scores. The following formula was employed:

$$P\ (\%) = (N \cdot 5 \sum f_i \cdot x_i\ ) \times 100$$

The results were categorized into three levels—high, moderate, and low—based on the final index values, in accordance with the guidelines of Best & Kahn (2006, p. 389). Interpretations were made in light of established pedagogical standards for quality assessment in foreign language education.

3.7. Study Delimitations

The topical delimitation of this study is limited to the evaluation system of the Communicative Arabic course, excluding other curricular components. The human scope includes only first-year students at the Faculty of Usuluddin, while the temporal delimitation is confined to the second semester of the 2024 academic year. The geographical delimitation is restricted to Sultan Sharif Ali Islamic University, Brunei Darussalam.

IV. Field Study: Analysis and Discussion of Results

This study seeks to evaluate the effectiveness of the assessment system employed in the teaching of Communicative Arabic at Sultan Sharif Ali Islamic University, relying on quantitative data collected from a representative sample of students. The analysis focuses on key statistical indicators, particularly the approval rates across essential dimensions such as alignment with instructional objectives, evaluative efficiency, result validity, and comprehensiveness.

a. Alignment with Curriculum Objectives
The findings revealed that 50% of participants strongly agreed that the assessments align with the curriculum objectives, while 41.7% agreed, and 8.3% remained neutral—resulting in a cumulative positive response of 88.3%. This indicates a robust alignment between assessment mechanisms and curricular goals, serving as a strong indicator of content validity, a core criterion in educational evaluation, as affirmed by Brown (2004, p. 49) and Nitko & Brookhart (2014, pp. 8–9).

b. Effectiveness in Assessing Oral Proficiency
A total of 41.7% strongly agreed and 41.7% agreed that the assessments effectively measure oral communicative skills, with 16.7% expressing neutrality, yielding an overall approval rate of 85%. This reflects reasonable effectiveness, although the notable neutrality suggests challenges in capturing oral performance. Luoma (2004, pp. 12–14) emphasized that assessing speaking skills demands authentic, performance-based approaches rather than traditional paper-based tools.

c. Assessment of Written Proficiency
Half of the participants (50%) strongly agreed and the other half agreed on the system's effectiveness in evaluating written skills, yielding a 90% approval rate. This strongly supports the validity of the writing assessment, aligning with Weir's (2005, pp. 21–24) argument that written assessments tend to offer greater stability and consistency.

d. Balanced Score Distribution
Regarding score balance between oral and written components, 41.7% strongly agreed, 41.7% agreed, and 16.7% remained neutral, totaling 85% approval. This balance is essential for ensuring academic fairness, especially in

communicative language programs. As Alderson, Clapham & Wall (1995, p. 107) note, imbalance can compromise test credibility.

e. Alignment with Instructional Content

Some 58.3% strongly agreed and 33.3% agreed that assessments correspond well with classroom lessons and exercises, while 8.3% remained neutral—culminating in a 90% approval rate. This reflects strong instructional validity, essential for ensuring alignment between assessments and instructional content (Gronlund & Linn, 2000, pp. 42–44).

f. Reflection of Actual Language Proficiency

Responses showed that 41.7% strongly agreed, 33.3% agreed, 16.7% were neutral, and 8.3% disagreed—resulting in an overall agreement of 81.7%. This demonstrates acceptable construct validity, confirming that the assessments fairly reflect actual student performance (Bachman & Palmer, 1996, pp. 18–21).

g. Overall Validity of Assessment Results

The strongest approval rate was observed here: 66.7% strongly agreed and 33.3% agreed, yielding a perfect 93.3% approval. This underscores participants' confidence in the fairness and precision of the system, resonating with Hughes' (2003, pp. 26–28) emphasis on result validity as a cornerstone of assessment quality.

h. Comprehensiveness of the Assessment System

Half of the respondents (50%) strongly agreed, 41.7% agreed, and 8.3% were neutral—totaling 88.3%. This suggests that the assessment system adequately encompasses the various dimensions of communicative language competence. Fulcher & Davidson (2007, pp. 91–92) identified comprehensiveness as a principal determinant of assessment quality.

i.   Integration of Theoretical Knowledge

Fifty percent strongly agreed, 33.3% agreed, and 16.7% remained neutral, with an overall approval of 86.7%. This demonstrates that the assessment incorporates conceptual knowledge, reinforcing the integration between theory and practice as proposed by Hughes (2003, p. 35).

j. Integration of Practical Application

This area showed mixed responses: 50% strongly agreed, 25% agreed, 8.3% were neutral, 8.3% disagreed, and another 8.3% strongly disagreed—resulting in an 80% approval rate. While acceptable, this variance indicates a need for enhanced real-world performance integration. Bachman (1990, pp. 142–143) advocates for more extensive inclusion of authentic communicative tasks to nurture true communicative competence.

Analytical Summary

The findings of this field study reveal a high degree of alignment between the assessment system and the curricular goals of the Communicative Arabic program at Sultan Sharif Ali Islamic University. The system demonstrates notable content and construct validity, as well as consistency in targeting relevant communicative skills. However, gaps persist in the integration of applied performance elements, necessitating the refinement of authentic assessment tools that better simulate real-world communicative scenarios.

V. Conclusions, Suggestions and Recommendations

a.   **Conclusions**

Based on the quantitative data obtained from the field study, the following key findings were identified:

-          High alignment between assessment and curriculum objectives, with 88.3% of respondents affirming that the test system corresponds well to the stated educational goals. This reflects strong content validity and the alignment between what is assessed and what should be assessed (Brown, 2004, p. 49; Nitko & Brookhart, 2014, pp. 8–9).

-          Satisfactory effectiveness in measuring oral proficiency, with an approval rate of 85%, though 16.7% of participants expressed reservations. This suggests the need to refine oral assessment tools to better reflect authentic communicative performance (Luoma, 2004, pp. 12–14).

-          High efficacy in assessing written proficiency, with a total agreement rate of 90%, indicating the effectiveness of the written component of the assessment system (Weir, 2005, pp. 21–24).

-          Balanced distribution of grades across oral and written skills, with an approval rate of 85%, demonstrating fairness in evaluating communicative performance components (Alderson, Clapham & Wall, 1995, p. 107).

-          Strong alignment between assessments and instructional content, with 90% of participants affirming this link, reflecting high instructional validity of classroom assessments (Gronlund & Linn, 2000, pp. 42–44).

-          Moderate reflection of actual language proficiency, with 81.7% agreeing that the test results accurately reflect students' real linguistic levels, supporting acceptable construct validity (Bachman & Palmer, 1996, pp. 18–21).

-          High confidence in result validity, with a 93.3% approval rate, suggesting a high degree of trust in the fairness and accuracy of the system (Hughes, 2003, pp. 26–28).

- Notable comprehensiveness of the assessment system, with 88.3% approval, indicating satisfaction with the coverage of various dimensions of communicative competence (Fulcher & Davidson, 2007, pp. 91–92).

- Adequate integration of theoretical components, with an 86.7% agreement rate, demonstrating the conceptual depth of the assessments (Hughes, 2003, p. 35).

- Moderate integration of applied components, with an 80% approval rate. Some comments suggest a need to enhance the practical implementation of assessment activities (Bachman, 1990, pp. 142–143).

b.   Sugestions
Based on the above findings, the study proposes the following actions:
- Redesign oral assessments to be more authentic and interactive by incorporating real-life communication tasks and performance-based interviews.
- Enrich test content with applied tasks using authentic assessment methods such as oral presentations, simulations, and collaborative writing projects.
- Strengthen the integration of theoretical and applied components to ensure comprehensive assessments aligned with the goals of communicative competence development.
- Implement periodic reviews of assessment tools using validity and reliability checks to ensure continued alignment with evolving curriculum objectives.
- Develop a multidimensional question bank that addresses various domains of communicative competence (listening, speaking, reading, writing) and accommodates individual learner differences.

c.   Recommendations
In light of the findings and proposed measures, this study recommends:
- Adopting an evidence-based assessment policy in Communicative Arabic courses to ensure credibility, fairness, and pedagogical effectiveness, as advocated by Bachman & Palmer (1996).
- Training faculty members in designing and implementing adaptive, authentic assessments that reflect real communicative contexts and elevate students' language proficiency, in line with Luoma (2004).
- Conducting ongoing revisions of assessment structures based on student feedback and updating them in accordance with both formative and summative evaluation outcomes, as recommended by Hughes (2003).
- Employing international language assessment frameworks such as the Common European Framework of Reference (CEFR) to standardize proficiency levels and promote functionally-oriented performance development.
- Promoting a self-assessment culture among learners by incorporating elements of self- and peer-assessment, thereby fostering learner autonomy and metacognitive awareness.

References:

[1].   Alderson, J. C., Clapham, C., & Wall, D. (1995). Language Test Construction and Evaluation. Cambridge University Press. p. 107.
[2].   Bachman, L. F. (1990). Fundamental Considerations in Language Testing. Oxford University Press. pp. 142–143.
[3].   Bachman, L. F., & Palmer, A. S. (1996). Language Testing in Practice: Designing and Developing Useful Language Tests. Oxford University Press. pp. 18–34.
[4].   Best, J. W., & Kahn, J. V. (2006). Research in Education (10th ed.). Pearson Education. p. 389.
[5].   Brown, H. D. (2004). Language Assessment: Principles and Classroom Practices. Pearson Education. p. 49.
[6].   Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics, 1(1), 1–47.
[7].   Creswell, J. W. (2012). Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research (4th ed.). Pearson Education. pp. 295–298.
[8].   Dörnyei, Z. (2007). Research Methods in Applied Linguistics. Oxford University Press. pp. 101–104.
[9].   Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). How to Design and Evaluate Research in Education (8th ed.). McGraw-Hill. pp. 91–92.
[10].   Fulcher, G., & Davidson, F. (2007). Language Testing and Assessment: An Advanced Resource Book. Routledge. pp. 91–92.
[11].   Gronlund, N. E., & Linn, R. L. (2000). Measurement and Assessment in Teaching (8th ed.). Prentice Hall. pp. 42–44.
[12].   Hughes, A. (2003). Testing for Language Teachers (2nd ed.). Cambridge University Press. pp. 26–35.
[13].   Hughes, A. (2003). Testing for Language Teachers (2nd ed.). Cambridge University Press. pp. 26–35.
[14].   Luoma, S. (2004). Assessing Speaking. Cambridge University Press. pp. 12–14.
[15].   Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50(9), pp. 741–749.
[16].   Muijs, D. (2004). Doing Quantitative Research in Education: With SPSS. SAGE Publications. p. 71.
[17].   Nitko, A. J., & Brookhart, S. M. (2014). Educational Assessment of Students (7th ed.). Pearson. pp. 8–9.
[18].   Weir, C. J. (2005). Language Testing and Validation: An Evidence-Based Approach. Palgrave Macmillan. pp. 21–24. Alderson, J. C., Clapham, C., & Wall, D. (1995). Language Test Construction and Evaluation. Cambridge University Press. pp. 106–109.