

Preface

Evaluation is everywhere. It has been attributed a great deal of attention because of its germane role for different stakeholders, such as educators, policy-makers, students, teachers, parents, administrators, curriculum developers, book designers, and evaluation practitioners in general. Evaluation has repercussions for the individual, societal, economic, cultural, and political levels. It also has an ethical side, and it is tailored to the needs of these different people to remain abreast of the effectiveness and efficiency of programs. It is only in this regard that evaluation has to be implemented carefully by the right people.

Evaluation plays a very important role in different fields, such as language programs, quality assurance, teaching, and testing. For instance, in any educational reform (Leung & Rea-Dickins, 2007), evaluation is needed to improve teaching, testing, accreditation, and curriculum reform based on students' self-assessment (Ladyshevsky & Taplin, 2015). Evaluation can also be implemented for sustainable development that could be linked to quality assurance. Kiley and Rea-Dickins (2005, p. 6) define evaluation as a "form of enquiry, ranging from research to systematic approaches to decision-making." Evaluation is implemented from different perspectives: Students' learning and evaluation (Golding & Adam, 2014; Nygaard & Belluigi, 2011), teachers being evaluated by students, faculty staff evaluation, conceptions and practices of assessment (Hidri, 2015), formative versus summative evaluation, dynamic assessment, language program evaluation, doctoral dissertations or theses (Kyvik & Thune, 2015), teacher professional development evaluation, classroom observation to improve the teaching practices (Wei, 2015), evaluation of text genres, test evaluation (placement, diagnostic, progress, achievement, CBT), evaluation of teaching methods and methodologies and assessment literacy, etc. All these evaluation perspectives are contrived for decision-making purposes, such as maintaining "evaluation practices" (West, 2015). Tests and test validations are used for evaluation purposes in that they are evaluated for their test usefulness qualities, such as validity, reliability, authenticity, practicality, interactiveness, and impact (Bachman & Palmer, 1996, 2010).

To cope with the changing needs of the different stakeholders, many types of evaluations have witnessed tremendous changes: Formative vs. summative evaluation in improving teaching practices (Wei, 2015), classroom-based assessment, large-scale assessment, course evaluation (Bailey & Brown, 1996; Brown & Bailey, 2008), standardized examinations, dynamic assessment in mediating the test-takers to overcome their testing difficulties (Hidri, 2014; Lantolf & Poehner, 2011), frameworks of reference, teacher evaluation, and quality assurance and its sustainable development (Shiel, Filho, do Paço, & Brandli, 2015). Such evaluation facets require well-defined standards deemed necessary and relevant to demarcate the different benchmarks against which any evaluation judgment could be made.

Many researchers (e.g., Campbell & Mark, 2015; Hart, Diercks-O'Brien, & Powell, 2009; Jin, 2010; Kiley & Rea-Dickins, 2005; Swanwick, 2007) have investigated evaluation because of its cultural and prominent role in preparing and evaluating language teachers, focusing on the quality of learning and teaching, addressing language programs, the curriculum and changing and improving the teaching and learning qualities and processes. To enhance the effectiveness of evaluation facets, a wide range of data collection methods could be utilized, such as questionnaires, interviews, observation checklists, text analyses, and examinations all of which are geared toward enhancing the effectiveness of such facets.

In the Middle East and North Africa (MENA) settings, many stakeholders (e.g., educators, policy-makers, students, teachers, parents, and administrators) have cautioned against the overlooking of evaluation. The MENA region has gone through different and significant educational changes with pinned hopes on attaining better education quality and, therefore, meeting international standards. While there is a plethora of studies on evaluation, there is a paucity of research on the status of evaluation in the MENA context, especially evaluation of English language programs, examinations, text genre analyses, assessment accountability (Hart, Diercks-O'Brien, & Powell, 2009), learning, teaching and quality assurance. Unfortunately, such types of evaluation have not gained momentum for practical and most often political reasons. A Google search on evaluation in the MENA context yields fewer, if not, poorer results that could not amount to official evaluation enterprises of educational programs. There is a dramatic scarcity of research on evaluation in the MENA region.

This book presents myriads of evaluation themes the first of which is *Teacher and Faculty Staff Evaluation*. In Chapter “[Teacher Evaluation: What Counts as an Effective Teacher?](#),” Mazandarani and Troudi talked about the multidimensional aspects of evaluation in shaping the teacher profile that is perceived as central to the attention of policy-makers and administrators alike. This exploratory study looked into teachers’ perceptions of how an “effective second/foreign language teacher” should be. Results showed five major categories of what it means to be an effective teacher and perhaps what is relevant about this chapter is its model of effective teaching that embraces teachers’ personal, cognitive, metacognitive, pedagogical, and professional skills deemed necessary for teacher development. In the same vein, Alamouldi and Troudi, in Chapter “[EFL Teacher Evaluation: A Theoretical Perspective](#),” tackled teacher evaluation and its relevant importance in signposting

the necessity for teachers to keep abreast of the different trends that promote teacher development. This chapter suggested different ways in which teachers could be evaluated. Al-Fattal, Chapter “[Faculty Performance Evaluation and Appraisal: A Case from Syria](#),” foregrounded performance evaluation of academic staff members in Syria and its relevance to the educational context. The study triangulated different sources of data collection tools and methods, such as interviews, documentary analysis, and observation. One of the results relevant to this study was the idea of “judgmental” evaluation. The author called for a recommendation tip: Standardizing evaluation procedures that would help faculty members to be aware of the relevance and importance of performance evaluation.

The second important part of this book is concerned with *Assessment Practices*. In Chapter “[Ethicality in EFL Classroom Assessment: Bridging the Gap between Theory and Practice](#),” Torky and Haider, in an empirical study in Egypt, stressed the notion of ethicality in EFL classroom-based assessment context and its relation to test fairness. Based on a fifty-item questionnaire on teachers’ perceptions of ethicality, findings of this study demonstrated that such concepts were controversial, thus calling for the use of multiple sources of measuring students’ ability to set up some matching between the assessment methods, curriculum objectives and classroom activities. For the authors, one of the practical ways to reach assessment fairness is teacher training. In Chapter “[Problematising Teachers’ Exclusion from Designing Exit Tests](#),” another equal ethical issue related to assessment fairness was investigated by Dammak who questioned the exclusion of teachers from designing examinations, such as exit tests in the UAE context. This issue was undertaken from the perspective of two main stakeholders: teachers and policy-makers. Qualitative results bespoke that teachers’ self-awareness of assessment was relevant in evaluating the course objectives. This assessment literacy stood in sharp contrast with the assumptions that teachers lacked testing competence, thus denying the motif to exclude them from test design. Albaiz, in Chapter “[The Voice of Classroom Achievement towards Native and Non-native Educators in English Language Teaching: An Evaluative Study](#),” evaluated the students’ class performance as undertaken by native and non-native speakers (NNS) in a Saudi context. Results indicated that NNS teachers were likely to face different problems related to the misconceptions of some subject-matter teaching key concepts. To remedy this, the author suggested the implementation of appropriate teaching strategies to help learners to develop their language ability in their learning environment.

Part III of the book addresses *Text Genre Analysis Evaluation* by presenting two controversial cases from Tunisia. In Chapter “[Evaluation of Generic Structure of Research Letters Body Section: Create a Research Letter Body Section Model](#),” Melliti focused on the Evaluation of the Organizational Structure of Research Letters sections. Based on a sentence-by-sentence content analysis, the author found that a “Create A Research Letter Body Model (CARL)” contained 58 sentences of which 49 are obligatory. Implications of this study could be related to curriculum design and implementation as well as the teaching of writing to ESP researchers, students, and curriculum developers. In the same vein, text genre

analysis evaluation, in Chapter “Genre Analysis and Cultural Variations: A Cognitive Evaluation of Anglo-American Undergraduate Personal Statements,” in a contrastive analysis of 60 British ($n = 30$) and American ($n = 30$) personal statements written by undergraduates in different disciplines, such as business, physics, and biology, Hajji maintained that both Anglophone undergraduates yielded rhetorical and linguistic similarities and differences in the analyzed corpora, which was due to the sociocultural context where these undergraduates were operating. This study is also relevant in probing the genre aspects of personal statements.

In Part IV, *Assessment of Productive Skills*, Ben Maad, in Chapter “Learner Differences: A Trojan Horse Factor in Task-Based Oral Production Assessment?,” tackled the role of learner differences in an oral production assessment mode from a task-based assessment approach. The author listed three assessment criteria in any oral performance: Fluency, accuracy, and complexity necessary for a successful speaking production. In Chapter “Assessing ESL Students’ Paraphrasing and Note-Taking,” Soheim, in assessing paraphrasing and note-taking, spotlighted the necessity to go through different steps to design a writing test for undergraduate ESL learners in Egypt, such as initial planning, test specs, and note-taking skills. All such phases were meant to make students aware of the necessity to avoid plagiarism, while concentrating on the obligation of academic writing through the implementation of paraphrasing and note-taking for Arab learners in an ESL program. In Chapter “Criteria for Assessing EFL Writing at Majma’ah University,” Yahya studied the necessity of implementing quality assurance standards in the Saudi universities considered as relevant for teaching and assessment purposes. The author wondered whether the students’ writings were standardized using defined marking rubrics from three perspectives: Instructional experience of the faculty, academic levels of writing courses, and type and nature of writing. Findings showed that there were no statistically significant differences in the implementation of the marking rubrics.

In Part V of the book, *Textbook and ICT Evaluation*, Hermessi, Chapter “An Evaluation of the Place of Culture in English Education in Tunisia,” evaluated the cultural instances found in official educational documents, eight textbooks, and seven teacher guides produced by Tunisians who were involved in designing curricula and materials design in English. Results pointed that curriculum developers and textbook writers did not have any preconceived ideas on excluding culture from the English program although approaching such cultural instances was done in a non-systematic way. In Chapter “Evaluation of ICT Use in Language Education: Why Evaluate, Where to Look, and with What Means?,” Derbel reviewed the status of ELT from a computer-assisted language learning (CALL) angle. The study highlighted the intricacies of learning and teaching in an ICT environment, and it concluded by mentioning future research venues for evaluating ICT in the MENA educational context.

Part VI deals with *Evaluation of ELT Certificates and Programs* in Sudan. In Chapter “Evaluating the Certificate of Teaching English as a Foreign Language (CTEFL): A Way to Quality,” Nur and ElSaid Mohamed probed the evaluation

of the *Certificate of Teaching English as a Foreign Language* (CTEFL) administered by the graduate unit in the English Language Institute in Khartoum, Sudan. Data from a questionnaire survey revealed that the service beneficiaries (students) were satisfied with the program, since it contained interesting modules that met their needs. Students also praised the quality of their instructors. In the same context, Alhassan and Holi Ali, in Chapter “[An Evaluation of the Challenges of Sudanese Linguistics and English Language-Related Studies’ Ph.D. Candidates: An Exploratory Qualitative Study](#),” in an exploratory qualitative study, evaluated ELT studies of the Sudanese Ph.D. candidates. The study put focus on different facets to improve the quality of Ph.D. supervision in this very context. Results suggested that the quality of this supervision was fraught with myriads of problems and challenges among which were scarcity of resources and organization. What was missing from this chapter was whether the nature of the supervisors’ expertise had a significant impact on the quality of dissertation.

Part VII of the book, evaluation of *Quality Assurance and ESP Needs Analysis*, Staub, Chapter “[Quality Assurance and Foreign Language Programme Evaluation](#),” addressed the relevance of quality assurance and foreign language program evaluation in Turkey. Many stakeholders, such as universities, faculties, and instructional programs, were striving to achieve institutional quality that was valued by external and internal stakeholders. In addition, the author stressed the fact that both quality assurance and evaluation were increasingly becoming “critical activities for EFL programs wishing to demonstrate their worth.” In Chapter “[Evaluation in Tunisia: The Case of Engineering Students](#),” Jamly investigated the status of evaluating needs analysis of engineering students from an ESP dimension. Based on data collection on the perceptions of the TOEIC test in an engineering program in Tunisia administered to students and teachers a results denoted that current employees, who were former students of this engineering program, did not think of the TOEIC as an important test to evaluate their language ability in English, since, according to their replies, the test contents did not meet their needs. The author recommended that perceptions of evaluation embrace the needs of all stakeholders from an outcome-based measuring perspective “to build the bridge between learning objectives and learner evaluation in ESP.”

In Part VIII of the book, *Assessment Literacy and Dynamic Assessment*, Bouziane, in Chapter “[Why Should the Assessment of Literacy in Morocco Be Revisited?](#),” investigated the reading and writing assessment literacy in Morocco, which could be a replica of the testing situation in North African, such as Tunisia, Libya, and Algeria. The author stressed the negative washback effect of assessment on students “both during and after school.” Unfair types of assessment, as the author claimed, rested upon two major aspects: The use of a very narrow range of skills in testing reading, which could be amounted to the notion of construct fuzziness, and teachers’ rating inconsistencies in writing. To remedy these shortcomings, the author called for a more comprehensive definition of the reading and writing constructs to include other types of test items that could be objectively scored. In addition, for this similar situation to be improved, the author highlighted the fact that there should be a reconsideration of ELT research and teacher training in how to produce useful and fair tests in similar-related contexts. In Chapter “[Specs Validation of a Dynamic Reading Comprehension Test for](#)

[EAP Learners in an EFL Context](#),” Hidri evaluated the theoretical and practical aspects of designing dynamic assessment specs of a reading comprehension examination for learners of English in an Omani context. Using both qualitative and quantitative approaches, results of the study showed that students’ performance improved in the presence of mediation and support. The study concluded with highlighting a list of specs that test designers, in similar contexts, might consider to help their learners overcome their testing difficulties through appropriate use of dynamic assessment.

In the MENA context, evaluation challenges are numerous. For instance, one of the challenges rests on the stakeholders’ ability to implement specific evaluation standards that would serve the interests of all parties. Evaluation has to be contextualized and relevant and it has to be regarded from this perspective. For instance, how can evaluation be employed to improve the curriculum and other related programs? The current situation in the MENA context should consider different parameters to enhance the relevance of evaluation to different stakeholders, such as teacher training, teaching content, teaching methodologies, students’ needs, curriculum design, and writing relevant test specs that meet the curriculum objectives and materials design. There should be a careful consideration of the people who should be well versed in evaluation to implement, collect, analyze, and report on data. Evaluation has to be contextualized by highlighting its purpose, content, usage, and method (Nygaard & Belluigi, 2011) and its results should be treated with caution to avoid any misuse or harm.

The MENA context boasts itself for being “unified” at the level of language, culture, and religion. However, there exists a big challenge on whether the different stakeholders of the MENA context, especially the Arab countries, are capable of developing a common Arab framework of reference to improve the educational standards and practices. Research in this field should focus on student evaluation, teacher evaluation, and summative versus formative evaluation. In addition, further research is needed to evaluate the quality of the graduate programs, given the eventuality of the increasing number of graduates sitting for their master’s and Ph.D. programs everywhere in the MENA context. There should be an open debate among these countries to maintain some educational sustainability and improvement. Many challenges for evaluation appear straightforward, such as the sociocultural context, the right people to implement evaluation, and the different ways to align evaluation to well-defined parameters that reflect the very sociocultural context of the MENA region. Even though this book bears great significance to evaluation in general, still addressing other facets of evaluation in the MENA context remains unexplored.

Sahbi Hidri
English Language Institute, University of Jeddah
Jeddah, Saudi Arabia; and
Faculty of Human and Social
Sciences of Tunis, Tunis, Tunisia

Christine Coombe
Higher Colleges of Technology
Dubai, United Arab Emirates

References

- Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Bailey, K. M., & Brown, J. D. (1996). Language testing courses: What are they? In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 236–256). Clevedon, UK: Multilingual Matters.
- Brown, J. D., & Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing*, 25(3), 349–384.
- Campbell, B., & Mark, M. M. (2015). How analogue research can advance descriptive evaluation theory: Understanding (and improving) stakeholder dialogue. *American Journal of Evaluation*, 36(2), 204–220. doi:[10.1177/1098214014532166](https://doi.org/10.1177/1098214014532166)
- Golding, C., & Adam, L. (2014). Evaluate to improve: Useful approaches to student evaluation. *Assessment and Evaluation in Higher Education*, 41(1), 1–14. doi:[10.1080/02602938.2014.976810](https://doi.org/10.1080/02602938.2014.976810)
- Hart, D., Diercks-O'Brien, & Powell, A. (2009). Exploring stakeholder engagement in impact evaluation planning in educational development work. *Evaluation*, 15(3), 285–306. doi:[10.1177/1356389009105882](https://doi.org/10.1177/1356389009105882)
- Hidri, S. (2014). Developing and evaluating a dynamic assessment of listening comprehension in an EFL context. *Language Testing in Asia*, 4(4). doi:[10.1186/2229-0443-4-4](https://doi.org/10.1186/2229-0443-4-4)
- Hidri, S. (2015). Conceptions of assessment: Investigating what assessment means to secondary and university teachers. *Arab Journal of Applied Linguistics*, 1(1), 19–43.
- Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27(4), 555–584. doi:[10.1177/026553220935143](https://doi.org/10.1177/026553220935143)
- Kiley, R., & Rea-Dickins, P. (2005). *Program evaluation in language education*. Palgrave: Macmillan.
- Kyvik, S., & Thune, T. (2015). Assessing the quality of Ph.D. dissertations. A survey of external committee members. *Assessment and Evaluation in Higher Education*, 40(5), 768–782. doi:[10.1080/02602938.2014.956283](https://doi.org/10.1080/02602938.2014.956283)
- Ladyshevsky, R., & Taplin, R. (2015). Evaluation of curriculum and student learning needs using 360-degree assessment. *Assessment and Evaluation in Higher Education*, 40(5), 698–711. doi:[10.1080/02602938.2014.950189](https://doi.org/10.1080/02602938.2014.950189)
- Lantolf, J. L., & Poehner, M. E. (2011). Dynamic assessment in the classroom: Vygotskian praxis for second language development. *Language Teaching Research*, 15(1), 1–23. doi:[10.1177/1362168810383328](https://doi.org/10.1177/1362168810383328)
- Leung, C., & Rea-Dickins, P. (2007). Teacher assessment as policy instrument: Contradictions and capacities. *Language Assessment Quarterly*, 4(1), 6–36. doi:[10.1080/15434300701348318](https://doi.org/10.1080/15434300701348318)
- Nygaard, C., & Belluigi, D. Z. (2011). A proposed methodology for contextualized evaluation in higher education. *Assessment and Evaluation in Higher Education*, 36(6), 657–671. doi:[10.1080/02602931003650037](https://doi.org/10.1080/02602931003650037)
- Shiel, C., Filho, W. L., do Paço, & Brandli, L. (2015). Assessing and evaluating sustainable development in higher education. *Assessment and Evaluation in Higher Education*, 40(6), 783–784. doi:[10.1080/02602938.2015.1073028](https://doi.org/10.1080/02602938.2015.1073028)
- Swanwick, T. (2007). Introducing large-scale educational reform in a complex environment: The role of piloting and evaluation in modernizing medical careers. *Evaluation*, 13(3), 358–370. doi:[10.1177/1356389007078624](https://doi.org/10.1177/1356389007078624)

- Wei, W. (2015). Using summative and formative assessments to evaluate EFL teachers' teaching performance. *Assessment and Evaluation in Higher Education*, 40(4), 611–623. doi:[10.1080/02602938.2014.939609](https://doi.org/10.1080/02602938.2014.939609)
- West, S. E. (2015). Evaluation, or just data collection? An exploration of the evaluation practice of selected UK environmental educators. *The Journal of Environmental Education*, 46(1), 41–55. doi:[10.1080/00958964.2014.973351](https://doi.org/10.1080/00958964.2014.973351)

Evaluation in Foreign Language Education in the Middle
East and North Africa

Hidri, S.; Coombe, C. (Eds.)

2017, XXII, 337 p. 28 illus., 18 illus. in color., Hardcover

ISBN: 978-3-319-43233-5