Formative assessment in review

This part presents the main characteristics of Formative Assessment (FA) with an emphasis on Language Learning (LL). It also exhibits the features of a Systematic Review (SR) and an Annotated Bibliography (AB) selected for the needs of this book. The aim is to set the background for the SR and AB on FA in LL, which is the focus of this book.

1.1. FA

Most of the history of assessment in education traces back to the Imperial Chinese system of examinations (Spolsky, 2008), which deals with testing. As Spolsky (2008) has argued, "language testing grew up against this background" (p. 5). Consequently, the developments and discussions about the history of language assessment deal mostly with the history of testing. A good example is O'Sullivan's (2012) chapter *A Brief History of Language Testing* (pp. 9-19).

Testing and assessment are often treated as synonyms. This is evident in cases, where the word *assessment*, and/or *testing* are used in publications' titles but in reality, the publication concentrates on testing. The following examples are indicative: Davies's (2013) *Fifty years of language assessment;* Spolsky's (2016) *Language assessment in historical and future perspective*; and Farhady's (2018) *History of language testing and assessment*.

In its history, language assessment has followed the developments in assessment in education and in theory. As described by Farhady (2018), various perspectives and issues resulted in drastic changes and in shifting the attention to finding alternatives to assessing language ability in the context

of and during the process of learning. One could observe this over the years as assessment moved from the prescientific to the scientific, structuralist, integrative sociolinguistic to the communicative era of language teaching and learning, and faced pedagogical implications. It moved from what the student knows to what the student can do. These alternative types of learning involved student cooperation, the assumption of more responsibility in their learning, and encouraged the application of knowledge to solving real-life problems. Alternative assessment goes beyond traditional forms of assessment such as tests and high-stakes examinations. Similarly, FA evaluates during the process in the form of, for example, classroom polls, exit tickets, and early feedback, and not in the form of midterm exams, end-of-unit or chapter tests, final projects or papers, district benchmark and scores, after the learning process as SA does; FA monitors the learning process rather than assigning grades, it aims to improve student's learning rather than evaluating student's achievements, it focuses on little content areas rather than complete chapters or content areas, and it considers evaluations as a process and not as a product as summative assessment does (Renard, 2017).

For the purposes of this book, we first examine the various definitions of FA and its characteristics given in the course of its history; the aim of this was to establish a definition that incorporates the most common characteristics discussed by researchers, which could then be the base for the SR and AB which constitute parts of this book.

Scriven (1967) was the first to suggest two roles evaluation may play. He suggested that evaluation

> "may have a role in the on-going improvement of the curriculum [… and] may serve to enable administrators to decide whether the entire finished curriculum refined by the use of the evaluation process in its first role, represents a sufficiently significant advance on the available alternatives to justify the expense of adoption by a school system" (Scriven, 1967, pp. 41-42).

To define these two roles in relation to curriculum evaluation and distinguish their differences, he proposed "the terms 'Formative' and 'Summative' evaluation" (Scriven, 1967, p. 43).

Not long after Scriven's (1967) definitions, Bloom (1969) stated that the same terms can be used not only to evaluate curriculum but also to evaluate students' learning. While acknowledging the value of summative evaluation of student learning, Bloom (1969) has also emphasised the value of formative evaluation. He saw 'formative evaluation' as a way "to provide feedback and correctness at each stage in the teaching-learning process" (Bloom, 1969, p. 48). Both Scriven (1967) and Bloom (1969) supported that the information given during formative evaluation helps in making changes in the teaching and learning activities during the learning process.

Since then, a lot of other researchers have attempted to define SA and FA and their differences. Saito and Inoi (2017) support that some classroom assessments, such as midterms, finals, and large-scale achievement tests, are 'inherently' summative because they are administered with summative intention, whereas most other types of classroom assessments are inherently formative. According to Bennett (2011) and Liu (2015), SA is a one-time opportunity where a student can demonstrate their knowledge. Brookhart (2010) maintains that the division between FA and SA is still blurred, as assessment can be used for both summative and formative purposes.

Black and Wiliam have been contributing to the evolution of FA since the 1990's. Their wide-ranging literature review claimed that "conclusively […] FA does improve learning" (Black & Wiliam, 1998, p. 61). They identified the main features of FA as sharing criteria with learners, developing classroom talk and questioning, giving appropriate feedback, and peer and self-assessment. In their review, FA "is to be interpreted as encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (Black & Wiliam, 1998, pp. 7-8). Black and Wiliam (1998)

and Boston (2002) referred to FA as ongoing: during the assessment process, teachers gather evidence of the student's learning, which they use to adapt their teaching so that it meets students' needs and diagnoses their progress toward a long-term objective.

FA has been described as an 'assessment for learning', while SA has been defined as an 'assessment of learning' (Rea-Dickins & Gardner, 2000). This means that students need to be given continuous information about their own learning, how they are progressing, the nature, scope, and level of their learning, and in which areas improvement is needed. Both types of assessment are equally valuable and significant in the learning process and complement each other. Gattullo (2000, p. 279) characterised FA as an ongoing multi-phase process that is carried out on a daily basis through teacher-pupil interaction with the provision of feedback.

'Assessment for learning' has also been described as "the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there" (Broadfoot et al., 2002, p. 3). In Wiliam (2011), 'assessment for learning' is defined as "any assessment for which the first priority in its design and practice is to serve the purpose of promoting students' learning. It thus differs from assessment designed primarily to serve the purposes of accountability, or of ranking, or of certifying competence" (p. 10). Stiggins (2005) recognised FA as a diagnostic test, however, Popham (2006) argued that FA is not a test, it is a process. Cizek (2010) has claimed that not all these characteristics should be met in order for assessment to be formative. According to Black and Wiliam (2018), the teacher elicits "evidence of students' understanding and based on that evidence takes decision for next steps for effective instruction" (p. 8). Teachers need to be aware of what students understand from the learning experience. Tan (2013) also aimed "to identify the minimal requirements for FA to succeed in terms of assessment standards, assessment design, and assessment feedback" (p. 1). Bahati, Tedre, Fors, and Evode (2016) support that assessment can only be considered *formative* if it can generate feedback that students can use to improve their learning and achievements. Additionally, it also needs to be used by teachers to re-evaluate and reflect on teaching strategies in response to

their learners' needs, which means FA occurs during the course of learning. It aims to determine the next steps by giving feedback to students and establishing students' needs and progress in order to modify aspects such as planning, curriculum design, content, learning experiences, and resources for the benefit of students' learning. It has been practised much less and only in the last decades during the history of assessment (Heineke & McTighe, 2018).

FA is an informal type of assessment, as opposed to formal assessment. Formal assessment typically means a test or examination that involves standardised administration, for example, end-of-chapter tests, end-of-semester tests, or high-stakes examinations. Informal assessment is a process of obtaining information that can be used to make judgements about students' progress and make improvements in the learning processes. Informal assessments include, e.g. projects, presentations, experiments, demonstrations, or performances (Ketabi & Ketabi, 2014). They can include portfolios, asking questions during class, or informal observations of interaction, quizzes, rubrics, discussions, and self and peer assessment techniques (Ketabi & Ketabi, 2014) for FA purposes in order to improve the learning processes and learning.

In reality, FA is considered an alternative type of assessment, which refers to assessments, alternative to traditional ones, that offer a variety of measurement ways designed to understand what a student can do rather than what they know. Alternative assessment measures proficiency in relation to knowledge application rather than recitation or memorisation and includes designated projects, portfolios, observations, performance tasks, exhibitions, demonstrations, journals, reflective pieces, case-based scenarios, reports, teacher-created tests, rubrics, and self- and peer-evaluation. Bahrani (2011) also mentions interviews and the implementation of a number of Web 2.0 tools. They encourage critical thinking, collaboration, and information synthesis. They derive, reflect, and focus more on what learners can do in authentic-like real-life like, contextualised tasks (Papadima-Sophocleous, 2017). Moreover, the elements of alternative assessment can provide the learner with the opportunity to show what they can do with the language with innovative teaching approaches and techniques (Rea-Dickins, 2004), learning experiences, resources, modified curriculum design,

and content; this process can be characterised as a response to the traditional test-based assessment.

FA is also considered a classroom assessment approach. According to this type of assessment, teachers attempt to find out what and how well the learners understand during the lesson and to improve the quality of students' learning by making new decisions, which would facilitate improvement in the learning process (Angelo & Cross, 2012). According to Heineke and McTighe (2018), the following have been recorded as used for FA purposes: formal data collection such as quizzes, academic prompts, and second informal like classroom observations, dialogues, self-, and peer-assessment checks for understanding. Teachers integrate multiple opportunities to collect evidence in order to monitor learners' progress throughout the learning process. Black and Wiliam (2009) mention in their article *Developing the theory of formative assessment*, FA can also give rise to effective changes with the integration of interactive feedback.

Since Black and Wiliam's (2009) review on classroom assessment and learning was published, the authors continuously contributed in the area of assessment. In one of their latest publications (Black & Wiliam, 2018), they propose a model design of educational activities and argue that assessment is influenced by a combination of the theories of pedagogy, instruction, learning, and the subject discipline, along with the wider context of education. This indicates that FA practices are considered as one of the most motivational modes to increase students' engagement and performance (Ketabi & Ketabi, 2014). In their critical review of research on FA, Dunn and Mulvenon (2009) have tried to capture the scientific evidence of the impact of FA in education. They have argued that, although FA is important, limited empirical evidence exists to support the 'best practises' for formative evaluation.

1.2.   FA in LL

The first publication that referred to FA in LL was by Rea-Dickins and Gardner in 2000. Before the year 2000, not much had been recorded regarding FA in

LL, neglect noted by Rea-Dickins and Gardner (2000). In the year 2000, Rea-Dickins and Gardner put emphasis on the characteristics of FA. They argued that if teachers' decisions are made responsibly during the language lesson, this will increase students' performance (Rea-Dickins & Gardner, 2000). Cheng, Rogers, and Hu (2004) also argued that the study of the assessment practices in the field of English LL and teaching were limited. The claims of neglect continued when Fakeye (2016) stated that FA is an overlooked type of assessment since language teachers pay more attention to SA that includes tests and scores. After the year 2000, studies began to focus on FA and English as a Second Language/English as a Foreign Language (ESL/EFL) and paid more attention to aspects such as the process and design of learning, the curriculum design, the learning experiences, content, and planning teachers needed to employ for this type of assessment. Bachman and Palmer were among the first who mentioned that giving feedback to language students may support their performance in formal tests. They also emphasised the relationship between FA and formal tests in language education (Bachman & Palmer, 2010).

Gradually, FA started being globally recognised as an essential element in the language area; however, it continued to face implementation issues and the need for more research continued to be evident. Heitink et al. (2016), for instance, have argued that there is limited scientific evidence on the positive impact of FA. Additionally, the different conceptualisations of FA and their understanding made FA applications more difficult and the research of FA applications in LL more challenging. At the same time, FA faced some practical issues; for example, emphasis was put by teachers who were used to practicing SA, marking and scores, especially in crowded classes instead of providing feedback, and did not agree on how FA is given (Ketabi & Ketabi, 2014).

As pedagogy progressed with time, the integration of technology in education added opportunities for a supportive environment to implement FA practices for LL, provided it is aligned with the learning theory, language teaching/ learning, and FA principles (Vassiliou & Papadima-Sophocleous, 2019). Technology-enhanced teaching and learning allow the teacher and the learner to use a number of tools for FA purposes that would enable to increase the

L2 learning outcomes (Perera-Diltz & Moe, 2014). Technology-enhanced learning can accommodate better environments for effective, instant, and meaningful feedback (Heinrich, Milne, & Moore, 2009), depending on its nature and quality. Furthermore, the proliferation of technology tools in L2 has the potential to support the role of FA and enable learners and educators to use technologies not only for score tests and exams but also for meaningful FA. E-assessment tools, like online quizzes, can also be used following the FA principles and provide comprehensive and on-time feedback to students, and most interestingly they can monitor their understanding (Baleni, 2015). According to the literature, examples of e-tools that can be used for FA purposes and can provide e-feedback are: (1) Turnitin and Grademark, (2) Electronic Feedback Software, (3) Questionmark Perception, (4) WebCTConnect, (5) MarkTool, (6) Markin (http://www.cict.co.uk/software/markin/index.htm), (7) Moodle Quiz, and (8) Markers Assistant (Heinrich et al., 2009). Another example of an online tool that offers the opportunity for instant and effective feedback and can be used for FA purposes, is Google Docs where teachers and students can discuss and exchange ideas synchronously on a shared document (Reimann, Halb, Bull, & Johnson, 2011).

Additionally, other e-applications offer opportunities to students for self and peer feedback as part of FA like the Online Peer Feedback (OPF) application, (Rosalia & Llosa, 2009). Furthermore, it has been argued that there are some technology-enhanced tools that can be used to support FA integrations. Examples of such tools include e-journals, e-reflections, e-portfolios, e-rubrics, e-can-do lists, and e-artefacts, with the use of Google documents and Google Sites (Papadima-Sophocleous, 2017). Google Forms can also be used as an excellent type of e-Exit ticket cards and/or as a type of self-reflections. Responses from such tools assist teachers in their planning of subsequent lessons according to students' understanding (Exit Ticket, n.d.).

According to the literature, the use of iPods and iPads can be used for FA in LL (Levy & Gertler, 2015; Medina & Hurtado, 2017). Students can make videos or audio recordings and improve their speaking and listening skills. Many studies highlight the importance of gamified quizzes and online assessment tools like

Socrative, Kahoot, Eclipse, Quizlet, Edmodo, Padlet, Storify, Google Doc, Google Forms, and Remind 101 as FA tools in LL, which can be used during the learning process, and give instant feedback to students; in order to modify and improve their curriculum design, content, their student's learning experiences and their learning (Heinrich et al., 2009; Reimann et al., 2011).

As established so far, FA is both conceptually and practically still shaping. However, in order to proceed in conducting an SR and an AB which focus on FA, some foundation needed to be set. According to Bennett (2011, p. 6), in order to provide the field with a meaningful definition of FA, we need (1) a theory of action and (2) a concrete instantiation. The theory of action can identify the characteristics and components of the entity we are claiming FA is, along with a rationale for each characteristic and component; and proposes how these characteristics and components can work together in order to create a desired set of outcomes.

Our theory of action stems from the literature review we conducted regarding different stands on FA, and revolves around the idea of involving both students and teachers during the students' learning: engaging in effective discussions, interaction, criteria development, giving feedback, engaging in teacher, self and peer assessment, and gathering evidence of learner's learning, which they would use to adapt both the teaching and the learning in order to diagnose students' needs and progress in a continuum, toward a long-term objective.

For the purpose of this book, we have adopted the following FA features, drawn from earlier research, in the hope that each reviewed and annotated publication would fully or partially reflect them.

FA characteristics taken into consideration in this volume are as follows.

- It is classroom (Can Daşkın & Hatipoğlu, 2019) and school based in contrast to high-stakes examinations which are externally based.

- It involves students, the teacher, and peers (Carless, 2002).

- It supports learning and assessment for forming learning in many ways; the theoretical background of FA is aligned with current learning theories, such as constructivism. Constructivist learning practices, for example, are in line with the nature of formative evaluation as described by Stiggins (2005), stating that formative evaluations enable learning to be guided according to the student's ability level. It provides students with opportunities for active involvement in their own learning in an environment where both students and teachers are engaged (Heitink et al., 2016).

- It is part of the learning and teaching process, it gathers information from them (Gan & Leung, 2020; Shepard, 2006) to further improve "the instructional decisions that are made by teachers, learners" (Wiliam, 2011, p. 13) or their peers, unlike SA which is usually administered to categorise students' performances or for accreditation (Cizek, 2010, p. 1).

- It helps students "understand learning objectives and become aware of strategies and steps to be undertaken in order to move their learning forward" (Gan & Leung, 2020, p. 2).

- It provides opportunities for giving feedback by an agent (e.g. teacher, peer, self), engaging in teacher, self and peer assessment, and gathering evidence of learner's learning, which they would use to adapt both the teaching and the learning in order to diagnose students' needs and progress in a continuum, toward a long-term objective (Bachman & Palmer, 2010).

- It establishes what students know while they are still in the process of learning it (Broadfoot et al., 2002).

- It materialises in classroom-based practices that range from e.g. observations, class discussions, peer- and self-assessment, feedback, moment-by-moment teacher decisions and responses, and construction

of artefacts, etc. These can be in paper or technology and web-based FA form, with the latter offering interactivity, real-time practice, multimedia features, timely feedback, variety of formative exercises, own pace learning, provision of multiple attempts (Buchanan, 2000; Jia, Chen, Ding, & Ruan, 2012), and gaming features and strategies (Wang, 2008).

- It includes a collection of evidence of performances over time to provide evidence of growth and learning; it is closely related to teachers' day-to-day work of teaching and learning and assessment *for* learning (Scarino, 2013, p. 312).

- It supports and facilitates the process of learning before SA comes to verify at the end of a learning process what learning has been achieved and whether the learning outcomes have been met.

As one can observe, the FA and SA boundaries are not clear. Some assessment applications can be used in both. What needs to be clear is that it is the purpose of the assessment that helps in defining and determining whether it is formative or summative. The above features aim to assist in deciding when the intention is to assess students formally.

The examination of the research conducted so far on FA has established that the definition of FA is not yet completed, it is in its making. However, in order to record both in the SR and the AB the research activities in the area of FA in LL during the designated period, we had to come up with a minimum framework of the main characteristics of FA as discussed so far in the literature. With the recording of the sources between 2000 and 2020, however, we established that this is an area that offers itself for further future research.

So far, we have presented the characteristics of FA in general and described its presence and contribution in LL in particular. Preliminary research helped in establishing that there is no comprehensive SR, nor any AB in FA in LL.

The results revealed the existence of some literature reviews as a part of research papers, or more generic research focused on FA implications and not on FA and LL (Allal & Lopez, 2005; Black & Wiliam, 1998; Dunn & Mulvenon, 2009; Gikandi, Morrow, & Davis, 2011). As a result, the aim of this research was to examine the area of FA in LL through first an SR and then an AB in FA in LL in order to establish a more solid background in the area. In order to conduct an SR and an AB, an investigation of their characteristics was conducted in order to clearly set the parameters of the study.

1.3.   SR

An SR is based on a research review design and can be of a qualitative, quantitative and/or mixed research approach. The main purpose of SR is to synthesise different studies which are related to a specific research area (Hanley & Cutts, 2013). SR is different from a narrative traditional type of research as it critically summarises and syntheses all data related to a topic, and focuses on systematic research of the literature (Štrukelj, 2018). As with other research designs, SR follows a specific protocol, meaning it has a set of characteristics that one follows. Some of these characteristics are: (1) a clearly stated set of objectives; (2) a presentation of one or more research questions; (3) an explicit, reproducible methodology; (4) a set of clearly defined criteria for inclusion/exclusion of the relevant studies; (5) a systematic search for identification of studies that would meet the eligibility criteria; (6) a systematic presentation and synthesis of the findings, making comparisons, associations, or identifications of new research areas; and (7) assessment of the validity of the findings (Hanley & Cutts, 2013). According to Norris and Ortega (2007), the strengths of systematically reviewing applied linguistics are promising, in comparison to narrative literature reviews. It can reveal gaps, weaknesses, and needs in a research area. For that reason, Norris and Ortega (2007) encourage applied linguists to adopt this research design and "to think and act systematically" (p. 813).

The above research review design characteristics and guidelines were followed in Part 2 of this book to conduct the SR on FA in LL.

1.4.   AB

An AB provides an overview of available research sources (Engle, 2017), of the main issues, arguments, and research within a particular area. This list of works is formatted according to a specific documentation style (e.g. MLA, APA, etc.) (Saint Mary's University, 2019). The content of the AB can be listed alphabetically by the author or arranged chronologically by publication date. In the introduction, the topic or subject area covered by the bibliography is described, and the method used to identify possible sources, the rationale for selecting the sources, and, if appropriate, an explanation describing the reasons for exclusions of some types of resources are explained (Harner, 2015). This introduction is then followed by the citation, according to the specific chosen documentation style, followed by an annotation, a summative paragraph that evaluatively describes the content of the source.

An AB focuses on the importance of each source in relation to the topic (Buttram, MacMillan, & Koch, 2012). It pays particular attention to the content and contribution of each individual source to the given area of research. Each entry can be defined as a brief explanatory or evaluative note of each reference or citation (Buttram et al., 2012). An annotation can be helpful to researchers in informing them about the source and evaluating whether the source is relevant to a given topic or line of inquiry (Engle, 2017).

The above characteristics and guidelines were followed in Part 3 of this book to conduct the AB on FA in LL.

1.5.   Why combine SR and AB?

The reason for combining the two different research designs conducted during the specific time period under study was to give as much information as possible about the publications on FA in LL. The research is based on specific research questions, and a systematic evaluation of FA in LL studies of 2000-2020, with the use of inclusion and exclusion criteria. The aims were:

- to synthesise, analyse, and interpret all the data by making comparisons, associations, or identifying new research areas (Hanley & Cutts, 2013); and

- to then focus on the importance of each source in relation to the topic (Buttram et al., 2012), pay particular attention to the content and contribution of each individual source, by briefly describing and evaluating explanatorily each reference or citation (Buttram et al., 2012).

The first step towards that was to have an overview of this activity through an SR of the research carried out during this period.

References

Allal, L., & Lopez, L. M. (2005). Formative assessment of learning: a review of publications in French. In J. Looney (Ed.), *Formative assessment: improving learning in secondary classrooms* (pp. 241-264). Organisation for Economic Cooperation and Development. http://www.oecd.org/edu/ceri/34488354.pdf

Angelo, T. A., & Cross, K. P. (2012). *Classroom assessment techniques*. Jossey Bass Wiley.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: developing language assessments and justifying their use in the real world*. Oxford University Press.

Bahati, B., Tedre, M., Fors, U., & Evode, M. (2016). Exploring feedback practises in formative assessment in Rwandan higher education: a multifaceted approach is needed. *International Journal of Teaching and Education*, *4*(2), 1-22. https://doi.org/10.20472/TE.2016.4.2.001

Bahrani, T. (2011). Technology as an assessment tool in language learning. *International Journal of English Linguistics*, *1*(2), 295-298. https://doi.org/10.5539/ijel.v1n2p295

Baleni, Z. G. (2015). Online formative assessment in higher education: its pros and cons. *The Electronic Journal of e-Learning*, *13*(4), 228-236. https://academic-publishing.org/index.php/ejel/article/view/1730

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5-25. https://doi.org/10.1080/0969594X.2010.513678

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, Principles, Policy & Practice*, *5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31. https://doi.org/10.1007/s11092-008-9068-5

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in education: principles, policy & practice, 25*(6), 551-575. DOI: https://doi.org/10.1080/0969594X.2018.1441807

Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In R. W. Tyler (Ed.), *Educational evaluation: new toles, new means: the 63rd yearbook of the National Society for the Study of Education, part 2* (pp. 26-50). University of Chicago Press.

Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research & Evaluation, 8*(9), 1-4. https://doi.org/10.7275/kmcq-dj31

Broadfoot, P., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart G. (2002). *Assessment for learning: 10 principles. Assessment Reform Group*. Nuffield Foundation and University of Cambridge. http://assessmentreformgroup.files.wordpress.com/2012/01/10principles_english.pdf

Brookhart, S. M. (2010). Mixing it up: combining sources of classroom achievement information for formative and summative purposes. In H. L. Andrade & G. J. Cizek (Eds), *Handbook of formative assessment* (pp. 279-296). Routledge.

Buchanan, T. (2000). The efficacy of a World-Wide Web mediated formative assessment. *Journal of Computer Assisted Learning, 16*(3), 193-200. https://doi.org/10.1046/j.1365-2729.2000.00132.x

Buttram, C., MacMillan, D., & Koch, R.T. (2012). *Comparing the annotated bibliography to the literature review.* UNA Center for Writing excellence. https://www.una.edu/writingcenter/docs/Writing-Resources/Comparing%20the%20Annotated%20Bibliography%20to%20the%20Literature%20Review.pdf

Can Daşkın, N., & Hatipoğlu, Ç. (2019). Reference to a past learning event as a practice of informal formative assessment in L2 classroom interaction. *Language Testing, 36*(4), 527-551. https://doi.org/10.1177/0265532219857066

Carless, D. R. (2002). The mini-viva as a tool to enhance assessment for learning. *Assessment & Evaluation in Higher Education*, *27*(4), 353-363. https://doi.org/10.1080/0260293022000001364

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: purposes, methods, and procedures. *Language Testing*, *21*(3), 360-389. https://doi.org/10.1191/0265532204lt288oa

Cizek, G. (2010). An introduction to formative assessment. In H. Andrade & G. Cizek (Eds), *Handbook of formative assessment* (1st ed., pp. 3-17). Routledge. https://www.routledge.com/Handbook-of-Formative-Assessment/Andrade-Cizek/p/book/9780415993203

Davies, A. (2013). Fifty years of language assessment. In A. J. Kunan (Ed.), *The companion to language assessment* (pp. 1-21). Wiley-Blackwell. https://doi.org/10.1002/9781118411360.wbcla127

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: the limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation*, *14*(7), 1-11. https://doi.org/10.7275/jg4h-rb87

Engle, M. (2017). *How to prepare an annotated bibliography: the annotated bibliography*. Cornell University Library. https://guides.library.cornell.edu/annotatedbibliography

Exit Ticket. (n.d.). The teacher tool kit. http://www.theteachertoolkit.com/index.php/tool/exit-ticket

Fakeye, D. O. (2016). Secondary school teachers' and students' attitudes towards formative assessment and corrective feedback in English language in Ibadan Metropolis. *Journal of Educational and Social Research, 6*(2), 141-148. https://doi.org/10.5901/jesr.2016.v6n2p141

Farhady, H. (2018). History of language testing and assessment. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching* (1st ed., pp. 1-7). John Wiley & Sons. https://doi.org/10.1002/9781118784235.eelt0343

Gan, Z., & Leung, C. (2020). Illustrating formative assessment in task-based language teaching. *ELT Journal, 74*(1), 10-19. https://doi.org/10.1093/ELT/CCZ048

Gattullo, F. (2000). Formative assessment in ELT primary (elementary) classrooms: an Italian case study. *Language Testing, 17*(2), 278-288. https://doi.org/10.1177/026553220001700210

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: a review of the literature. *Computers & Education*, *57*(4), 2333-2351. https://doi.org/10.1016/j.compedu.2011.06.004

Hanley, T., & Cutts, L. A. (2013). What is a systematic review? *Counselling Psychology Rev*iew, *28*(4)*,* 3-6. https://www.research.manchester.ac.uk/portal/files/32908640/FULL_TEXT.PDF

Harner, J. L. (2015). *On compiling an annotated bibliography* (2nd ed.). Modern Language Association of America.

Heineke, A. J., & McTighe, J. (2018). *Using understanding by design in the culturally and linguistically diverse classroom*. ASCD.

Heinrich, E., Milne, J. D., & Moore, M. (2009). An Investigation into e-tool use for formative assignment assessment-status and recommendations. *Educational Technology & Society*, *12*(4), 176-192. https://www.researchgate.net/publication/220017580_An_ Investigation_into_E-Tool_Use_for_Formative_Assignment_Assessment_-_Status_and_ Recommendations

Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, *17*, 50-62. https://doi.org/10.1016/j. edurev.2015.12.002

Jia, J., Chen, Y., Ding, Z., & Ruan, M. (2012). Effects of a vocabulary acquisition and assessment system on students' performance in a blended learning class for English subject. *Computers & Education, 58*(1), 63-76.

Ketabi, S., & Ketabi, S. (2014). Classroom and formative assessment in second/foreign language teaching and learning. *Theory & Practice in Language Studies*, *4*(2), 435-440. https://doi.org/10.4304/tpls.4.2.435-440

Levy, T., & Gertler, H. (2015). Harnessing technology to assess oral communication in Business English. *Teaching English with Technology*, *15*(4), 52-59. https://files.eric. ed.gov/fulltext/EJ1138436.pdf

Liu, X. (2015). The application of formative assessment in College English Teaching. *Proceedings of the 1st International Conference on Arts, Design and Contemporary Education* (pp. 875-879)*. ICADCE, 2015. https://doi.org/10.2991/icadce-15.2015.211

Medina, E. G., & Hurtado, C. (2017). Kahoot! A digital tool for learning vocabulary in a language classroom. *Revista Publicando, 4*(12)*, 441-449.

Norris, J. M., & Ortega, L. (2007). The future of research synthesis in applied linguistics: beyond art or science. *TESOL Quarterly*, *41*(4), 805-815. http://www.jstor.org/ stable/40264408

O'Sullivan, B. (2012). A brief history of language testing. In C. Coombe, P. Davidson, B. O'Sullivan, & C. Stoynoff (Eds), *The Cambridge guide to language assessment* (pp. 9-19). Cambridge University Press.

Papadima-Sophocleous, S. (2017). L2 assessment and testing teacher education: an exploration of alternative assessment approaches using new technologies. In K. Borthwick, L. Bradley & S. Thouësny (Eds), *CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017* (pp. 248-253). Research-publishing.net. https://doi.org/10.14705/rpnet.2017.eurocall2017.721

Perera-Diltz, D. M., & Moe, J. L. (2014). Formative and summative assessment in online education. *Journal of Research in Innovative Teaching, 7*(1), 130-142. https://digitalcommons.odu.edu/chs_pubs/37/

Popham, W. J. (2006). Phony formative assessments: buyer beware! *Educational Leadership*, *64*(3), 86-87.

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing, 21*(3), 249-258. https://doi.org/10.1191/0265532204lt283ed

Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: disentangling the construct of formative assessment. *Language Testing, 17*(2), 215-243. https://doi.org/10.1177/026553220001700206

Reimann, P., Halb, W., Bull, S., & Johnson, M. (2011). *Design of a computer-assisted assessment system for classroom formative assessment* [Conference paper]. *2011 14th International Conference on Interactive Collaborative Learning* (pp. 465-472). IEEE. https://doi.org/10.1109/ICL.2011.6059627

Renard, L. (2017). The differences between formative and summative assessment – BookWidgets interactive learning. https://www.bookwidgets.com/blog/2017/04/the-differences-between-formative-and-summative-assessment-infographic

Rosalia, C., & Llosa, L. (2009). Assessing the quality of online peer feedback in L2 writing. In R. de Cássia Veiga Marriott & P. Lupion Torres (Eds), *Handbook of research on e-learning methodologies for language acquisition* (pp. 322-338). IGI Global. https://doi.org/10.4018/978-1-59904-994-6.ch020

Saint Mary's University. (2019). *Annotated bibliographies*. https://www.smu.ca/academics/writing-an-annotated-bibliography.html

Saito, H., & Inoi, S. I. (2017). Junior and senior high school EFL teachers' use of formative assessment: a mixed-methods study. *Language Assessment Quarterly*, *14*(3), 213-233. https://doi.org/10.1080/15434303.2017.1351975

Scarino, A. (2013). Language assessment literacy as self-awareness: understanding the role of interpretation in assessment and in teacher learning. *Language Testing, 30*(3), 309-327. https://doi.org/10.1177/0265532213480128

Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagné & M. Scriven (Eds), *Perspectives of curriculum evaluation (AERA Monograph Series on Curriculum Evaluation* no. 1 (pp. 39-83). Rand McNally.

Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 623-646). Praeger.

Spolsky, B. (2008). Language assessment in historical and future perspective. In E. Shohamy & N. H. Hornberger (Eds), *Encyclopedia of language and education* (pp. 445-454). Springer. https://doi.org/10.1007/978-0-387-30424-3_192

Spolsky, B. (2016). History of language testing. In E. Shohamy, I. Or & S. May (Eds), *Language testing and assessment. Encyclopedia of language and education* (3rd ed.). Springer. https://doi.org/10.1007/978-3-319-02326-7_32-1

Stiggins, R. (2005). From formative assessment to assessment for learning: a path to success in standards-based schools. *Phi Delta Kappan*, *87*(4), 324-328. https://doi.org/10.1177/003172170508700414

Štrukelj, E. (2018, January 3). Writing a systematic literature review. *JEPS Bulleting Blog*. https://blog.efpsa.org/2018/01/03/writing-a-systematic-literature-review/

Tan, K. (2013). A framework for assessment for learning: implications for feedback practices within and beyond the gap. *International Scholarly Research Notices*, *2013, 1-6*. https://doi.org/10.1155/2013/640609

Vassiliou, S., & Papadima-Sophocleous, S. (2019). A systematic review and annotated bibliography of second language learning formative assessment: an overview. *Conference Proceedings*, *12th International Conference Innovation in Language Learning* (pp. 352-362). https://conference.pixel-online.net/ICT4LL/files/ict4ll/ed0012/Conference%20Proceedings.pdf

Wang, T.-H. (2008). Web-based quiz-game-like formative assessment: development and evaluation. *Computers & Education, 51*(3), 1247-1263. https://doi.org/10.1016/j.compedu.2007.11.011

Wiliam, D. (2011). What is assessment for learning? *Studies In Educational Evaluation, 37*(1), 3-14. https://doi.org/10.1016/j.stueduc.2011.03.001

Formative assessment in second language learning: a systematic review and an annotated bibliography
Written by Skevi Vassiliou, Salomi Papadima-Sophocleous, and Christina Nicole Giannikas