

A systematic review of formative assessment in language learning

2.1. Introduction

A Systematic Review (SR) is "the art and science of identifying, selecting, and [synthesising] primary research studies to provide a comprehensive and trustworthy picture of the topic being studied" (Oakley, 2012, cited in Crompton, Burke, & Gregory, 2017, p. 5). According to Crompton, Burke, and Lin (2019), an SR "can uncover new trends and additional findings" (p. 5). In the case of Formative Assessment (FA) research, scholars aggregate findings to gain a better understanding of how FA is supporting Language Learning (LL) (Bachelor & Bachelor, 2016; Graham, Harris, & Hebert, 2011; Jian & Luo, 2014; Saoud, 2016; Widiastuti & Saukah, 2017; Wolf, Shore, & Blood, 2014).

With more and more frequent use of FA in LL, and as stated by Crompton et al. (2017), "it is critical to maintain an updated synthesised collection of research so that the scholarly community can remain current in their understanding of [FA] and its impact on student learning. Furthermore, it is of [great importance] that researchers continue to add to this growing base of scholarly knowledge by investigating unexamined or under-examined questions surrounding [FA in LL]" (p. 1), such as the type of publications in the researched period of time, the research purposes and types, the research designs, methods, and data collection tools and expected outcomes of these studies; the languages, the types of participants, and the education levels (primary, secondary, tertiary) in the publications studied; the types of FA applications used in LL in the research studied; the focus of the LL; the learning theories and teaching methods supporting the FA LL researched; and the major geographic distribution in the studies of FA in LL.

Based on this rationale, and on the fact that there is a gap in the area of SR in FA in LL since the beginning of FA in LL, the purpose of this SR is to deliver the scholarly community with a current sum of evidence and a synthesis of FA in LL research conducted in the period of 20 years (2000-2020). It aims to offer an analysis of the specific LL FA practices by answering specific research questions and suggesting further research in the area. In order to have an overview of the main characteristics of the studies involving FA in LL from 2000 to 2020, this SR was guided by the below tentative research questions.

- What were the major publication types, research purposes, research design types, methods, data collection tools, and outcomes?
- What were the languages, participant types, and educational levels involved in the studies?
- Which types of FA applications were used in LL research studies?
- Which was the LL focus formatively assessed?
- What learning theories and language teaching methods were used to support FA in LL?
- What was the geographic distribution of LL FA studies?

These tentative research questions framed eight features that were identified for analysis as it was conducted in other SRs and used as predetermined codes for the qualitative analysis (Crompton et al., 2019; Spolaôr & Benitti, 2017). These main categories/codes were the following.

Publication type: the types of publications were classified as scientific
peer-reviewed: journal articles, conference proceedings papers, short
papers, book chapters, books, handbooks, doctoral or master theses,
or reports.

- Research type: the research type aimed to identify the main research designs, methodologies (e.g. qualitative, quantitative, mixed; case study, action research, etc.), and data collection tools that were applied in FA in LL studies
- Research purpose and outcomes: this category aimed to identify the main purpose and outcomes of the publications reviewed.
- The language(s) studied, participant types and educational level: another category aimed to identify, establish, and record (1) the languages explored in research in FA in LL; (2) the types of participants (students, teachers, or both) in the research under examination; and (3) the level of educational studies (primary, secondary, and tertiary) FA in LL research was carried out for, and the potential research gaps in specific educational levels.
- FA types: this category aimed to investigate and record the most commonly used types of FA applied in LL.
- Language focus: this classification aimed to bring light into the LL focus of the FA in LL research carried out.
- Learning theories and language teaching methods: this category aimed to record the learning theories supporting the use of FA in LL.
- Geographic distribution of the use of FA in LL studies: this categorisation
 aimed to provide a chart of the countries where the studies took place
 and demonstrate the activity on FA in LL at country level worldwide.

2.2. Methodology

This book focused on the years 2000 (Rea-Dickins & Gardner, 2000: first

publication referring to FA in LL) to 2020. Searches included scientific reviewed publications written in English and published between 2000 and 2020.

2.2.1. How the literature search was conducted

This SR is informed by methodology procedures followed in other SR (e.g. Crompton et al., 2019; Poole & Clarke-Midura, 2020; Spolaôr & Benitti, 2017). The SR follows a qualitative research method in order to summarise the studies conducted on the subject. During the search, particular sections of the publications were looked at in order to inform the selection process. These sections were looked at in the following order: titles, abstract, table of contents, whole text. If the title was relevant to the topic, the abstract was looked at, and then the rest of the aspects. If one of the aspects was irrelevant, the publication was not considered for this review.

An initial search was conducted in electronic databases. The following table presents the electronic databases and electronic resources that were used for the purpose of this research.

Table 2.1. Databases used in SRs

Bibliographic databases	Database URL
EBSCOhost	http://web.a.ebscohost.com
ERIC	https://eric.ed.gov/
ResearchGate	https://www.researchgate.net/
Google Scholar	https://scholar.google.com/

The study selection consisted of applying search strings from the above bibliographical databases described (Table 2.1). The search included strings such as below

 (Formative Assessment) AND (Second Language Teaching) OR (Second Language Learning)

- (Classroom Assessment) AND (Second Language Teaching) OR (Second Language Learning)
- (Alternative Assessment) AND (Second Language Teaching) OR (Second Language Learning)
- (Portfolio Assessment) AND (Second Language Teaching) OR (Second Language Learning)

2.2.2. Inclusion/exclusion criteria

After a close study of SR inclusion/exclusion criteria (Meline, 2006; Piper, 2013; Saldaña, 2015; The University of Melbourne, 2021), a number of criteria were specified to select studies for inclusion in the review. The following 12 criteria were used to determine which types of publications to include in the review:

- the publication was published between 2000-2020;
- the publication was peer-reviewed (journal articles, conference proceedings papers, short papers, book chapters, books, handbooks, Doctoral or Master theses, research reports);
- the publisher or type of publication (journal, book, etc.) where the
 research was found is distinguished and reviewed by professionals in
 the field;
- the publication included at least two of the search terms;
- the publication matched the predetermined characteristics used in other SR and served the purposes of this SR;
- author(s) had credibility with institutional affiliation, educational background, past writing experience as it relates to research, or text

written in the author's area of expertise, and the author has been cited in other sources or bibliographies;

- the publication reported the application of FA in LL;
- the information is verified in another source or the author gives evidence to their findings;
- the publication presented FA in LL in a primary, secondary, or tertiary context;
- the publication presented a quantitative or a qualitative or a mixed research approach to researching FA in LL;
- the publication was well written; and
- the publication contributed to the aim of this book, which is to give an overview of the activity in L2 FA in the last 20 years.

Seven criteria for exclusion of articles were also identified:

- the publication was not published between 2000-2020;
- the publication did not match the predetermined characteristics, and emerging code used in other SRs, and did not serve the purposes of this SR;
- the publication was hosted in web pages that are not freely accessed and only abstracts were accessed;
- the publication is composed of only one page (abstract papers), posters, scientific events programmes, and tutorial slides;

- the publication duplicated other publications by the same author (similar title, abstract, results, or text). In such a case, only one was included in this review;
- the publication's writing style did not meet academic expectations; and
- the publication was written in a language other than English.

As mentioned earlier, studies considered for inclusion within the SR were first identified from titles and abstracts generated from one of two sources: electronic databases, and other electronic resources. The study search was applied for work published between 2000-2020 using the electronic databases presented in Table 2.1. As a result, a total number of 16,475 research publications were identified. The number was considerably decreased to 104 after applying the inclusion/exclusion criteria. Table 2.2 presents the number of publications that were found in each electronic database before and after applying the inclusion/exclusion criteria.

Table 2.2. Number of papers before and after the inclusion/exclusion criteria

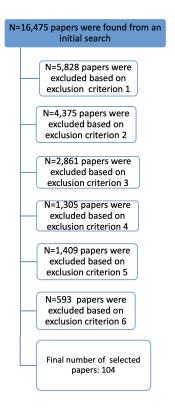
Database	Frequency	Frequency	
	Initial Search	After the inclusion/exclusion criteria	
EBSCOhost	151	11	
ERIC	351	19	
Google Scholar	15,721	57	
Research Gate	252	16	

Most of the publications were found via Google Scholar (57.44%) and ERIC (16.49%) databases. The exclusion criteria significantly decreased the number of documents that were related to the use of FA in LL.

An initial search in databases resulted in 16,475 papers. The first exclusion criterion eliminated 5,828 publications, as they were not published in the chronological period 2000-2020; 10,647 publications remained. With the second

criterion, 4,375 were removed since they dealt with FA generically and were not related to the use of FA in LL. The third exclusion criterion eliminated 2,861 publications since they were not freely accessed and only abstracts could be accessed by the researcher. The remaining 3,590 studies were further evaluated to ensure that they fulfilled the other inclusion/exclusion criteria. Another 1,305 studies were also removed according to the fourth exclusion criterion, since they were composed only of one page. The fifth exclusion criterion eliminated another 1,409 publications since they were duplicated. Moreover, according to the sixth criterion, 593 publications were removed since they were written in languages other than English. The search process is presented in Figure 2.1.

Figure 2.1. The search strategy and review process



The 104 papers that were included in the SR and AB met all the inclusion criteria.

2.2.3. Coding

Once the selection of the publications was completed with the use of the set of inclusion/exclusion criteria, and a specific number of publications was determined, methodology procedures used in SRs were explored in order to analyse the data. As SRs mainly follow a qualitative research method, a study of coding procedures was carried out to decide on the ones that would best suit this SR. According to Creswell (2009, p. 187), there are three approaches to coding:

- developing codes only on the basis of the emerging information collected from participants;
- using predetermined codes and then fitting the data to them; or
- using some combination of predetermined and emerging code.

The combination of the predetermined and emerging code approach was followed to collect and analyse the data. A deductive coding or a concept-driven approach with code themes was first applied. These code themes were predefined by the researchers based on the inclusion and exclusion criteria (Medelyan, 2020). During the coding procedure, however, other interesting themes emerged. These were also recorded and analysed.

Data were analysed by following a coding process, which is a procedure of analysing qualitative inquiries; it is a way of mapping or tagging data that are related to a particular research question (Elliott, 2018; Saldaña, 2015). Qualitative data were then graphed in various ways, e.g. using figures (Qualitative Variables, 2009). Such figures are used in order to identify percentages of the themes identified.

2.2.4. Analysis framework

The coding enabled the researchers to obtain concrete data, in order to better synthesise and interpret the results. The data examination and analysis also helped in refining and improving the original research questions. Moreover, it enabled the researchers to evaluate and compare the data and come to some new suggestions and recommendations for future research (Saldaña, 2015).

2.3. Results

One hundred and four (104) publications were screened, assessed for eligibility by the three SR authors, and included in this review. This process was carried out from 2019 to February 2022. The analysis of the data helped finalise the formation of the research questions. Some substantial considerations and conclusions about the implementation of FA in LL were noted, based on the research questions and the data analysis. The following is a description of the results.

2.3.1. Major types of publications, research purposes, research types, methods, data collection tools, and outcomes identified in the studies involving FA in LL

2.3.1.1. Type of publications

The following figure presents the types of publications for FA in LL from 2000 to 2020 included in the SR (Figure 2.2).

The majority of publications were research articles with a percentage of 80%, followed by book chapters with 3%, Master dissertations with 3%, conference papers with 3%, reports with 3%, short papers with 3%, books with 2%, Doctoral dissertations with 2%, and symposium papers with 1%.

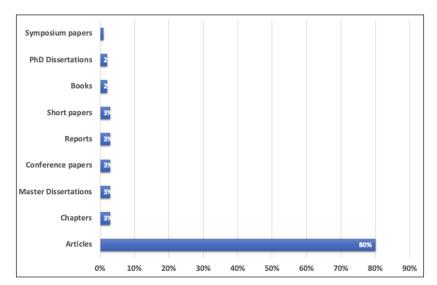


Figure 2.2. Types of publications in FA in LL practices

2.3.1.2. Research purposes

A large number of research studies up to 97% aimed to investigate or collect data about the use of FA in learning and teaching a L2. The rest included literature reviews or other publications. More specifically, 35% of them aimed to display the important role of providing feedback to students which is the core characteristic of FA (Bachelor & Bachelor, 2016; Bahati, Tedre, Fors, & Evode, 2016; Burner, 2016; Chen, May, Klenowski, & Kettle, 2014; Fakeye, 2016; Levy & Gertler, 2015; Titova, 2015; Tsagari & Michaeloudes, 2013). Also, 24% of the studies sought to identify if PA for FA purposes is considered effective for L2 learning and teaching (Burner, 2014, 2016; Cummins & Davesne, 2009; Little, 2002; Papadima-Sophocleous, 2017; Phung, 2016; Rezaee, Alavi, & Shabani, 2013). Furthermore, 14% of studies investigated the different types of assessment practices of L2 instructors in different countries (Leung & Rea-Dickins, 2007); 10% of the publications discussed that, although there is interest globally in FA in LL and is applied in many cases,

there was still a dominance of test-driven formal assessment environments in many countries and that these caused stress and anxiety to students (Leung & Rea-Dickins, 2007; Rea-Dickins, 2004; Tsagari, 2004). Additionally, a significant number of 7% of publications aimed to identify if students were trained in FA practices. Furthermore, 10% revealed that researchers, on the one hand, seek to discover if the integration of digital tools enables teachers to introduce better FA practices, and on the other hand, if students accept and better understand the importance of FA through digital applications.

2.3.1.3. Research types

Research types were identified by taking into consideration the research designs reported in the sources examined. They have been recorded and coded as mentioned in the publications. The major research type, as indicated from the data analysis, is 'study' with a percentage of 42%, either mentioned as a study or comparative study, case study, or pilot study. For that reason, the authors have investigated the research methodologies and research tools that were mentioned in the publications in-depth so as to collect more information about the type of study. However, the only information for the research type or design that was mentioned was the word 'study'. The percentage of research papers mentioned only as a study is about 22%, followed by a case study with 8.25%, a comparative study with a percentage of 6.25%, and a pilot study with a percentage of 5.5%. This also happened with the use of the word 'research'. Twenty per cent (20%) were described as only 'research' with no further indication as to what type of research it was in each case, 'survey' was represented with a percentage of 12.4%, 'action research' with a percentage of 8.8%, 'literature review' with a percentage of 7%, 'exploratory analysis' with a percentage of 2.7%, 'experiment project' with 2%, 'review' with 3.2%, and 'systematic review' with 1.9%. The main characteristic of the types of sources investigated that used study research design (either as a case study or comparative study or pilot study) were interviews and observations to identify the impact of language FA integrations into their classes. Overall, eight different research types were reported. Table 2.3 presents the major research types that were applied in FA in LL publications, as reported in the publications.

Table 2.3. Major research designs that were used for FA in LL practices

Study	42%
Research	20%
Survey	12.4%
Action Research	8.8%
Literature Review	7%
Review	3.2%
Exploratory Analysis	2.7%
Experiment Project	2%
Systematic Review	1.9%

This data analysis revealed that the majority of research publications did not clearly state their type of research: 42% of the publications did not clearly indicate what type of study/research the publication was. Twenty-two per cent of publications were mentioned as just study and 20% as just research. The rest of them clearly stated what type of research they were ('survey' 12.4%, 'action research' 8.8%, 'literature review' 7%, 'review' 3.2%, 'exploratory analysis' 2.7%, 'experiment project' 2%, and 'systematic review' 1.9%).

2.3.1.4. Research methods and data collection tools

The dominant research method that was applied by researchers FA in LL, as extracted from the data, is the qualitative method with a percentage of 39.4%, followed by the quantitative method with a percentage of 32.4%, and the mixed method with 28.2%. Moreover, it seems that the most common research tool that was applied by researchers was interviews with a percentage of 29%, followed by questionnaires with 24%, observations with 20%, prepost tests with 17%, video-audio recordings with 6%, transcripts of students' work with 4%.

2.3.1.5. Research outcomes

The outcomes of the research publications examined included the impact of FA application to both teachers and students.

A total of 95% publications reported a positive impact of FA in LL practices on students' motivation and progress. A low 3.7% did not mention any impact of FA in LL practices on students or teachers. Moreover, it has been stated that the positive impact of FA on students may increase cooperative learning by integrating peer assessment (Hansson, 2015).

Another theme that emerged from the qualitative data analysis in some research papers was that of reduction of students' anxiety (Bayat, Jamshidipour, & Hashemi, 2017; Ketabi & Ketabi, 2014; Tang, 2016). It has been noted that FA practices and multimedia tools can reduce students' anxiety and improve oral performance (Tang, 2016). Also, in Bayat et al. (2017) it was concluded that formative quizzes can also reduce students' anxiety and improve their listening skills.

In 19% of the publications, it was stated that teachers had a positive attitude towards using FA in their practice. Some studies focused on specific FA aspects which had a negative effect, for example, in some cases it has been argued that portfolios used for FA purposes are time-consuming for students (Guadu & Boersma, 2018). They have been considered time-consuming as well for teachers to provide feedback, especially in classes with a big number of students (Tsagari, 2004).

Furthermore, a number of studies have highlighted the fact that the educational system was still largely based on formal traditional testing assessment and there were fewer cases where alternative types of assessment were implemented (Rezaee et al., 2013). Additionally, in other studies, it has been argued that language teachers still preferred using summative based testing assessment tasks due to the belief that SA provides them with a clearer picture of their students' performances (Tsagari, 2016).

In addition, some other publications have indicated that there are some significant considerations regarding the lack of teachers' knowledge and training in LAL (Crusan et al., 2016; Lam, 2015; Tsagari & Michaeloudes, 2013; Tsagari et al., 2018). Numerous studies have concluded that pre-service and in-service L2

teachers need specific training in LAL in implementing FA practices and their level in LAL was low (Crusan et al., 2016; Lam, 2015; Tsagari & Michaeloudes, 2013).

2.3.2. Languages, types of participants, and educational levels in the publications studied, involving FA in LL

2.3.2.1. Languages

Figure 2.3 indicates that the dominant language in LL FA research, which was conducted in English, was for EFL/ESL. This was reflected in 84% of the research publications. Then, it was followed by research for EAL with 6%, for Spanish with 3%, French with 3%, Italian with 3%, and Sign Language as an L2 with 1%.

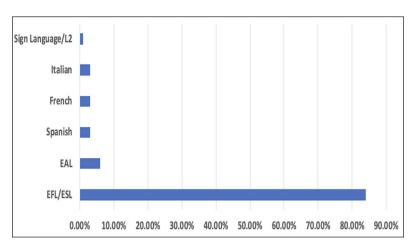


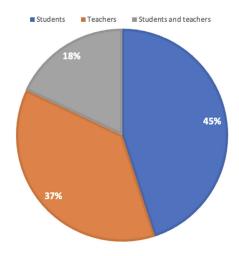
Figure 2.3. Languages in FA in LL applications

2.3.2.2. Types of participants

Another theme that emerged from the data was the type and number of participants who took part in the various research projects. It is clear from the

information presented in Figure 2.4 that the majority of the studies were carried out with participants as students, with a percentage of 45%; then as teachers, with a percentage of 37%, followed by both teachers and students with a percentage of 18%.

Figure 2.4. Participants in LL FA research implementations



2.3.2.3. Types of educational levels

Figure 2.5 clearly indicates that the majority of research for FA practices has been carried out at the tertiary level with 62%, at the secondary level with 18%, and at primary level with 20%.

Figure 2.5. Language FA practices at different educational levels



The data reveal that tertiary settings were the most frequent. Therefore, it is evident that college and university students and teachers as participants experienced the most language FA practices compared to primary and secondary education students and teachers.

2.3.2.4. Types of FA applications in LL studies

The information in Figure 2.6 displays the types of activities used as FAs in LL that were reported in the investigated publications.

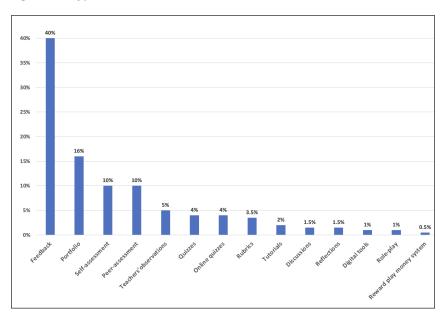


Figure 2.6. Types of FAs used in LL

The most frequently reported was the provision of feedback with 40%. Feedback was reported in papers as 'corrective feedback', 'online feedback', 'peer feedback', 'diagnostic feedback', 'audio feedback' and 'criterion feedback'. The portfolio was next in frequency with 16%, followed by self-assessment with 10%, then peer assessment with 10%, reflections with 1.5%, quizzes with 3%, online

quizzes (Socrative, Kahoot, Edmodo, Padlet, Storify, Quizlet) with 3%, rubrics with 3.5%, teachers' observations with 5%, tutorials with 2%, discussions with 1.5%, digital tools (iPods and iPads) with 1%, role-play with 1%, and reward play money systems with 1%. It is worth mentioning that technology-enhanced activities are gaining ground in LL FA (Cummins & Davesne, 2009; Levy & Gertler, 2015; Pinto-Llorente, Sánchez-Gómez, García-Peñalvo, & Martín, 2016; Saglam, 2018).

The information collected from the qualitative analysis of the data about FA types in LL also indicates the significant role of digital tools in LL FA practices; 25.65% of the FA types involved digital tools or applications that were used by teachers. These included online quizzes (Socrative, Kahoot, Edmodo, Padlet, Storify, Quizlet) with 5.5%, online portfolios with 7%, and digital tools (iPods and iPads) with 0.9%. Moreover, online feedback was used by either the teacher or by a peer (12.25%). Also, it is stated that learners benefit from receiving computer formative feedback and their speaking and writing skills improve (Levy & Gertler, 2015). Learners are benefitted from the instant feedback and the positive environment that digital tools offer. The game-based characteristic of these digital tools can enhance their language skills according to the findings from the publications examined (Cummins & Davesne, 2009; Levy & Gertler, 2015; Pinto-Llorente et al., 2016; Saglam, 2018).

2.3.2.5. FA LL focus

Language teaching, learning, and assessment have a long history since ancient times. The focus depended on the reasons people learned a language. According to Howatt and Smith's (2014) synoptic overview, modern language teaching in Europe included the following periods: the Classical Period (1750-1880), which focused on emulating the teaching of classical languages (based on the model of teaching Latin and Greek), and was associated with grammar-translation and classical methods of teaching; the Reform Period (1880-1920), which focused on teaching the spoken language and was associated with various reform methods such as the natural method, the berlitz methods and the direct method; the Scientific Period (1920-1970), which focused on a

scientific basis for teaching and was associated with teaching methods such as the oral method, the multiple line of approach, the situational approach, the oral approach and the audiolingual method; the Communicative Period (1970-2000+), which aimed to teach for 'real-life' communication and was associated with communicative language teaching and task-based language teaching methods. One can notice through the study of these periods, that the focus changed direction from the traditional language teaching coverage of specific language skills such as reading, speaking, listening, and writing and language aspects such as grammar and vocabulary to a more integrative way of dealing with all these for 'real life' communication. One of the aims of this review was to identify the focus of the research conducted in the area of FA in LL in the period 2000-2020, in other words, to find out whether it followed this trend. Figure 2.7 displays the percentages of the focus of these publications.

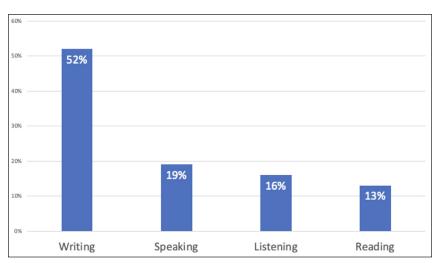


Figure 2.7. Language focus of FA in LL

According to the data shown in Figure 2.7, most FA in LL research publications were carried out for writing skills with 52%. Oral communication skills come second with 19%, then listening skills with 16%, and reading with 13%.

According to the publications examined, most of the research on FA for writing skills was implemented in an academic writing environment (65%) and 35% at the secondary educational level. The most common tool that was used for writing skills for FA purposes was the portfolio as it was considered as alternative evidence of students' writing performance (Burner, 2014; Cummins & Davesne, 2009; Lam, 2015; Little, 2002; Papadima-Sophocleous, 2017; Phung, 2016; Rezaee et al., 2013). Other tools that were used to assess writing skills formatively were blogs, peer- and self-feedback, AWE tools, and journals.

For speaking communication skills, 60% of the research has been carried out at the tertiary level; 25% in primary educational settings, and 15% in secondary educational settings.

For listening skills, 55% of the studies have been carried out in tertiary settings, 25% in secondary settings, and 20% in primary settings. The most frequent tools used to formatively assess listening were iPods, iPads, online quizzes, and portfolios.

According to the publications examined, most of the research on FA for reading skills was carried out at the tertiary level as well (68%) followed by 32% at the secondary educational level. Some of the tools used to formatively assess reading skills were ELFA and portfolios.

2.3.2.6. Language aspects

Most publications on FA in LL treated either language skills or other aspects (72%). Vocabulary was mentioned in 8% of the studies and grammar in 4%. Many publications mentioned the role of gamified quizzes for the acquisition of L2 vocabulary and grammar and the increase of students' performance with a percentage of 15.6% (Pinto-Llorente et al., 2016). The significant role of feedback in the acquisition of vocabulary was also highlighted (Titova, 2015). It is worth mentioning that in some educational systems, more emphasis is put on assessing grammar and vocabulary than on language skills (Tsagari, 2016). Some particular software like Grammarly and Turnitin Quickmark were recorded as

tools that provide opportunities for electronic feedback in grammar, vocabulary, and pronunciation (Williamson & Sadera, 2016).

2.3.2.7. Other aspects: gender, student anxiety reduction

Another aspect that has been noted during the data analysis was that participants in almost all of the studies presented so far were both male and female, with the exception of only one case, where participants were only females (Chen et al., 2013).

Another theme that emerged from the qualitative data analysis in some research papers was that of reduction of students' anxiety (Bayat et al., 2017; Ketabi & Ketabi, 2014; Tang, 2016). Researchers noted that FA practices and multimedia tools used reduce students' anxiety and improve oral performance (Tang, 2016). Also, Chen et al. (2013) have argued that participants from an urban and regional university in China shared the same understanding of FA, in addition to differences which are related to their sociocultural conditions, beliefs on teachers' and students' roles, and expectations in English.

2.3.3. Learning theories and teaching methods used to support FA in LL

LL, teaching, and assessment are based on some teaching approach or method, which is based on some learning theory. Very often, in our teacher training programmes, we noticed that language practitioners tend to base their teaching more on practice and are often not clear of the learning theory and teaching methods their teaching is based on. For this reason, the publications were also reviewed to explore this aspect in the research carried out on FA in LL. The data analysis revealed that only a small number of studies (4.8%) mentioned the learning theories, learning approaches and learning methods that support the concept of the use of FA in teaching and learning (Chen & Zhang, 2017; Davison, 2019; Kuo, 2015; Little, 2002; Poehner & Lantolf, 2005). The dominant learning theory, as found in 2% of the publications, that supports FA practices in L2 environments is Vygotsky's sociocultural theory of learning. According to

Vygotsky, learning occurs as a result of interaction with others where alternative types of assessment are suggested (Poehner & Lantolf, 2005). Moreover, another important language theory mentioned is the constructivism theory of learning, where students have an active involvement in their learning and in their own assessment with a percentage of 1.5% (Buyukkarci, 2010). Additionally, a process-oriented approach was mentioned in one paper. According to this approach, the learner practices writing skills for authentic purposes and the whole procedure includes pre-writing, writing, revising, editing and sharing (Kuo, 2015).

2.3.4. Major geographic distributions in studies involving FA in LL

This SR also investigated the geographic distribution of research in the use of FA in LL practices. The aim of this was to establish the current research practices in FA in LL and identify possible needs and themes that may need to be further researched in the future. Location is related to where research was conducted. If the publication was not related to a research paper, the location of the researcher's affiliation was recorded as 'location'.

The SR reveals that studies took place in all continents except Antarctica. It was found that the continent with the highest percentage was Asia with a percentage of 40%, followed by Europe with 25%, America (South and North) with 22%, Oceania with 10%, and Africa with 3%. The data showed that Asia is the most active continent, with China and Iran presenting 19% of the total amount of publications (Figure 2.8).

Within these continents, a total number of 39 countries were represented. From Figure 2.9, it is evident that the US has the highest percentage (18%) of studies in the use of FA in LL, followed by China with 10%, Iran with 9%, Turkey with 8%, Australia with 8%, Norway with 5%, the UK with 5%, Canada with 4%, Taiwan with 3.5%, Cyprus with 3%, Malaysia with 3%, Japan, Canada, New Zealand with 2% each, and South Korea, Spain, Greece, Indonesia, Thailand, Colombia, Ethiopia, Chile, Israel, Italy, Nigeria, Netherlands, and Russia with

Oceania

1% each. The significance of these findings is elaborated further on in the discussion section.

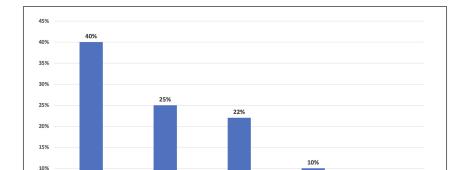
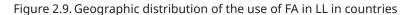


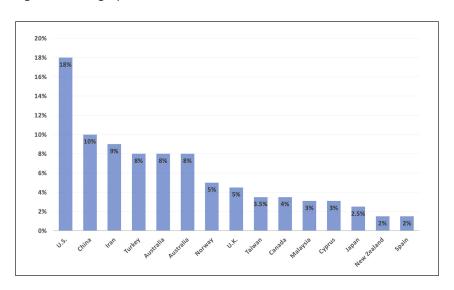
Figure 2.8. Geographic distribution of FA in LL in continents



America (North and South)

Europe

Asia



3%

Africa

The findings of this SR reveal a global interest in the use of FA in LL, and the distribution of this interest (Figure 2.8 and Figure 2.9).

2.4. Discussion

The combination of predetermined and emerging code approaches used for the collection of the data (Creswell, 2009) was also used for the qualitative analysis and discussion of the results. The analysis and discussion evolved around the research questions and the eight features that were identified (Crompton et al., 2019; Spolaôr & Benitti, 2017), as well as the emerged themes.

The SR of these 104 publications and the method used brought to light the following insights related to research in FA in LL during the specific period under review.

2.4.1. The increasing interest in the use of FA in LL in recent years

The data analysis revealed that FA is gaining ground in LL research in recent years. Figure 2.10 illustrates the number of publications per year.

According to the literature, during the first attempts of FA application in LL, between 2000 to 2011, there was somewhat low activity in the area. This is not surprising as stakeholders need some time to get used to new concepts such as FA, and engage in its application and research. During the next half of the period examined, the data reveal a growing interest, with the years 2016 and 2017 recording contributing the largest number of publications to the literature. This can be attributed to the increase in the holders' understanding of current learning theories, language teaching methods, and deriving assessment development and applications.

The availability of digital tools in LL may have also offered additional opportunities, for example, digital tools can contribute to giving instant feedback,

and online FA tools can improve students' language skills (Pinto-Llorente et al., 2016; Williamson & Sadera, 2016). The implementation of technology enables teachers to collect data during lessons from students' performances and helps them to assess their progress on a continuous basis (Joyce, 2018; Levy & Gertler, 2015; Pinto-Llorente et al., 2016; Williamson & Sadera, 2016).

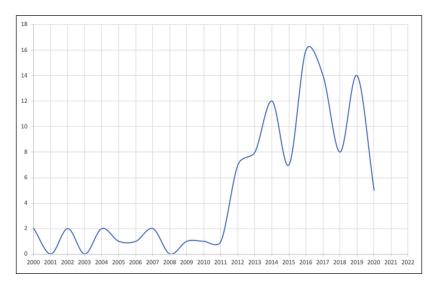


Figure 2.10. Number of publications in each year the use of FA practices in LL

2.4.2. Types of publications

The findings revealed that from 2000 to 2020, the types of publications that were published were mostly articles (80%). This reveals the interest in research in this area and the urge of researchers to inform and share such research with the FA in LL research community and other stakeholders in recognised scientific publications. The rest (books, book chapters, reviewed conference papers, doctoral and Master dissertations to short papers and reports) were less than 5%. The number of publications in the form of scientific and reviewed books and book chapters indicate that researchers may need to be encouraged to engage more in such activity in this form. More research by postgraduate students at

MA and PhD level is also recorded to be needed (Buyukkarci, 2010; Kuo, 2015; Meissner, 2018; Radford, 2014; Saoud, 2016; Vågen, 2017). This book may prove to be useful to those interested in the different types of publications, their value and characteristics, and work as a guide as to what form of research has been conducted and what needs to be further researched.

2.4.3. Research purposes

The examination of the data regarding the purposes of the research publications revealed that these evolved around the following topics: (1) the importance of the role of providing feedback to students (Bachelor & Bachelor, 2016; Bahati et al., 2016; Burner, 2016; Chen et al., 2014; Fakeye, 2016; Levy & Gertler, 2015; Titova, 2015; Tsagari & Michaeloudes, 2013), (2) the identification of the effectiveness of PA for L2 learning and teaching (Burner, 2014, 2016; Crusan et al., 2016; Cummins & Davesne, 2009; Little, 2002; Papadima-Sophocleous, 2017; Phung, 2016; Radford, 2014; Rezaee et al., 2013), (3) the identification of the different types of assessment practices in different countries (Leung & Rea-Dickins, 2007), (4) the fact that 10% of the publications discussed that there was still a dominance of test-driven formal assessment environments, (5) the extent of test-driven formal assessment environments in many countries and their cause of stress and anxiety to students even though FA applications are being used (Crusan et al., 2016; Leung & Rea-Dickins, 2007; Phung, 2016; Radford, 2014; Rea-Dickins, 2004; Tsagari, 2004), (6) the identification of student training in FA in LL practices, (7) the identification of the role of technologies in the introduction of better FA practices by the teachers, and (8) the acceptance and better understanding from students of the importance of FA through digital applications. Although the data analysis made a number of FA in LL research purposes known, it also revealed that there are a lot of FA in LL research purposes that need further examination or have not vet been examined, areas such as a clearer definition of formative assessment, a clearer understanding of formative and summative assessments and the way some assessments are used differently for formative or summative assessment purposes, the distance between FA in LL policies, and their actual application during the learning process, etc.

2.4.4. Research types

A close look at the types of research identified in the 104 publications reviewed confirms that the majority tends not to clearly state the types of study/research the publications are based on. For example, where in some cases it is clearly stated that it is a case study or a pilot study, in many others it is referred to only with the word 'study'. This may suggest that the authors may not think it is important to mention what type their research is, or they may not be clear on their research type. The results can also suggest that this phenomenon has not been much researched. The lack of such reference in earlier research is indicative. This could be a focus of future studies.

2.4.5. Research methods and data collection tools

The qualitative method was most practised with a percentage of 39.4%, (Bahati et al., 2016; Guadu & Boersma, 2018; Haines, Meima, & Faber, 2013; Lam, 2015; Rezaee et al., 2013; Saliu Abdulahi, 2017; Vågen, 2017; Widiastuti, Mukminatien, Prayogo, & Irawati, 2020), followed by the quantitative method with a percentage of 32.4% (Caruso, Gadd Colombi, & Tebbit, 2017; Pinto-Llorente et al., 2016; Seyyedrezaie, Ghansoli, Shahriari, & Fatemi, 2016), and the mixed method with a percentage of 28.2% (Chen & Zhang, 2017; Cotter & Hinkelman, 2019; Guadu & Boersma, 2018; Naghdipour, 2017; Tang, 2016; Tsagari & Michaeloudes, 2013; Yarahmadzehi & Goodarzi, 2020). Some common research outcomes emerged from the qualitative data analysis of the types of sources examined. These findings established that all three methods were nearly equally used. Furthermore, the types of research methods used in these publications reflect a balanced application of all three. It seems that collecting data through qualitative research tools was preferable to many researchers. However, the great number of papers with a mixed method approach, where researchers combine qualitative and quantitative data cannot be ignored, as their outcomes are considered more valid and offer a deeper and clearer understanding of the findings (McKim, 2017).

The examination of data collection tools also recorded the use of a variety ranging from interviews, questionnaires, observations, pre- and post-tests, video-

audio recordings, and transcripts of students' work. The findings also revealed a widespread use of these tools by many researchers. Future research may focus on the quality of the use of these research methods and tools.

2.4.6. Research outcomes

The outcomes of the research publications under examination affirmed mostly a positive impact of FA in LL (Bayat et al., 2017; Caruso et al., 2017; Chen & Zhang, 2017; Fakeye, 2016; Naghdipour, 2017). This is a good finding as it may encourage teachers reluctant to integrate FA in their teaching to do so. The majority of the studies also suggested that the integration of FA for learning an L2/FL may increase students' motivation, and as a result improve of their LL (Alzaid & Alkarzae, 2019; Ammar, 2020; Vassiliou & Papadima-Sophocleous, 2019). It may also enable teachers to collect more data and information about their students' progress, and use it to improve the learning process (Huang, 2016; Little, 2002; Smith & Davis, 2014).

In the publications examined, the use of assessment activities for FA purposes such as peer assessment creates a more comforting feeling to students towards LL. It makes them feel more comfortable through cooperative learning, and it reduces their anxiety (Bayat et al., 2017; Buyukkarci, 2010; Tang, 2016; Zhao, 2014). Some publications also mentioned the improvement of their listening skills (Bayat et al., 2017; Caruso et al., 2017; Cummins & Davesne, 2009; Pinto-Llorente et al., 2016).

The use of technologies was mentioned as a factor supporting all of the above (Alam, 2019; Buyukkarci, 2010; Caruso et al., 2017; Karagianni, 2012; Phung, 2016; Tsagari, 2004). This reinforces the student-centred focus of current learning theories and LL approaches which aim to involve students in their language acquisition (Chen & Zhang, 2017; Davison, 2019; Papadima-Sophocleous, 2017). Moreover, two other common findings were the effective provision of feedback and the increase of students' motivation and performances in the target language (Burner, 2016; Chen et al., 2014). It has been highlighted by researchers that FA environments can reduce students' anxiety and make them

feel more comfortable using the target language. According to the publications, the focus of the teachers has been on various aspects such as specific assessments and tools used for FA purposes, time considerations in giving feedback, the tendency in some countries for SA practices/preferences, and the lack and need of teacher training in the use of FA in LL. These findings support earlier findings by researchers (Chen et al., 2014; Crusan, Plakans, & Gebril, 2016; Cummins & Davesne, 2009; Kuo, 2015; Vassiliou & Papadima-Sophocleous, 2019; Zhao, 2014). They also indicate the importance of further research in these areas.

The outcomes of the reviewed research also highlighted the lack of teachers' knowledge and training in the application of FA practices, thus reinforcing earlier research on these topics (Crusan et al., 2016; Lam, 2015; Tsagari & Michaeloudes, 2013; Tsagari et al., 2018) and the need for clearer understanding of the relation and or distinction between FA and SA (Leung & Rea-Dickins, 2007).

2.4.7. Languages

Another feature investigated was which languages were represented in research carried out in English in the area of the use of FA in LL (Alam, 2019; Alharbi & Meccawy, 2020; Burner, 2016; Chen & Zhang, 2017; Cotter & Hinkelman, 2019; Joyce, 2018; Levy & Gertler, 2015; Tang, 2016; Vågen, 2017). The results indicated that the dominant language was English as an FL, L2, or EAL, followed by Spanish (Bachelor & Bachelor, 2016; Carreira, 2012; Radford, 2014). One could expect that, in mainly English publications, this is expected. However, in other research written in English and focusing on other aspects related to LL and assessment (testing), languages other than English are more researched. More research is needed in FA in the teaching of other languages beyond English and Spanish. A review of research written in other languages would also enrich our knowledge in the area in further domains as well.

2.4.8. Types of participants

In the 104 publications reviewed in this book, the results indicated a balanced distribution of research in these three different categories of participants: (1)

students (Alam, 2019; Cotter & Hinkelman, 2019; Joyce, 2018; Ranalli, Link, & Chukharev-Hudilainen, 2017), (2) teachers and students (Burner, 2016; Kuo, 2015), and (3) teachers (Guadu & Boersma, 2018; Papadima-Sophocleous, 2017; Wang, 2017). This established that the publications reviewed considered and included all classroom shareholders in their research. As most research recorded was mainly conducted at tertiary education level, future research may bring interesting information to light regarding participants from the other education levels, namely primary and secondary. Future research may also explore other aspects related to FA in LL and participants, such as preferences, similarities, and differences in preferences, etc.

2.4.9. Types of educational levels

It is obvious from the data collected that the dominant educational setting where language FA applications were carried out was that of higher education (Alam, 2019; Guadu & Boersma, 2018; Joyce, 2018; Kızıl & Yumru, 2019; Lam, 2015; Naghdipour, 2017; Papadima-Sophocleous, 2017; Seyyedrezaie et al., 2016; Wang, 2017; Williamson & Sadera, 2016). It is clear that college and university students and teachers had the opportunity to experience more language FA practices than teachers and students in other educational levels, namely primary and secondary. Consequently, it is suggested that more research should be carried out to investigate the reasons for the focus of language FA studies at the tertiary level and to encourage researchers, practitioners, and teachers to include more language FA practices in primary and secondary settings. Findings from tertiary environments could also be made known or shared with primary and secondary educators. Training in FA in LL could also be offered to primary and secondary educators.

2.4.10. Types of FA applications in LL studies

The data revealed that a variety of types of activities and tools such as feedback, portfolio, online quizzes, rubrics, teacher observations, peer-work, tutorials and questioning, discussions, and digital voice recordings were already used in L2 learning for FA purposes (Alam, 2019; Bayat et al., 2017; Carreira, 2012;

Karagianni, 2012; Lam, 2018; Pinto-Llorente et al., 2016). However, since theories of learning, L2 teaching methodologies and technologies constantly develop, more and continuous research is required in the future in underresearched areas and in areas where no research has been carried out so far. A deeper discussion is also needed for the relation of these with the actual nature of FA, as very often they are automatically assumed as FA.

2.4.11. LL focus formatively assessed

To determine the LL focus of the publications under review when analysing the results, the development of this focus and the current trend were identified. Throughout the history of modern language teaching in Europe, the focus of language teaching, learning and assessment depended on the reasons people learned languages and changed through the years (Howatt & Smith, 2014). The findings revealed that the reviewed publications mostly focused on the FA of languages skills and specific aspects such as grammar and vocabulary, based on the traditional language teaching approaches; it did not follow the change of direction which focuses on learning a language to be able to participate in 'real-life' communication, and as a consequence assess formatively 'real-life' communication activities. It is important to also notice that most of this research focused on writing (Alam, 2019; Kızıl & Yumru, 2019; Tavakoli et al., 2018) followed by much less research on speaking (Black & Jones, 2006; Colby-Kelly & Turner, 2007; Cummins & Davesne, 2009; Tuttle & Tuttle, 2013), listening (Bayat et al., 2017; Caruso et al., 2017), reading (Ponce, Mayer, Figueroa, & López, 2018), and other aspects such as vocabulary (Pinto-Llorente et al., 2016; Ponce et al., 2018; Titova, 2015) and grammar (Gan & Leung, 2020; Karagianni, 2012; Titova, 2015).

It would be useful to conduct further research to establish the reason for this preference to writing skills and the reasons why other language skills like reading, speaking, and listening, and other aspects such as grammar and vocabulary have not been yet explored in a great extent. Research should also be carried in the future by researchers to reveal the potentials of FA tools in other aspects and in a more integrative way, reflecting the 'real-life' communication situations which

constitute the expected outcome of language education in recent years. 'Real-life' communication FA practices have not yet been investigated to a satisfactory extent. Another interesting finding was that technology was discussed in many publications which dealt with the acquisition of L2 vocabulary and grammar. Further research would also shed light on the use of technologies in FA LL in other aspects such as FA applications for real-life communication.

2.4.12. Learning theories, teaching, and learning approaches and methods in LL FA

The results revealed that only a small number of studies (4.8%) mentioned learning theories, LL, and teaching approaches and methods underpinning the language FA practices they investigated (Chen & Zhang, 2017; Davison, 2019; Kuo, 2015; Little, 2002; Poehner & Lantolf, 2005). This may suggest that specific learning theories, LL, and teaching approaches are implied; it may also suggest that they are not clear to the researchers or they may not feel it is necessary to mention them in their research reports. This lack could be because of other reasons that could be of interest to be identified and discussed. It is suggested that these aspects should be further explored.

2.4.13. Geographic distribution of studies

The study of this aspect unveiled that although research in the practice of FA in LL is carried out globally (in all continents except Antarctica, and in 39 countries), this is not sufficient and extensive compared to the research in SA. However, when it comes to countries, the results indicate that much more research is required globally, in the countries where research already occurs but also in other countries around the world, where such research has not started yet, and establish the reasons why this is not occurring in both cases. It may be either because of their educational systems or because of lack of availability of research funding, it could be because of other reasons that need to be identified and discussed. For example, although the US comes first in research in the area, the percentage of research conducted (22%) is still low; therefore, more research is required in language FA in the US, compared to that conducted for SA. The

same goes for all the countries where research in language FA is carried out, as such research is even less than that in the US.

More research should also be carried out to investigate why in countries in Asia, and in specific countries, the rate of interest in language FA practices is higher than in other countries in, for example, Northern America, and Europe. This could be identified by exploring language assessment practices in these countries. According to Chen, Kettle, Klenowski, and May (2013) and Jian and Luo (2014), China's educational system, for example, is based on SA. Added to that, according to Jian and Luo (2014), teachers' level of FA understanding was very low. According to them, this phenomenon could be due to the fact that in China there is low financial support for FA implementations. It could also be because of the relation of the language FA practices to the sociocultural, historical, political, and geographic conditions of China. As stated by Wang (2017), Chinese EFL teachers have difficulty inputting FA theories in action and their assessment practices are dominated by SA with tests and exams. Another explanation may be due to the fact that China belongs to the Confucian Heritage Culture (CHC) with some other Asian countries. In CHC, countries such as China are dominated by examination-oriented environments. According to CHC, a teacher is the authority and plays a key role in the learning process (Chen et al., 2013).

Also, in some studies, it is reported that in traditional educational systems like in China and Japan, interaction is not encouraged and students are sceptical and reluctant to give and accept peer feedback. Silence and listening are more common for them than interacting (Chen et al., 2013). Also, it was found that 90% of Iranian teachers use testing to assess their students as they perceive it as a fairer approach of assessment (Rezaee et al., 2013).

On the other hand, the high number of studies of language FA that took place in China, for example, can explain the quest to change the testing-centred assessment educational system of China. Some research outcomes already support that the inclusion of FA in LL can benefit both students and teachers (Chen et al., 2013; Jian & Luo, 2014; Jiang, 2014; Zhao, 2014). Also, educational systems such

as that of Iran connect the alternative type of assessment with a need for more democratic and ethical environments by giving the chance to LL for real-life contextualised practices (Rezaee et al., 2013).

Further and deeper study of the reasons why these countries conduct research in FA in LL may provide useful information which can be shared.

In general, the interest in language FA appears to be gradually taking place globally with a close proportion approximately in all continents. However, more research would be beneficial. Aspects such as the knowledge and practices of FA by the different stakeholders can be further explored, the pros and cons can be discussed, the further development can be further examined, comparison can be made and useful conclusions can be drawn.

2.4.14. Emerged themes

2.4.14.1.Types of technologies in the use of FA in LL

Although the use of technologies in FA in LL did not constitute part of any research question, this emerged as an added theme during the data analysis. This finding complements earlier findings (Cummins & Davesne, 2009; Levy & Gertler, 2015; Pinto-Llorente et al., 2016; Saglam, 2018), which support that technologies can enhance the practice of FA in LL. The findings of this research (15% of language FA types involved technologies) indicate the need for further practices and research in this area.

2.4.14.2.The necessity for training language teachers in language FA practices

Another emerging code was the necessity of in-service and pre-service training for language teachers in LAL with a percentage of 7.5% according to Vogt and Tsagari (2014), who explored the Literacy Assessment (LA) of L2 teachers in seven European countries with 854 participants, as well as the necessity of training L2 Teachers in the use of FA in LL. They have stressed the need to address the current insufficient training in LAL. This has also been

confirmed by other researchers. Lam (2015) stated that L2 teachers had little or no training in LA practices. According to Lam (2015), there is a lack of classroom-based assessment practices in Hong Kong. Crusan et al. (2016) have indicated that only 26% of the language teachers who participated had nothing or little training on writing assessment practices and most of them did not assimilate the distinction of FA and SA.

It can be concluded that language teachers could benefit more from FA practices in the future in order to know what and how to effectively assess their students' progress.

2.4.14.3.The necessity for training students in language FA practices

Another emerging theme was that of student training. Some studies reviewed in the research, such as those of Zhao (2010), Restrepo and Nelson (2013), Tsagari and Michaeloudes (2013), Chen et al. (2014), Kuo (2015), Lam (2015), and Crusan et al. (2016) also referred to the importance of training students in FA practices. Also, it was stated that many learners were reluctant to write self-reflections, and teachers admitted that they found it hard to convince them (Chen et al., 2013). Other studies referred to the benefits of such training and the importance of students' awareness and understanding of FA and its practices and benefits in LL. Formative types of assessment enable learners to identify their own strengths and weaknesses and to acquire an awareness of their own skills (Restrepo & Nelson, 2013).

Furthermore, in many exam-oriented countries, it is stated that learners do not feel comfortable or confident in providing peer feedback because of their cultural values or cultural background. Learners are more passive in SA environments and do not accept such positive FA implementations which require critical thinking and self-control (Chen et al., 2013; Jian & Luo, 2014; Jiang, 2014; Liu, 2015; Zhao, 2010).

Therefore, the more proficient and trained the learners are, the more positive and willing they will be in participating in language FA activities that are often used

to support FA. According to Herrera and Macías (2015) LAL teacher education should balance both classroom (which also includes formative assessment) and accountability assessments (e.g. large scale standardised tests).

Some suggestions derived from the studies are based on how teachers can support students' training in FA practices. An initial step should start from teachers' clear understanding of FA features (Burner, 2016; Crusan et al., 2016). Then, teachers could guide students on FA practices and let them experience FA through, for example, the use of FA as a means of checklists and self-assessment rubrics, and show them examples of effective feedback before they will apply it in learning. Moreover, follow-up discussions and oral feedback after the provision of feedback with learners will enhance the validity of the provided feedback (Saliu Abdulahi, 2017). All these aspects could be the subject of further research in the use of FA in LL.

2.4.14.4.Other emerged themes

The review of the 104 publications brought to light that beyond the FA of languages skills such as writing, speaking, listening, reading, and aspects such as vocabulary and grammar, research also explored other aspects related to the practice of FA in LL such as the reduction of students' anxiety (Bayat et al., 2017; Ketabi & Ketabi, 2014; Tang, 2016), students' shared understanding of FA, differences related to students' sociocultural conditions, beliefs on teachers' and students' roles, and expectations in the target language (Chen et al., 2013). It would be interesting to see further research conducted in these areas and in other areas still not dealt with in research such as mediation and multilingualism, etc.

2.5. Limitations

A significant limitation of the study was the difficulty in accessing all the papers that were detected. Although a substantial number of papers (16,475) were found, a considerable number (2,816) were not accessible. Their inclusion in this SR would have made a difference in the research findings and given further and

more informed and inclusive future directions to the researchers, practitioners, and language teachers. That being said, it would also be safe to say that not all 2,816 would have probably met the criteria of FA as forming the learning of students during the learning process.

Another limitation of the study was the fact that in this SR, only papers written in English were included. One could argue that there is a substantial number of papers published in other languages that have reached important research outcomes related to the use of FA in LL. The present study was considered as a starting point to a more comprehensive study of this area, and a blueprint in terms of the method used.

2.6. Conclusions

The SR, which constitutes Part 2 of this research, can be considered a snapshot of current studies of language FA practices from 2000 to 2020. It has identified, selected, and synthesised primary research studies and provided a picture of the topic being studied (Oakley, 2012). Furthermore, the SR highlighted the impact of FA application in support of language teaching and learning. More specifically, the SR has identified the main trends and issues that are related to language FA: the raise of interest in language FA in recent years; the main research methodologies and research tools that are preferred by the researchers in the area of language FA; the purpose of language FA in the research conducted from 2000 to 2020; the types of research conducted; the main research methodologies and data collection tools that are preferred; the types of outcomes reported in these publications; the languages explored in language FA research; the types of participants; the preferred education levels where language FA was carried out; the types of FA applications practised in LL; the input and the extent of input of technologies; the language focus examined; the locations globally where language FA research was carried out; and the need for training of both students and language teachers in FA. The present SR provides researchers, practitioners, and other interested stakeholders considerable information and background regarding research in the area between 2000 to 2020; it also provides new directions for further research of these issues and supports the role of FA in LL.

The annotated bibliography that follows complements the snapshot provided by the SR, by giving an evaluative description of the main features of each of the 104 reviewed publications.

References

- Alam, M. (2019). Assessment challenges & impact of formative portfolio assessment (FPA) on EFL learners' writing performance: a case study on the preparatory English language course. *English Language Teaching*, 12(7), 161-172. https://doi.org/10.5539/elt.v12n7p161
- Alharbi, A. S., & Meccawy, Z. (2020). Introducing Socrative as a tool for formative assessment in Saudi EFL classrooms. *Arab World English Journal*, 11(3), 372-384.
- Alzaid, F., & Alkarzae, N. (2019). The effects of paper, web, and gamebased formative assessment on motivation and learning: a literature review. https://files.eric.ed.gov/fulltext/ED594189.pdf
- Ammar, A. (2020). Impact of formative assessment on raising students' motivation: case of third year EFl students at the university of El-Oued. Doctoral dissertation. University of Tlemcen.
- Bachelor, J. W., & Bachelor, R. B. (2016). Classroom currency as a means of formative feedback, reflection, and assessment in the world language classroom. *NECTFL Review*, 78, 31-42. https://files.eric.ed.gov/fulltext/EJ1256488.pdf
- 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
- Bayat, A., Jamshidipour, A., & Hashemi, M. (2017). The beneficial impacts of applying formative assessment on Iranian university students' anxiety reduction and listening efficacy. *International Journal of Languages' Education and Teaching*, 5(2), 1-11.https://doi.org/10.18298/ijlet.1740
- Black, P., & Jones, J. (2006). Formative assessment and the learning and teaching of MFL: sharing the language learning road map with the learners. *Language Learning Journal*, 34(1), 4-9. https://doi.org/10.1080/09571730685200171

- Burner, T. (2014). The potential formative benefits of portfolio assessment in second and foreign language writing contexts: a review of the literature. *Studies in Educational Evaluation*, *43*, 139-149. https://doi.org/10.1016/j.stueduc.2014.03.002
- Burner, T. (2016). Formative assessment of writing in English as a foreign language. Scandinavian Journal of Educational Research, 60(6), 626-648. https://doi.org/10.1080/00313831.2015.1066430
- Buyukkarci, K. (2010). The effect of formative assessment on learners' test anxiety and assessment preferences in EFL context. Unpublished doctoral dissertation. Cukurova University.
- Carreira, M. M. (2012). Formative assessment in HL teaching: purposes, procedures, and practices. *Heritage Language Journal*, 9(1), 100-120. https://doi.org/10.46538/hlj.9.1.6
- Caruso, M., Gadd Colombi, A., & Tebbit, S. (2017). Teaching how to listen. Blended learning for the development and assessment of listening skills in a second language. *Journal of University Teaching & Learning Practice*, 14(1), 14. https://files.eric.ed.gov/fulltext/ EJ1142367.pdf
- Chen, D., & Zhang, L. (2017). Formative assessment of academic English writing for Chinese EFL learners. *TESOL International Journal*, *12*(2), 47-64. https://files.eric.ed.gov/fulltext/EJ1247811.pdf
- Chen, Q., Kettle, M., Klenowski, V., & May, L. (2013). Interpretations of formative assessment in the teaching of English at two Chinese universities: a sociocultural perspective. *Assessment & Evaluation in Higher Education*, 38(7), 831-846. https://doi.org/10.1080/02602938.2012.726963
- Chen, Q., May, L., Klenowski, V., & Kettle, M. (2014). The enactment of formative assessment in English language classrooms in two Chinese universities: teacher and student responses. Assessment in Education: Principles, Policy & Practice, 21(3), 271-285. https://doi.org/10.1080/0969594X.2013.790308
- Colby-Kelly, C., & Turner, C. E. (2007). AFL research in the L2 classroom and evidence of usefulness: taking formative assessment to the next level. *Canadian Modern Language Review*, 64(1), 9-37. https://doi.org/10.3138/cmlr.64.1.009
- Cotter, M., & Hinkelman, D. (2019). Video assessment module: self, peer, and teacher post performance assessment for learning. In F. Meunier, J. Van de Vyver, L. Bradley & S. Thouësny (Eds), *CALL and complexity short papers from EUROCALL 2019* (pp. 94-99). Research-publishing.net. https://doi.org/10.14705/rpnet.2019.38.992

- Creswell, J. W. (2009). Research design: qualitative, quantitative, and mixed methods approaches (3rd ed.). Sage Publications.
- Crompton H., Burke D., & Gregory, K. H. (2017). The use of mobile learning in PK-12education: a systematic review. *Computers & Education (2017)*. https://doi.org/10.1016/j.compedu.2017.03.013
- Crompton, H., Burke, D., & Lin, Y. C. (2019). Mobile learning and student cognition: a systematic review of PK-12 research using Bloom's taxonomy. *British Journal of Educational Technology*, 50(2), 684-701. https://doi.org/10.1111/bjet.12674
- Crusan, D., Plakans, L., & Gebril, A. (2016). Writing assessment literacy: surveying second language teachers' knowledge, beliefs, and practices. *Assessing Writing*, 28, 43-56. https://doi.org/10.1016/j.asw.2016.03.001
- Cummins, P. W., & Davesne, C. (2009). Using electronic portfolios for second language assessment. *The Modern Language Journal*, 93(1), 848-867. https://doi.org/10.1111/j.1540-4781.2009.00977.x
- Davison C. (2019). Using assessment to enhance learning in English language education. In X. Gao (Ed.), Second handbook of English language teaching. Springer International Handbooks of Education. Springer. https://doi.org/10.1007/978-3-030-02899-2 21
- Elliott, V. (2018). Thinking about the coding process in qualitative data analysis. *The Qualitative Report*, 23(11), 2850-2861. https://doi.org/10.46743/2160-3715/2018.3560
- 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
- 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
- Graham, S., Harris, K. R., & Hebert, M. A. (2011). *Informing writing: the benefits of formative assessment. A Carnegie Corporation Time to Act Report.* Alliance for Excellence in Education. https://www.carnegie.org/publications/informing-writing-the-benefits-of-formative-assessment/
- Guadu, Z. B., & Boersma, E. J. (2018). EFL instructors' beliefs and practices of formative assessment in teaching writing. *Journal of Language Teaching and Research*, 9(1), 42-50. https://doi.org/10.17507/jltr.0901.06
- Haines, K., Meima, E., & Faber, M. (2013). Formative assessment and the support of lecturers in the international university. In *International experiences in language testing and* assessment (pp. 177-190). Peter Lang.

- Hansson, S. (2015). Benefits and difficulties in using peer response for writing in the EFL classroom. Göteborgs Universitet. https://gupea.ub.gu.se/bitstream/2077/38436/1/gupea 2077 38436 1.pdf
- Herrera L., & Macías, D. F. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal*, 17(2), 302-312. https://doi.org/10.14483/udistrital.jour.calj.2015.2.a09
- Howatt, A. P. R., & Smith, R. (2014). The history of teaching English as a foreign language, from a British and European perspective. *Language & History*, *57*(1), 75-95. https://doi.org/10.1179/1759753614Z.00000000028
- Huang, S. C. (2016). No longer a teacher monologue involving EFL writing learners in teachers' assessment and feedback processes. *Taiwan Journal of TESOL*, *13*(1), 1-31. http://www.tjtesol.org/attachments/article/402/04 TJTESOL-273.pdf
- Jian, H., & Luo, S. (2014). Formative assessment in L2 classroom in China: the current situation, predicament and future. *Indonesian Journal of Applied Linguistics*, 3(2), 18-34. https://doi.org/10.17509/ijal.v3i2.266
- Jiang, Y. (2014). Exploring teacher questioning as a formative assessment strategy. *RELC Journal*, 45(3), 287-304. https://doi.org/10.1177/0033688214546962
- Joyce, P. (2018). The effectiveness of online and paper-based formative assessment in the Learning of English as a second language. *PASAA*, *5*, 126-146. https://www.culi.chula.ac.th/publicationsonline/files/article/W8Mo0m4nyBMon110741.pdf
- Karagianni, E. (2012). Employing computer assisted assessment (CAA) to facilitate formative assessment in the State Secondary School: a case study. *Research Papers in Language Teaching and Learning*, *3*(1), 252-268.
- 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
- Kızıl, V., & Yumru, H. (2019). The impact of self-assessment: a case study on a tertiary level EFL writing class. *Mevzu Sosyal Bilimler Dergisi*, *1*, 35-54.
- Kuo, C. L. (2015). A quasi-experimental study of formative peer assessment in an EFL writing classroom. Unpublished doctoral dissertation. Newcastle University. http://theses.ncl. ac.uk/jspui/handle/10443/2863
- Lam, R. (2015). Language assessment training in Hong Kong: implications for language assessment literacy. Language Testing, 32(2), 169-197. https://doi. org/10.1177/0265532214554321

- Lam, R. (2018). Understanding assessment as learning in writing classrooms: the case of portfolio assessment. *Iranian Journal of Language Teaching Research*, 6(3), 19-36. https://doi.org/10.30466/ijltr.2018.120599
- Leung, C., & Rea-Dickins, P. (2007). Teacher assessment as policy instrument: contradictions and capacities. *Language Assessment Quarterly*, 4(1), 6-36. https://doi. org/10.1080/15434300701348318
- 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
- Little, D. (2002). The European Language Portfolio: structure, origins, implementation and challenges. *Language Teaching*, 35(3), 182-189. https://doi.org/10.1017/S0261444802001805
- 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
- McKim, C. A. (2017). The value of mixed methods research: a mixed methods study. *Journal of Mixed Methods Research*, 11(2), 202-222. https://doi.org/10.1177/1558689815607096
- Medelyan, A. (2020). *Coding qualitative data: how to code qualitative research*. InSights. https://getthematic.com/insights/coding-qualitative-data/
- Meissner, M. C. (2018). Formative assessment at the intersection of principles, practice and perceptions. Master's thesis. Faculty of Education and Natural Sciences, Inland Norway University of Applied Sciences.
- Meline, T. (2006). Selecting studies for systematic review: inclusion and exclusion criteria. *Contemporary issues in Communication Science and Disorders*, *33*, 21-27. https://doi.org/10.1044/cicsd_33_S_21
- Memorial University Libraries. (n.d.). *How to evaluate information resources*. https://www.library.mun.ca/researchtools/guides/doingresearch/evaluateall/
- Naghdipour, B. (2017). Incorporating formative assessment in Iranian EFL writing: a case study. *The Curriculum Journal*, 28(2), 283-299. https://doi.org/10.1080/09585176.2016 .1206479
- Oakley, A. (2012). Foreword. In D. Gough, S. Oliver & J. Thomas (Eds), *An introduction to systematic reviews* (pp. vii-x). Sage Publications.

- 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
- Phung, H. V. (2016). Portfolio assessment in second/foreign language pedagogy. *Hawaii Pacific University TESOL Working Paper Series 14*, 90-107. https://www.hpu.edu/research-publications/tesol-working-papers/2016/07HuyPhung.pdf
- Pinto-Llorente, A. M., Sánchez-Gómez, M. C., García-Peñalvo, F. J., & Martín, S. C. (2016, November). The use of online quizzes for continuous assessment and self-assessment of second-language learners. In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 819-824). https://doi.org/10.1145/3012430.3012612
- Piper, R. J. (2013). How to write a systematic literature review: a guide for medical students. NSAMR. University of Edinburg. https://sites.cardiff.ac.uk/curesmed/files/2014/10/NSAMR-Systematic-Review.pdf
- Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, 9(3), 233-265. https://doi.org/10.1191/1362168805lr1660a
- Ponce, H. R., Mayer, R. E., Figueroa, V. A., & López, M. J. (2018). Interactive highlighting for just-in-time formative assessment during whole-class instruction: effects on vocabulary learning and reading comprehension. *Interactive Learning Environments*, 26(1), 42-60. https://doi.org/10.1080/10494820.2017.1282878
- Poole, F. J., & Clarke-Midura, J. (2020). A systematic review of digital games in second language learning studies. *International Journal of Game-Based Learning (IJGBL)*, 10(3), 1-15. https://doi.org/10.4018/IJGBL.2020070101
- Qualitative Variables. (2009). In *Nelson Mathematics Secondary Year Two, Cycle One* (pp.1-4). Nelson Education. http://www.nelson.com/school/elementary/mathK8/quebec/0176237879/documents/NM8SB_3A.pdf?fbclid=IwAR3DNvF45J8yvfoyvyHPNSd-Eg3M1s1aPAlLu7LvE77tL91hySintkarto
- Radford, B. W. (2014). The effect of formative assessments on languageperformance. Unpublished doctoral dissertation. Brigham Young, Provo. http://scholarsarchive.byu.edu/etd/3978

- Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, 37(1), 8-25. https://doi.org/10.1080/01443410.2015.1136407
- 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
- Restrepo, A., & Nelson, H. (2013). Role of systematic formative assessment on students' views of their learning. *Profile Issues in Teachers Professional Development*, *15*(2), 165-183. http://www.scielo.org.co/scielo.php?pid=S1657-07902013000200011&script=sci_arttext&tlng=pt
- Rezaee, A. A., Alavi, S. M., & Shabani, E. A. (2013). Alternative assessment or traditional testing: how do Iranian EFL teachers respond? *Teaching English Language*, 7(2), 151-190.
- Saglam, A. L. G. (2018). The integration of educational technology for classroom-based formative assessment to empower teaching and learning. In A. Khan & S. Umair (Eds), *Handbook of research on mobile devices and smart gadgets in K-12 rducation* (pp. 321-341). IGI Global. https://doi.org/10.4018/978-1-5225-2706-0.ch020
- Saldaña, J. (2015). The coding manual for qualitative researchers. Sage.
- Saliu Abdulahi, D. (2017). Scaffolding writing development: how formative is the feedback? *Moderna språk, 111*(1), 127-155. https://www.duo.uio.no/handle/10852/59613
- Saoud, A. (2016). Impact of formative assessment on raising students' motivation: case of third year EFL students at the University of El-Oued. Unpublished doctoral dissertation. University of Tlemcen.
- Seyyedrezaie, Z. S., Ghansoli, B., Shahriari, H., & Fatemi, A. H. (2016). Examining the effects of Google docs-based instruction and peer feedback types (implicit vs. explicit) on EFL learners' writing performance. *CALL-EJ*, *17*(1), 35-51. http://callej.org/journal/17-1/Seyyedrezaie_Ghonsooly_Shahriari_Fatemi2016.pdf
- Smith, D. H., & Davis, J. E. (2014). Formative assessment for student progress and programme improvement in sign language as L2 programmes. In D. McKee, R. S. Rosen & R. McKee (Eds), *Teaching and learning signed languages* (pp. 253-280). Palgrave Macmillan. https://doi.org/10.1057/9781137312495 12

- Spolaôr, N., & Benitti, F. B. V. (2017). Robotics applications grounded in learning theories on tertiary education: a systematic review. *Computers & Education*, 112, 97-107. https://doi. org/10.1016/j.compedu.2017.05.001
- Tang, L. (2016). Formative assessment in oral English classroom and alleviation of speaking apprehension. *Theory and Practice in Language Studies*, 6(4), 751-756. https://doi. org/10.17507/tpls.0604.12
- Tavakoli, E., Amirian, S. M. R., Burner, T., Davoudi, M., & Ghaniabadi, S. (2018). Operationalization of formative assessment in writing: an intuitive approach to the development of an instrument. *Applied Research on English Language*, 7(3), 319-344. https://doi.org/10.22108/ARE.2018.112373.1340
- The University of Melbourne. (2021). *Common inclusion/exclusion criteria*. https://unimelb.libguides.com/c.php?g=492361&p=3368110
- Titova, S. (2015). Use of mobile testing system PeLe for developing language skills. In F. Helm, L. Bradley, M. Guarda & S. Thouësny (Eds), *Critical CALL proceedings of the 2015 EUROCALL Conference* (pp. 523-528). Research-publishing.net. https://doi.org/10.14705/rpnet.2015.000387
- Tsagari, D. (2004). Is there life beyond language testing? An introduction to alternative language assessment. *CRILE Working Papers*, *58*, 1-23. https://pdfs.semanticscholar.org/19ad/ddb4879992814f8ebbc323a8d6f2dd491a4f.pdf
- Tsagari, D. (2016). Assessment orientations of state primary EFL teachers in two Mediterranean countries. *Center for Educational Policy Studies Journal*, *6*(1), 9-30. https://doi.org/10.26529/cepsj.102
- Tsagari, D., & Michaeloudes, G. (2013). Formative assessment patterns in CLIL primary schools in Cyprus. In S. Ioannou-Georgiou, S. Papadima-Sophocleous & D. Tsagari (Eds), *International experiences in language testing and assessment* (pp. 75-93). Peter Lang Edition.
- Tsagari, D., Vogt, K., Froelich, V., Csépes, I., Fekete, A., Green A., Hamp-Lyons, L., Sifakis, N., & Kordia, S. (2018). *Handbook of assessment for language teachers*. https://taleproject.eu/pluginfile.php/2129/mod_page/content/12/TALE%20Handbook%20-%20colour.pdf
- Tuttle, H. G., & Tuttle, A. (2013). *Improving foreign language speaking through formative assessment*. Routledge. https://doi.org/10.4324/9781315854854
- Vågen, M. T. (2017). Formative assessment in EFL writing: a case study of pupils' perceptions of their feedback practice and attitudes to receiving and using feedback. Master's thesis. University of Bergen.

- 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%20 Proceedings.pdf
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: findings of a European study. *Language Assessment Quarterly*, 11(4), 374-402. https://doi.org/10.10 80/15434303.2014.960046
- Wang, X. (2017). A Chinese EFL teacher's classroom assessment practices. *Language Assessment Quarterly*, 14(4), 312-327. https://doi.org/10.1080/15434303.2017.1393819
- Widiastuti, I. A. M. S., Mukminatien, N., Prayogo, J. A., & Irawati, E. (2020). Dissonances between teachers' beliefs and practices of formative assessment in EFL classes. *International Journal of Instruction*, *13*(1), 71-84. https://doi.org/10.29333/iji.2020.1315a
- Widiastuti, I. A. M. S., & Saukah, A. (2017). Formative assessment in EFL classroom practices. *Bahasa dan Seni: Jurnal Bahasa, Sastra, Seni, dan Pengajarannya*, 45(1), 50-63. https://doi.org/10.17977/um015v45i12017p050
- Williamson, K., & Sadera, E. (2016). Electronic formative feedback and its effect on the writing skills of Asian L2 postgraduate students. *DEANZ2016: Conference Proceedings* (pp. 208-210). https://kiwibelma.files.wordpress.com/2016/05/deanz16-conferenceproceedings11-april.pdf
- Wolf, M. K., Shore, J. R., & Blood, I. (2014). *English learner formative assessment (ELFA):* a design framework. ETS.
- Yarahmadzehi, N., & Goodarzi, M. (2020). Investigating the role of formative mobile based assessment in vocabulary learning of pre-intermediate EFL learners in comparison with paper based assessment. *Turkish Online Journal of Distance Education*, 21(1), 181-196.
- Zhao, H. (2010). Investigating learners' use and understanding of peer and teacher feedback on writing: a comparative study in a Chinese English writing classroom. *Assessing writing*, 15(1), 3-17. https://doi.org/10.1016/j.asw.2010.01.002
- Zhao, H. (2014). Investigating teacher-supported peer assessment for EFL writing. *ELT Journal*, 68(2), 155-168. https://doi.org/10.1093/elt/cct068



Published by Research-publishing.net, a not-for-profit association Voillans, France, info@research-publishing.net

© 2022 by Authors

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

Publication date: 2022/11/21

Rights: this book is published under the Attribution-NonCommercial-NoDerivatives International (CC BY-NC-ND) licence. Under this licence, the book is freely available online (https://doi.org/10.14705/rpnet.2022.60.9782383720133) for anybody to read, download, copy, and redistribute provided that the authors and publisher are properly cited. Commercial use and derivative works are, however, not permitted.

Disclaimer: Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it was not under consideration for publication elsewhere. While the information in this book is believed to be true and accurate on the date of its going to press, neither the editorial team nor the publisher can accept any legal responsibility for any errors or omissions. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

Trademark notice: product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Copyrighted material: every effort has been made to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net Cover design and photo by © Raphaël Savina (raphael@savina.net)

ISBN13: 978-2-38372-013-3 (Ebook, PDF, colour)

ISBN13: 978-2-38372-014-0 (Ebook, EPUB, colour)

ISBN13: 978-2-38372-012-6 (Paperback - Print on demand, black and white)

Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

British Library Cataloguing-in-Publication Data. A cataloguing record for this book is available from the British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: 2022 novembre.