



Ca' Foscari
University
of Venice

Master's Degree
in Language Sciences

Final Thesis

A Reflective Approach to Learning- oriented Assessment of Young English Language Learners

A Diary Study in an Italian Primary English Classroom

Supervisor

Ch. Prof. Carmel Mary Coonan

Assistant supervisor

Ch. Prof. Marcella Menegale

Graduand

Elena Simioni

Matriculation Number 852331

Academic Year

2020 / 2021

TABLE OF CONTENTS

ABSTRACT	4
INTRODUCTION	5
1. LITERATURE REVIEW	8
1.1 Introduction	8
1.2 Assessment	9
1.2.1 Theories of Educational Assessment	9
1.2.2 Assessment in Second and Foreign Language	16
1.2.3 A Multidimensional Approach to Classroom-based Assessment	20
1.2.4 Diagnostic Competence and Teachers as Assessment Agents	25
1.2.5 Assessing Young Language Learners	30
1.2.6 Assessment in Italian Primary Schools: a new legislation	34
1.3 Reflective Teaching	39
1.3.1 Reflective practice	39
1.3.2 Reflective Teacher Diary	49
1.4 Conclusion	56
2. RESEARCH PROJECT	57
2.1 Introduction	57
2.2 Context and Research Questions	57
2.3 Method	59
2.3.1 Research methodology	59
2.3.2 Participants and Language Teaching History.....	61
2.3.3 Instruments and Procedure	61
2.3.4 Method of Data Analysis	65

3. ANALYSIS OF DATA	69
3.1 Deductive Category Application	71
3.2 Inductive Category Development	84
3.2.1 Emotional Challenges of Assessment	85
3.2.2 Reflective cycle	88
3.3 Quantification	90
4. PRESENTATION OF RESULTS	94
4.1 Discussion	94
4.2 Limitations and Further research	99
CONCLUSION	101
REFERENCES	102

ABSTRACT

The present study aims at discussing the use of the teacher diary as a tool for reflective teaching and for the formative assessment of young pupils of English as a Foreign Language (EFL) in an Italian primary school, in compliance with the national reform of the evaluation system at the primary level, published in December 2020. A search of the literature revealed few studies that adopt a teacher diary as a tool to investigate the practice of a foreign language teacher, but no studies have been found in the context of Italian primary education.

The diary study was conducted over a period of four months, and introspective data were collected through a diary kept by the teacher-researcher. The entries were carefully examined and analysed both during the study, in order to identify and correct any issues pertaining to assessment, and at the end of the study, through a process of qualitative content analysis (QCA). The findings confirm established research suggesting that young language learners have special needs when it comes to assessing their proficiency in a foreign language, and that teachers require specific diagnostic competences in order to meet their assessment needs. Reflective practice seems to be a viable way to support teachers in adapting and implementing learning-oriented assessment methods.

Despite the emergence of some issues related to keeping a teacher diary, overall the action-research proved to be successful, in that the teacher diary served as a means of initiating the teacher-researcher's self-reflection about her professional role.

INTRODUCTION

In today's Europe, there is a pressing need to assess and certify the language proficiency and the linguistic abilities of all citizens from an increasingly younger age, a phenomenon that is also due to the influence of the Common European Framework of Reference (CEFR) (Council of Europe, 2001) on National educational institutions' curricula. Considering the major implications for the learner's professional and personal future prospects, focussing on the special assessment needs of young language learners is of paramount importance (McKay, 2006).

Identifying the best way to assess and evaluate progress and performance of learners of English as a Foreign Language (EFL) is one of the most complex and challenging aspects in the field of teaching English as a foreign language (Loh & Liew, 2016). Even more so if the target group is formed by Young Language Learners (YLL) enrolled in the primary (elementary) level of education, as they require a different set of teaching skills and methodologies in order to best tackle their specific learning needs (Pinter, 2011).

In the Italian context, these influences have recently led to the publication, in December 2020, of a new Ministerial Order¹ which established a new system for the evaluation and the assessment of primary school pupils. The guidelines² provided by the Ministry of Education direct the teachers towards a change of perspective, from assessment *of* learning to assessment *for* learning. This fresh perspective has a strong formative stance, as the pupils' learning needs and learning styles are now the prime concern of every educator. The Ministry recognizes the complexity of the learning process and invites the teachers to adopt instruments suited to collecting the necessary elements to determine the level of acquisition of the learning objectives. This represents a major challenge for primary school teachers,

¹ M.O. n. 172, 4th December 2020. Available at: https://www.istruzione.it/valutazione-scuola-primaria/allegati/ordinanza-172_4-12-2020.pdf (accessed 31/08/2021).

² Ministero dell'Istruzione, dell'Università e della Ricerca (2020). Linee Guida. La formulazione dei giudizi descrittivi nella valutazione periodica e finale della scuola primaria. Retrieved April 27, 2021, from: https://www.istruzione.it/valutazione-scuola-primaria/allegati/Linee_Guida.pdf (accessed 31/08/2021).

who need to radically change their assessment approach in order to comply with the current legislation, which is also the main reason for conducting this research. The idea for this research project, therefore, stems directly from a real need of the teacher-researcher to adapt her classroom practice in order to implement the legislative provisions issued by the Ministry.

The objective of the research is that of exploring a potential solution that might help other EFL teachers to provide high-quality formative assessment for their young language learners. Although some studies adopt the teacher diary to investigate different aspects of teaching, the context of teaching English as a foreign language at the primary school level remains largely unexplored. In addition, very little research in the field of learning-oriented assessment and reflective teaching has been conducted in Italy, as the majority of studies has been conducted abroad.

In order to carry out this research project, the selected approach is that of reflective teaching, an 'action-oriented' (Kemmis, 1986 quoted in Bailey et al., 2001: 37) process of introspection carried out by both experienced and novice teachers who wish to focus on their teaching practices by collecting and analysing their thoughts in a critical manner (Bailey et al., 2001). More specifically, the study is centred on the adoption of a reflective teacher diary as the main tool to collect data and implement reflective practice-oriented assessment methodologies within the research context, namely two EFL classrooms of a primary school located in the metropolitan city of Venice.

Building on the fundamental idea that the language classroom is the 'crucible' of language learning (Allwright and Bailey 1991: 18, citing Gaies, 1980), the chosen type of research is that of an action-research-like spiral process of action, observation and critical reflection. The data is then further analysed by performing an interpretive analysis. The chosen approach is that of qualitative content analysis (QCA), developed by Mayring (2000; 2015). As categories, it was chosen to adopt the different levels of diagnostic competence, as theorized by Edelenbos & Kubanek-German (2004) in addition to some of the interrelated dimensions of learning-oriented assessment (LOA), as described in the working framework developed by Turner and Purpura (2015).

Ultimately, this study shows that journaling can help foreign language teachers to cope with many different issues related to assessing pupils that may be encountered within the classroom context. Furthermore, the teacher diary also proves to be a valid tool to comply with the new regulations in the area of assessment and evaluation, imposed by the Italian Ministry of Education since the 2020-2021 school year.

1. LITERATURE REVIEW

1.1 Introduction

This first chapter will provide an overview of the practice of reflective teaching applied to the context of foreign language teacher-based assessment of young learners. The aim is that of building a robust theoretical framework that will support the selection and design of the instrument (teacher diary) and the subsequent analysis, discussion and exploration of the information emerging from the data.

Paragraph 1.2.1 will provide an overview of assessment, with particular attention to the debate involving summative and formative assessment. Relevant constructs will be examined, such as the dynamics at play within Assessment for Learning (AfL), Assessment for Learning (AfL) and Assessment as Learning (AaL) as discussed by Black and William (1998), Stiggins (2002), Earl (2003), Taras (2005) and Newton (2007).

In paragraph 1.2.2, a pertinent body of research on second and foreign language assessment will be utilised to describe the main issues concerning features and purposes of language assessment and the role of the Common European Framework of Reference (CEFR) in the evaluation process. In paragraph 1.2.3, the focus will be on the classroom as the 'crucible' of learning, therefore various studies on classroom-based assessment will be examined. Given its multidimensional nature, Turner and Purpura's (2015) model will be used to describe the complex and dynamic aspects of learning and assessing a foreign language in a classroom context. Paragraph 1.2.4 is centred on the teacher as agent of assessment (Rea-Dickins, 2004). In exploring the teacher's role as assessment practitioner, particular attention will be given to the emotional challenges faced by teachers-assessors (Brown et al., 2018). More specifically, the constructs of diagnostic competence (Edelenbos & Kubanek-German, 2004) and assessment literacy will be looked at in detail. In paragraph 1.2.5, the factors involved in assessing young language learners (YLLs) will be explored.

Paragraph 1.2.6 will provide a historical outline of assessment in the Italian educational system, by referring to a body of research on assessment mainly focusing on the evolution of assessment and evaluation practices and procedures which are most common in the primary (elementary school) level of instruction.

In the second part of the literature review, more specifically in paragraph 1.3.1, a definition of reflective teaching will be provided, presenting the purposes and benefits of adopting this approach to classroom-based assessment (Pollard, 2002; Farrell, 2012), and attention will be devoted to related constructs of reflection-on, in- and before-action (Wilson, 2008). Wallace (1991)'s reflective practice model will be used as a framework to describe the reflective cycle at the core of this approach. Finally, in paragraph 1.3.3, the teacher diary as a reflective practice tool will be defined and analysed, and issues about its implementation in a research design will be discussed.

1.2 Assessment

1.2.1 Theories of Educational Assessment

Assessment is a 'natural activity', a social mechanism that happens spontaneously in our everyday life, whenever we judge or reflect upon our work, and therefore it is widely considered as a crucial component of any educational system and 'an intrinsic part of teaching and learning' (Rea-Dickins & Germaine, 1992: 3).

Brown (2018: 1) defines 'educational assessment' as

the set of methods and processes by which evidence about student learning is designed, collected, scored, analysed, and interpreted. These processes are meant to support decisions about teaching (e.g., What material needs to be taught again or differently?), learning (e.g., What material does a student need to revise?), administration (e.g., What students are ready for promotion to the next grade?), policymaking (e.g., What areas of the curriculum need teacher development?) and accountability (e.g., Which schools add more value? or, Which student should be given the supreme scholarship prize?). The processes depend on expert judgement and statistical analysis of the quality of the assessment methods, their relationship to intended objectives or outcomes, and the validity of consequences.

According to Brown (2018), factors such teaching and learning processes, and socio-cultural norms have a strong influence on educational stakeholders' beliefs, assumptions and opinions about assessment. In his view, for teachers, adopting a professional stance on assessment means acknowledging and being open to identify the fallacy of their assessment practices or the inefficacy of one's teaching, instead of blaming students or other entities for unsatisfactory learning outcomes (Brown, 2018).

A common distinction emerging from the literature on assessment is that between summative and formative assessment. Scriven (1967, as cited in Scriven,1991) is widely cited as being the first to draw the divide, referring to the procedure of evaluating educational programmes. As summarized by Newton (2007), other scholars like Sadler (1989), Black (1998) and Harlen and James (1997) contributed to this debate, but Bloom, Hastings and Madaus (1971, quoted in Taras, 2009) were the first to transfer the notion of 'formative' from evaluation to assessment (Dunn and Mulvenon, 2009; Taras, 2009). More specifically, in their *Handbook on Formative and Summative Evaluation of Student Learning*, they adopt new reference criteria (purpose, time, generalization level) – based on function rather than on process – to differentiate between formative and summative assessment (Taras, 2009).

Summative Assessment (SA) and Formative Assessment (FA) have often been equated to the notions of Assessment of Learning (AoL) and Assessment for Learning (AfL), respectively (Earl, 2003; Black and William, 1998; Taras, 2005).

A definition of AoL has been provided by Earl, who wrote that

its purpose is summative, intended to certify learning and report to parents and students about students' progress in school, usually by signalling students' relative position compared to other students. [...] in classrooms is typically done at the end of something [...] and takes the form of tests or exams [...](2003: 27).

According to its use, summative assessment can be divided into internal or external with reference to a particular institution, where internal uses include grading or reporting to parents and students, whilst external uses consist of certification or accreditation processes or selection for educational or professional purposes (Harlen, 2005). The high-stakes use of

tests for examination is particularly widespread in higher education contexts, where they have an impact on student learning and motivation, but issues about reliability, validity and utility of 'summative, feedout' high stakes practices are widely discussed in the literature (Knight, 2002).

As for the contrasting concept, that of Assessment for Learning, the term appears in a pamphlet published by the Assessment Reform Group (1999), and in Black and William (1998)'s seminal paper on formative assessment. In a later work on formative assessment, Black et al. (2004: 10), provide a comprehensive definition of AfL:

Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting students' learning. [...] An assessment activity can help learning if it provides information that teachers and their students can use as feedback in assessing themselves and one another and in modifying the teaching and learning activities in which they are engaged. Such assessment becomes "formative assessment" when the evidence is actually used to adapt the teaching work to meet learning needs.

In a later publication by the ARG entitled *Assessment for Learning: 10 Principles* (2002), AfL is defined 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'. Furthermore, they identify ten guiding principles of AfL:

Assessment for learning

- is part of effective planning
- focuses on how students learn
- is central to classroom practice
- is a key professional skill
- is sensitive and constructive
- fosters motivation
- promotes understanding of goals and criteria
- help learners know how to improve

- develops the capacity for self-assessment
- recognizes all educational achievement (ARG, 2002, quoted in Gardner, 2012).

Gardner (2012) comments on this definition and writes that ‘unpacking this deceptively simple definition, in terms of classroom practice, reveals a complex weave of activities involving pedagogic style, student-teacher interaction, self-reflection (teacher and student), motivation and a variety of assessment processes’ (2012: 3).

The definition provided by the research group brings to mind the re-conceptualization of Ramaprasad (1983)’s definition of feedback provided by William and Thompson, who subdivide the formative assessment practice into three main instructional steps to be undertaken by instructors/assessors:

- Establishing where the learners are in their learning
- Establishing where they are going
- Establishing what needs to be done to get them there (2007: 14).

Stiggins embraces the teachers’ perspective and claims that ‘when they assess for learning, teachers use the classroom assessment process and the continuous flow of information about student achievement that it provides in order to advance, not merely check on, student learning’ (2002: 761). While denouncing the lack of AfL practices in US schools, he admits that in order to provide a far-reaching kind of assessment aimed at maximizing student achievement, both AoL and AfL are useful and needed (ibid, 2002).

Of particular interest for the current research project is the fact that within the field of AfL, close attention is paid to the role of teachers. Teachers’ beliefs and attitudes toward this practice in relation to professional development have been investigated by Marshall and Drummond (2006), who elicited teachers’ reflections and found that in some instances their beliefs impede an effective application of AfL practices. Other teachers involved in the study, however, appeared to be more in line with ‘the spirit of AfL’, because in their reflections emerged a flexible attitude towards the lesson and a sense of agency (ibid, 2006: 147).

Gardner (2006, quoted in Lee et al. 2019) claims that one key difference between AoL and AfL is that in the latter, the teacher's role is decidedly downsized, as the students are encouraged to shoulder part of the responsibilities about the outcomes of the learning process.

In recent years, the field of formative assessment has seen the rise of a third notion, that of Assessment as Learning (AaL), an assessment practice that is considered as a subcategory of formative assessment (Earl, 2013) that 'highlights the student-centred dimension of AfL' (Lee et al. 2019: 73). By implementing AaL, teachers let students activate their critical thinking skills in order to understand and connect new information to prior knowledge thus fuelling the 'regulatory process in meta-cognition', a process where 'students personally monitor what they are learning and use the feedback from this monitoring to make adjustments, adaptation, and even major changes in what they understand' (Earl 2003, as cited in Earl, 2013: 3). A defining feature of AaL is the stress that is put on developing students' active involvement in assessment as part of their learning process, with particular emphasis on feedback, self-regulation, self-efficacy and metacognition (Dann, 2014).

The clear-cut distinction between summative and formative assessment, however, has been disputed by several sources, who point out the straightforward idea that summative assessment can be used for formative purposes (Black et al., 2004; Taras, 2001). In this regard, Wininger (2005) introduced the notion of formative summative assessment (FSA), an assessment method aimed at providing students with both qualitative and quantitative feedback on their exams in the context of higher education. Similarly, Dunn and Mulvernon (2009)'s critical review of research methods on formative assessment focuses on the terminological misreading linked to assessment and evaluation, and maintain that the notion of formative or summative evaluation needs to be clearly distinguished from that of assessment. In outlining their Practical Model of Assessment and Evaluation System (2009), the authors claim that

Assessments are instruments for collecting information [...]. Evaluation is a separate, but related issue that has to do with the use of assessment-based data. Although an assessment may be designed to be formative or

summative, the data acquired by the administration of either type of assessment may be used for formative or summative purposes. In other words, formative evaluation or summative evaluation may be applied to either formative or summative assessment data (2009: 3-4).

Taras (2005) taps into this discussion by posing the concept of feedback at the centre of the dispute. In her opinion, as 'most SA for formal assessment purposes requires feedback, therefore the only real requirement in order to integrate FA into practice is to engage learners with using this feedback for learning in future work' (2005: 475). In other words, she sees summative assessment as a central and neutral construct in the field of educational assessment. Feedback plays a fundamental role in formative assessment theory and its evolution in this context traces back to the work of Ramaprasad (1983), who provided the well-known definition of what has been labelled *formative feedback*: 'feedback is information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way' (1983: 4).

As appears from the above, when discussing the different types of assessment, it is fundamental to consider their purpose. Newton (2007: 150) provides an in-depth analysis of three different interpretations of the term 'assessment purpose', which correspond to three levels of interpretation:

- the *judgement level*, when the aim is to obtain a 'standards-referenced judgement'
- the *decision level*, when the aim is to use the assessment in order to make decisions
- the *impact level*, when the aim is to influence something or to obtain a pre-determined effect

Newton (2007) stresses how important it is to discriminate these levels. More specifically, he refers to the common distinction between formative and summative assessment as a misconception, because, from his perspective, these two types of assessment belong to different categories:

the term 'summative' can only meaningfully characterize a type of assessment judgement (i.e. it operates at the judgement level of discourse), while the term 'formative' can only meaningfully characterize a type of use

Newton (2007)'s view that, if seen through this lens, summative judgements can be used for formative purposes is supported by Taras (2002), who postulates that summative assessment tools such as marks and grades can be used formatively, but only if they are preceded by feedback.

1.2.2 Assessment in Second and Foreign Language

At a global level, demands in terms of second or foreign language (L2) proficiency have increased, and in order to succeed in their personal and professional life, L2 users have to develop their linguistic mastery in relation to new knowledge, skills and abilities (KSAs) and this shift has had a huge impact on many aspects of language education, including assessment (Purpura, 2016). Multiple stakeholders, from public institutions to individuals, increasingly rely on assessment practices for many aspects of language education, ranging from diagnosing learning needs to certify the achievement of standards (Purpura, 2016).

In its broader sense, language assessment is defined by Purpura as 'a systematic procedure for eliciting test and nontest data (e.g., a teacher checklist of student performance) for the purpose of making inferences or claims about certain language-related characteristics of an individual' (ibid: 191). According to Purpura, the objective of L2 assessment is to obtain 'performance consistencies' that can be recorded and subsequently 'used as evidence for making decisions' (ibid: 191).

In other words, according to Bachman and Palmer (2010, quoted in Purpura 2016), L2 assessments are carefully planned starting from some sort of syllabus, they have specific purposes even if such purposes appear to be implicit and they can be arranged in a continuum, from controlled (standardized tests such as *TOEFL* or unit tests) to informal (teacher observations during classes).

An important point raised by Purpura (2016) in his definition of language assessment, is the aspect of interpretation, that is to say the use performance records in order to make high stakes or low stakes decisions. In this sense, Purpura (2016: 192) draws a distinction between assessment and what he refers to as 'evaluation', namely 'the systematic use of evidence for the purpose of making decisions' in at least three different contexts:

- classroom context: decisions about readiness (i.e., diagnosis), progress, mastery, retention, and promotion or decisions about teaching, learning, and curricula.
- school context: decisions about proficiency, selection, and placement
- society context: decisions about certification and accountability (ibid, 2016: 192).

According to Purpura (2016), the most challenging aspect of L2 assessment is the comprehension and the definition of the construct of L2 proficiency.

In Europe, the Common European Framework of Reference for Languages (CEFR), elaborated by the Council of Europe (2001), carries great influence on how languages are learnt, taught and assessed (Lim, 2013). The aim of the CEFR, as stated in the document, is to provide 'a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe' (2001:1) in an effort to promote plurilingualism, a concept that underpins the framework. The promotion of plurilingualism implies a paradigm shift for language education, which adopts the new aim: to develop a lifelong 'linguistic repertory' based on a communicative competence aimed at integrating and interrelating different language experiences (2001: 4-5). Of particular relevance for this study, is the coverage of assessment in the CEFR. If the majority of the space is given to language use, language users and teaching, Chapter 9 – which is the final chapter of the document – is entirely devoted to assessment. The chapter defines three main uses of the CEFR:

- specify the content of tests and examinations.
 - set the criteria for the attainment of a learning objective, both in relation to the assessment of a particular spoken or written performance, and in relation to continuous teacher-, peer- or self-assessment.
 - describe the levels of proficiency in existing tests and examinations thus enabling comparison to be made across different systems of qualifications.
- (2001: 19).

Furthermore, the chapter considers the four fundamental issues of validity, reliability, feasibility and selectivity of assessment systems as forming the substratum of the framework. By validity, it is meant the degree of accuracy to which a test or assessment

represents the proficiency of the learner, while reliability refers to the replicability of results and is considered in light of the accuracy of judgements based on the criteria made by assessors. Feasibility involves some basic constraints (time, external limits) and the need to select a few criteria in order to produce a practical assessment procedure that can easily be administered (2001: 177-178). Finally, selectivity refers to the fact that users may choose to adopt simpler operational schemes, thus merging together different categories that appear in the Framework as separate units (2001: 178).

At the beginning of the chapter, it is stated that language tests are one among many forms of assessment, such as continuous assessment or informal observation, and the chapter proceeds to outline the difference between different types of assessment, as summarised in Figure 2:

1	Achievement assessment	Proficiency assessment
2	Norm-referencing (NR)	Criterion-referencing (CR)
3	Mastery learning CR	Continuum CR
4	Continuous assessment	Fixed assessment points
5	Formative assessment	Summative assessment
6	Direct assessment	Indirect assessment
7	Performance assessment	Knowledge assessment
8	Subjective assessment	Objective assessment
9	Checklist rating	Performance rating
10	Impression	Guided judgement
11	Holistic assessment	Analytic assessment
12	Series assessment	Category assessment
13	Assessment by others	Self-assessment

Figure 2 – Types of assessment (CEFR, 2001:183).

This categorization has been discussed and reorganized in the context of the *Encouraging the Culture of Evaluation Project (ECEP)*, supported by the Council of Europe aimed at investigating the impact of CEFR on assessment attitudes and practices of language teachers

(Piccardo, 2012). The data that emerged from the study suggested that teachers were overwhelmed and disoriented by the classification of the thirteen different assessment pairs, as it requires a lot of targeted and contextualised decisions (ibid, 2012). Therefore, the researchers re-grouped the pairs into four macrocategories (Figure 3):

MACROCATEGORY	PAIRS
Distinguishing competence and action	1, 6,7
Assessment: a question of timing?	2, 3, 4
The issue of objectivity in assessment	8, 10, 11, 12
Assessing: reasons, actors, modalities	5, 9, 13

Figure 3 – Macrocategories of Assessment types (Piccardo, 2012).

In the researchers' view, this system maintained the complexity but made it more manageable, besides being 'a sort of guiding map to help practitioners understand the 13 pairs, to link them to other layers of the assessment process and to see in what way all of these distinction would inform and improve their own practice' (Piccardo, 2012: 49).

According to Piccardo (2012), the main objective of the CEFR is to encourage an organic reflection on assessment as an integral part of teaching and learning and the vision it promotes is that of a multidimensional approach to assessment. The CEFR is a 'complex document', but it merely reflects the complexity of the teaching/learning process in general, and of the assessing process in particular, so the approach to the tools it offers needs to be a receptive one that embraces the complexity and transforms it into a viable asset point (ibid, 2012: 51-52).

Despite being a European document, the CEFR is used in many countries around the world, but its use is not homogeneous. In fact, it can be located in a continuum, from a cursory alignment of certification of proficiency and admission test for language requirements at university level, to a deeper role in curriculum reform and instructional innovation (Piccardo, 2019). According to the CEFR, assessment is a dynamic and composite process that goes beyond the use of standardized tests to measure a learner's language proficiency,

and according to Piccardo (2019: 2), considering the CEFR as a mere assessment tool 'reveals a limited vision of what the CEFR is'.

From the perspective of the field of second and foreign language, the disciplinary area of assessment benefits and is influenced by second language acquisition (SLA) research and theories of learning (Alderson et al., 2017). Within this intersection, among the various paradigms that have been proposed over recent years, the current research project focuses on the area of learning-oriented assessment (Turner & Purpura, 2015).

1.2.3 A multidimensional approach to Classroom-based Assessment

From the perspective of the current piece of research, classroom-based assessment (CBA) is the most interesting area of investigation, because it taps into the connection between teaching, assessing and learning.

The classroom, as a physical educational context, has been defined as the 'crucible' of language learning (Allwright & Bailey 1991: 18, citing Gaies, 1980), that is to say the place where the agents of learning (students and teachers) converge and interact with each other. Bailey and Allwright point out that along with their initial reaction, 'their constant *interaction*' is a key element for a successful learning experience (1991: 18, emphasis in original) and this relationship is illustrated in Figure 4. Every lesson happens in a social theatre of sorts and is a 'co-production' of teachers and learners, who have to operate with different factors – including participants' turn distribution, topic, task, tone and code – in order to create meaningful opportunities so that every student has the chance to manage their own learning (ibid, 1991: 19). According to Bailey and Allwright (1991), there are three main aspects that teachers consider when planning their classroom language lessons: syllabus, method and classroom atmosphere. This last aspect, according to them, relies strongly on cooperation between teacher and students, and as the final product is the result of interdependent mechanisms, each lesson will be unique, even if the planning is the same (ibid, 1991). Bailey and Allwright (1991) identify three components of a lesson's outcome (input, practice opportunities and receptivity), that 'refer to what actually happens in the classroom, regardless of whatever was planned to happen' (ibid, 1991: 24).

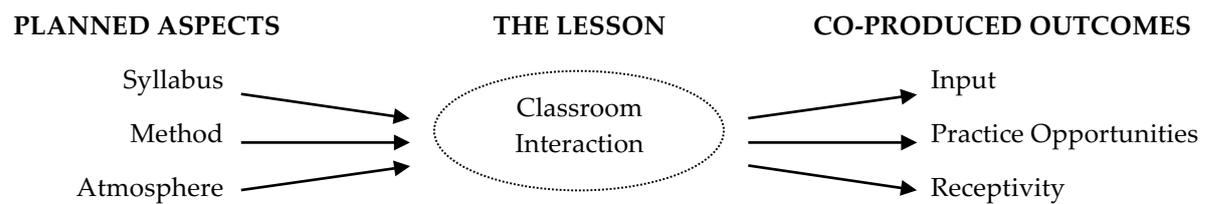


Figure 4 – The relationship between plans and outcomes (Bailey and Allwright, 1991: 25).

The model outlined by Bailey and Allwright (1991) focuses on the classroom because classroom processes determine and shape students' learning opportunities and the goal of classroom research is to deepen our knowledge of such processes. For this reason, a better understanding of classroom-based assessment is vital in order to sustain effective language learning.

Research into CBA has been deeply shaped by Torrance and Pryor's concept of the 'micro-sociology' of classroom assessment, described as the study of 'how assessment of young children is carried out in the classrooms, and with what possible consequences for their understanding of schooling and the development of their learning in particular subject areas' (Torrance & Pryor, 1998: 3). Their innovative approach to classroom assessment entailed regarding 'interactive assessment incident' as the basic analysis unit, and drawing from the fields of classroom interaction studies, social constructivist learning theories and motivation and attribution theories (ibid, 1998), an approach that appears to be still relevant for a contemporary enquiry on the subject.

As pointed out by Rea-Dickins, if formal testing occurs regularly in the school year, CBA is 'routinely embedded' within daily teaching activities that 'allow teachers to make decisions about their learners [...] and may be influential in determining what is taught next and how that material is taught' (2001: 434). Some of the mentioned strategies include questioning and probing, single and group interactions with the teacher, observed interactions among learners and feedback on learners' submitted work (ibid, 2001).

In second and foreign language (L2) classrooms, assessment has the potential to become a bridging device between teaching and learning (Colby-Kelly & Turner, 2007). Starting from

this premise, the learning-oriented assessment (LOA) approach focuses on the learning process and aims at finding new ways for teachers and learners to benefit from the information generated by planned and unplanned assessment opportunities (Purpura and Turner, 2015). Zeng et al. (2018) in their critical review of literature on LOA state that the term learning-oriented assessment (LOA) was coined by Carless (2007), even though the term appears previously in a book on assessing grammatical ability, authored by Purpura (2004)³.

Carless (2007) underlines the fact that ‘in LOA, learning comes first, both in the way the term is literally constructed, and as a matter of the principle of emphasising the learning aspects of assessment’ (2007: 5) and introduces a framework for LOA (Figure 5).

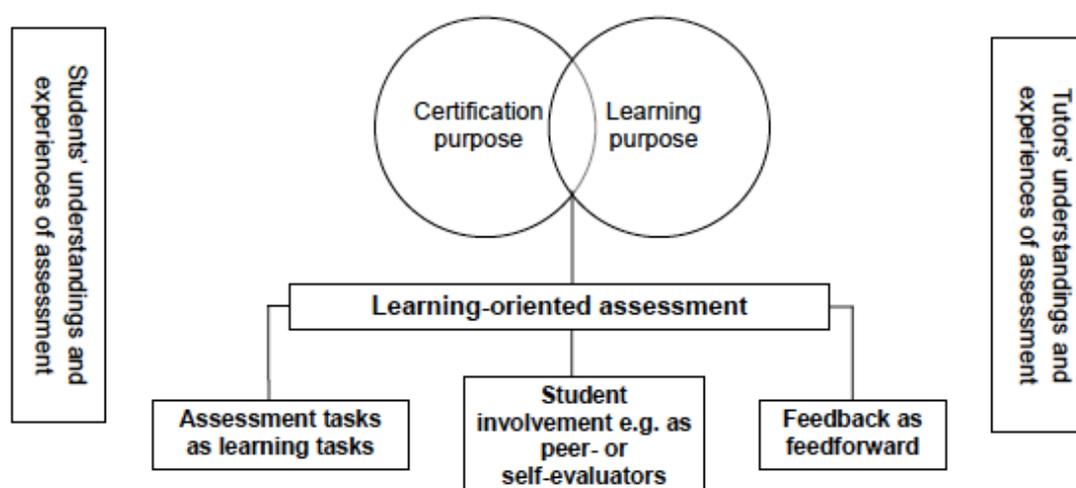


Figure 5 – Framework for learning-oriented assessment (Carless, 2007: 8).

In the model, the two main purposes of assessment are overlapping, meaning that efficient assessment is taking place; furthermore, the model identifies three interactive elements of LOA, (a) assessment tasks as learning tasks, (b) student involvement, and (c) feedback as feedforward (Carless, 2007). Three principles summarize this model:

1. Assessment tasks should be designed to stimulate sound learning practices amongst students;

³ In particular, Chapter 8 is titled “Learning-oriented assessment of grammatical ability”.

2. Assessment should involve students in actively engaging with criteria quality (their own and/or peers' performance);
3. Feedback should be timely and forward-looking so as to support current and future student learning (Carless, 2007: 8).

Learning-oriented assessment in second and foreign language classrooms has been investigated by Turner and Purpura (2015: 261), who proposed the following working framework of LOA:

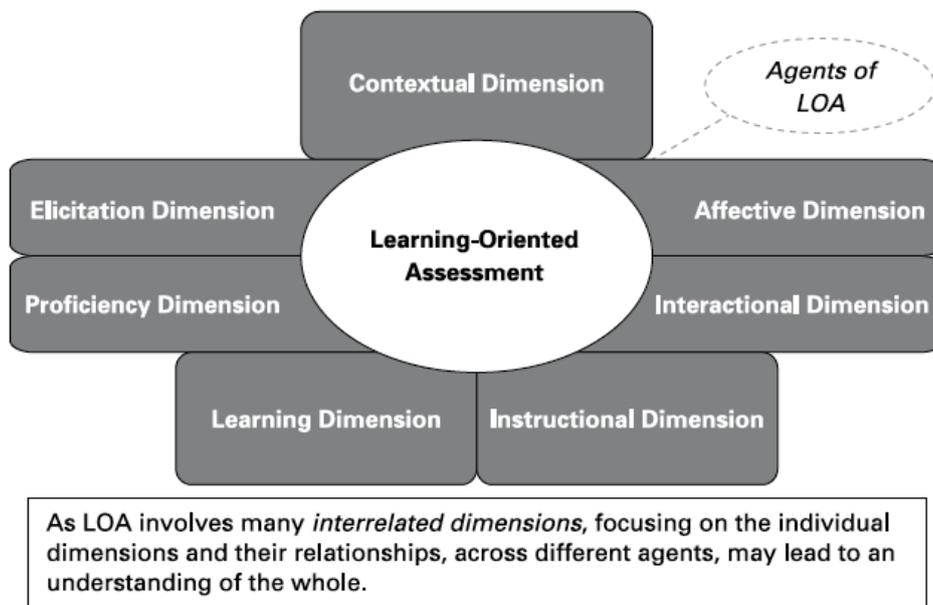


Figure6– Working Framework for learning-oriented assessment (Turner & Purpura, 2015: 261).

As for the contextual dimension, the factors that influence instruction and assessment include socio-political forces, sociocultural norms, personal attributes of teachers, teacher choices and a variety of students' personal attributes (ibid, 2015). The elicitation dimension represents a fundamental part of LOA, as it includes all activities (planned and unplanned) aimed at eliciting language inside the classroom, and is detailed in Figure 7 (ibid, 2015: 264):

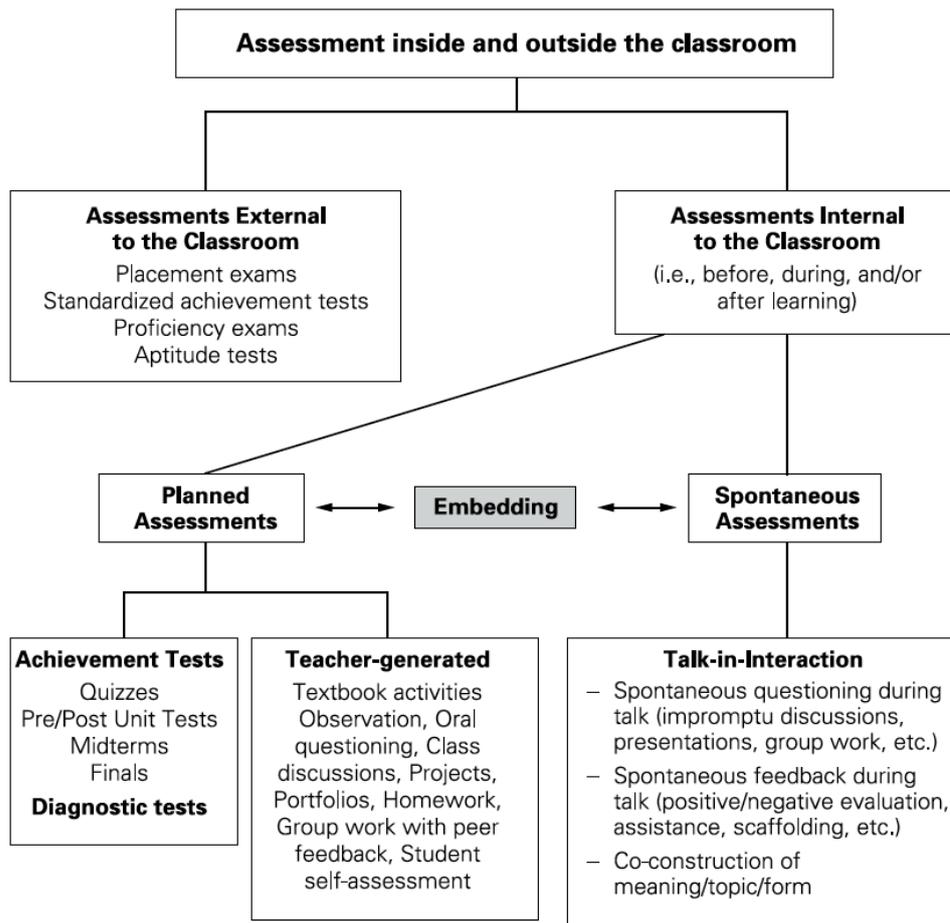


Figure 7 – Ways to elicit language for assessment (Turner & Purpura, 2015: 264).

The L2 proficiency dimension represents the cornerstone of LOA because teachers need to focus on the learning objectives, and they need to have a thorough awareness of L2 content knowledge, knowledge of SLA, L2 pedagogical content knowledge and topical content knowledge (ibid, 2015). The learning dimension is concerned with how learners learn the L2, and deals with the role of feedback and self-regulation in promoting L2 learning, while the instructional dimension focuses on teachers' knowledge of the language and subject matters, and their knowledge of pedagogical procedures (ibid, 2015), including assessment literacy (Purpura, 2016). The interactional dimension is connected to the elicitation dimension, as elicitation activities happen during talk-in-interaction contexts which result in 'exchange patterns' evaluating (positively or negatively) learners' achievement of a learning objective and 'assessment-type repair sequences' (Turner & Purpura, 2015: 268).

Finally, the affective dimension is another core element of LOA, and it includes learners' emotions, beliefs, attitudes, personal features and motivation (ibid, 2015).

1.2.4 Diagnostic competence and Teachers as Assessment Agents

The role of teachers in assessment practices is particularly evident within theories of teacher-based assessment, described as 'nonstandardized local assessment carried out by teachers in the classroom' (Butler, 2009: 418, citing Leung, 2005: 871). This assessment process represents a professional challenge for teachers, who often find themselves between hammer and anvil with regard to their dual role as facilitators and examiners (Stiggins, 1999; Rea-Dickins, 2004). Indeed, according to Butler (2009), the challenges linked to teacher-based assessment seem to stem from 'the dilemma between the pedagogic and measurement aspects of such assessments', besides the fact that teachers are increasingly being asked to adhere to accountability requirements based on national standards (2009: 418).

Furthermore, research findings suggest that teachers experience a strong tension between their inner values and many different external variables (such as national testing, district accountability and parents' expectations) and this pressure has major implications for teachers' evaluating and grading choices (McMillan, 2003). McMillan (2003) argues that teachers' reflective decision-making plays a fundamental role in classroom assessment and assessment skills such as reliability and validity influence teachers' interpretations and assumptions about student work. Teachers' assessment decision-making has been defined as 'a process in which teachers balance the demands of external factors and constraints with their own beliefs and values to determine classroom assessment practices' (McMillan, 2003: 42).

The paradigm shift from AoL to AfL and AaL that has been discussed in the previous sections (1.2.1. and 1.2.2.), has strongly influenced the notion of teachers' assessment abilities (Brookhart 2003, quoted in Firoozi et al., 2019). In greater detail, the notion of teachers' assessment literacy encompasses 'a core professional requirement across educational systems [...] involv[ing] the ability to construct reliable assessments, and then

administer and score these assessments to facilitate valid instructional decisions anchored to state or provincial educational standards' (DeLuca et al., 2016: 251-252).

Throughout their career, foreign language teachers are confronted with multiple assessment techniques ranging from high-stakes tests and external tests⁴ to informal formative assessments, other than innovative forms of assessments such as self-assessment or peer-assessment, as indicated in the CEFR (2001) and in the European Language Portfolio (2000) and are consequently asked to acquire new assessment skills (Vogt & Tsagari, 2014).

In the field of language testing and assessment (LTA) research, FL teachers' assessment literacy has been questioned with respect to the scarcity of training and also concerning the overall propensity of teachers to apply the same assessment techniques they experienced in their school history (Vogt & Tsagari, 2014).

Starting with the premise that teachers play a central role in assessment, Rea-Dickins focuses on teacher assessment, i.e. the teacher as *agent* of the assessment (2004: 252, emphasis in original). In her view, the teacher-assessor interacts with assessment at different levels (individual, cultural, professional and institutional and social), each of which has the power to enhance or hinder the processes of classroom assessment (ibid, 2004: 255).

The prominent role of FL teachers in assessment is of paramount importance for the present research project, and a crucial construct that will be central to the analysis process (see paragraph 2.3.4) is that of 'diagnostic competence'; it is a term introduced by Edelenbos and Kubanek-German, who defined it as 'the ability to interpret students' foreign language growth, to skilfully deal with assessment material and to provide students with appropriate help in response to this diagnosis' (2004: 260).

In other terms, this kind of assessment skill has to do with the monitoring of students' progress rather than performance, and this aspect of assessment in language education is also featured in the Barcelona strategy, developed by the European Commission (2002).

⁴ In the Italian context, for example, teachers of English as a foreign language have to deal with English examinations provided by the national institute for the evaluation of the educational system (INVALSI). The INVALSI tests for English are administered to children in grades 5, 8 and 13.

Through exploring assessment activities in FL primary classrooms, Edelenbos and Kubanek-German (2004) have identified six levels of diagnostic competence whose descriptors merit to be quoted at length:

1. Intuitive observations, isolated observations; observations geared to the whole class; observation geared towards the dominant students
2. Common knowledge about children as language learners; comments based on own learning history; first realizations that more contextual clues would be needed to interpret the situation
3. Ability to administer and correct ready-made test sheets in an appendix of a textbook, to classify students as good, average or weak learners; ability to bring together various observations (e.g., those gained through exchanges with colleagues), seeing language growth with a pedagogical eye ('he likes the chants best').
4. Ability to write a report with many prefabricated formulas; emerging ability of the teacher to create a narrative about an individual learner; ability to understand categories of diagnosis; ability to compare children using these categories; ability to judge a child's silence; ability to situate in terms of his/her own language acquisition trajectory; mental ongoing diagnosis leading to expansion, code-switching, etc.
5. Teacher becomes aware of how his/her categories and diagnosis are influenced by teaching style and professional biography (reflective practitioner); ability to understand basic features of simple tests; ability to select and adapt such assessment instruments as contained in the various teacher handbooks; ability to correct and explain his/her expectations
6. 'Pädagogischer Takt': ability to promote optimum student learning, i.e., to further autonomy, language awareness, self-assertiveness; to be able to give a rich hermeneutic interpretation of a language learning situation,

which becomes even more complex through dialogue, as an ongoing process (Edelenbos & Kubanek-German, 2004: 278-279).

The researchers explain that the sixth level is more of an ideal stage, but even in the actual classroom there are instances of 'fruitful moments' (ibid, 2004: 279) of learning apogee. As is apparent from the description of this last level, diagnostic competence appears to be a prerequisite for gainful assessment, it is 'the power to see more' in the classroom (ibid, 2004: 277).

Another area of assessment research that is closely linked to the present project entails the emotional challenges teachers face when evaluating or assessing pupils. Teaching is a social act that requires 'emotional knowledge' other than content knowledge, and the affective elements of teaching play an important role in shaping their relationship with the social contexts they work in and in defining their sense of self (Zembylas, 2005).

Teachers present different emotions towards assessment procedures, but generally they feel positively about formative assessments aimed at sustaining learning, while negative emotions involve assessment that evaluate pupils or the quality of their teaching thus affecting teachers' reputation (Brown et al. 2018). More specifically, formative assessment puts teacher in the contradicting position of being both supportive teachers who collect data in order to sustain learning and objective evaluators who judge the same learners against some kind of standard (Brown et al. 2018). Teachers experience positive emotions such as satisfaction, pride or joy when student reach the learning goals (Xu, 2013, quoted in Brown et al., 2018), but in case of negative performances, they struggle with how to communicate poor results and how to face students after giving them negative feedback, and they feel guilty for their failure (Brown et al., 2018). Another interesting aspect is that 'the emotions of students create challenging conditions for teachers' (Brown et al., 2018: 211), and teachers put great effort in establishing positive relations with their students. Brown et al. (2018), underscores the fact that while teachers of other subjects use the students' native language and assess more objective concepts, L2 teachers share their colleague's emotions but they face additional challenges, as 'their work circumstances press [...] negative feelings to the

surface more readily than in other subjects' (2018: 218). Brown et al. (2018) also found that teachers feel judged if their students fail to achieve and rely on their colleagues to cope with negative emotions derived from the assessment process.

The research carried out by Loh and Liew (2016) aims at highlighting the discipline-specific nature of the 'emotional labour' of English teachers, and it shows that assessment practices are stressful experiences that require significant emotional investments due to high expectations about language requirements needed to progress through the school levels. The authors recommend teacher educators to provide their students with real-life case studies or teacher-authored narratives, as they might prove useful for student teachers, who can 'become more critically aware of the emotional dimensions of teachers' practices, behaviours, and beliefs' (2016: 277).

Starting from the assumption that classroom-based assessment is 'teacher dependent' (Webb, 2004: 3) Pellegrino et al. (2001, quoted in Webb, 2004) suggested a model of evidence-based classroom assessment involving three interrelated elements (Figure 5).

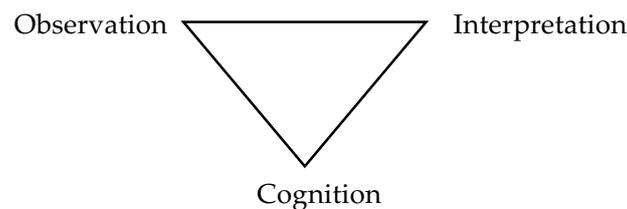


Figure 8 – Assessment Triangle (Webb, 2004: 3).

Cognition refers to teachers' idea of how learners depict understandings and progress in the discipline, interpretation has to do with how teachers understand student performance and observation implies eliciting student answers. In other words, internal (classroom-based) assessment 'largely depends on teacher judgments about student thinking' (Webb, 2004: 7) but this procedure is influenced by a variety of factors such as lesson duration, number of students, and other aspects of school organization. Other factors, such as the willingness of students to share their thinking hinges on teachers' ability to create a safe environment that produces effective assessment opportunities (Webb, 2004).

1.2.5 Assessing Young Language Learners

Before providing some theoretical references about the young language learners (YLL) assessment field, it is important to bear in mind that young learners represent a special and unique learning group. In order to understand its features and complexities, in discussing the theoretical background to second language acquisition and pedagogy in young children, Pinter (2011) refers to fundamental theories of child development promoted by Piaget and Vygotsky, besides examining theories of information processing and emotional development, the breadth of which exceeds the scope of the present research project.

The focus of this paragraph will be restricted to theories and issues pertaining to assessment needs of young learners and to their special characteristics as learners.

Within the field of foreign and second language (L2) teaching, the expression 'young learners' has been utilized to define very different target populations according to their age group, but essentially it goes from early childhood up to pre-adolescence or adolescence (Timpe-Laughlin and Cho, 2021). In some studies, the term has been used to refer to children from about 5/6 to 12/13 years old (Rea-Dickins, 2000; Hasselgreen, 2005; McKay, 2006; Nikolov, 2016), while in others the term encompasses learners from 3 up to 17 years old (Pinter, 2011). The discussion reported in the present research will be limited to young language learners (YLL), as defined by Hasselgreen:

Young language learners (YLLs) are defined here as being school pupils up to around 13 years old, thus incorporating the primary school population in most European countries, and impinging on the lower levels of secondary schooling in many (2005: 338).

Young language learners resemble adult learners in bringing to the class 'their own personalities, likes and dislikes and interests, their own cognitive styles and capabilities and their own strengths and weaknesses' (McKay, 2006: 5), but their individuality as learners is associated to some specific characteristics that can be grouped into three main areas: growth, literacy and vulnerability (ibid, 2006).

Regarding the first area, 'children are in a state of constant cognitive, social, emotional and physical growth' (McKay, 2006: 6), they tend to present limited attention span, a newly

emerging meta-language and a narrow experience of the world and therefore the assessment tasks should be limited and tangible (ibid, 2006). As for their social growth, their emotional sphere is affected by the feedbacks they receive from adults and peers, so they might suffer if they feel criticized or excluded (ibid, 2006). These feelings, associated with a gradual understanding of group dynamics, need to be taken into consideration when structuring oral tasks or interactive tasks such as individual presentation or team games (ibid, 2006). Teachers also need to consider physical development when designing language assessments, with regard to motivation, tiredness, hand-eye coordination and the ability to sit still (ibid, 2006).

As for the area of literacy, it is important to be aware that children might be developing their literacy skills in both their L1 and the target language, but their first language literacy skills might be a boost or a hindrance for their foreign or second language acquisition (ibid, 2006). Finally, in discussing issues of vulnerability, McKay (2006) writes that 'children have a heightened sensitivity to praise, criticism and approval and their self-esteem is strongly influenced by experiences at school', and in order for all children to experience a sense of success and progress, assessment tasks should be flexible and include data from observation (ibid, 2006).

When examining best practices in the assessment of children, Serragiotto highlights the need to pair a '*learning by doing*' teaching methodology with a '*assessing and evaluating by doing*' (2016: 131, emphasis in original) approach. Therefore, he suggests assessing children's initial stages of learning by asking them to act out their language knowledge by drawing, using gestures, doing things and so on (ibid, 2016). Another important feature of assessing young learners is to do so by adopting a playful approach and using icons such as smiles or emoticons in order to motivate them (ibid, 2016). Finally, implementing continuous monitoring and self-assessment practices seems to be vital in order to verify the acquisition and keep track of their emotional and psychological attitudes towards the foreign language (ibid, 2016). For this purpose, the teacher may want to keep a '*diario di bordo*' (teaching log) to annotate the activities carried out in the classroom, the issues, and the children's behaviours and reactions during the tasks (Serragiotto, 2016: 132).

Monitoring and observation is singled out as efficient ways of assessing children within a 'natural approach', so as to avoid any interference while the learners are playing and producing language (Novello, 2014: 29). As for the object of these observations, Novello suggests three main areas to be assessed by teachers:

- la motivazione nei confronti della lingua;
- la curiosità e l'attitudine verso nuovi bisogni linguistici;
- le abilità comunicative legate agli obiettivi della programmazione (2014: 28).

In fact, Novello (2014) specifies that when working with younger learners the communicative goals should mainly concern the speaking skills developed in a known context and their evolution in the progressive introduction of new content.

Rea-Dickins' (2000) seminal contribution on assessment in early years language learning contexts paved the way for a fruitful debate on the definition of some urgent areas of concerns related to the practice of assessing learners in their early years of language education. Some of the first issues in assessing young learners included:

- processes and procedures used by teachers to inform teaching and learning;
- assessment of achievement at the end of the primary phase of education; and
- teachers' professional development (Rea-Dickins, 2000: 119).

Almost two decades later, the field of YLLs assessment has branched off to investigate specific aspects connected to young learners research, and has also expanded geographically, by including populations from non-European backgrounds such as Asia, North America, Africa, Japan, Armenia to name a few (Timpe-Laughlin & Cho, 2021).

Moving on to consider the specific features of assessment in relation to this peculiar target group of pupils, Hasselgreen points out that the assessment techniques should comply with the following principles:

1. Tasks should be appealing to the age group, interesting and captivating, preferably with elements of game and fun.
2. Many types of assessment should be used, with the pupil's, the parents' and

the teacher's perspectives involved.

3. Both the tasks and the forms of feedback should be designed so that the pupil's strengths (what he or she can do) are highlighted.
4. The pupil should, at least under some circumstances, be given support in carrying out the tasks.
5. The teacher should be given access to and support in understanding basic criteria and methods for assessing language ability.
6. The activities used in assessment should be good learning activities in themselves (Hasselgreen, 2005: 338-339).

The literature suggests that the development of YLLs foreign language competence is slower if compared to pupils who start at a later stage (Nikolov & Mihaljević Djigunović, 2011). An early start into language learning could also result in difficulties for the teachers to maintain pupils motivated over a longer period of time, because 'the earlier L2 learning is introduced, the sooner typical classroom activities and topics become boring for young learners' (Nikolov, 2016: 3). However, young language learners tend to show more positive attitudes toward FL learning than their older peers, which tend to decline over time (Henry & Apelgren, 2008). Other studies focusing on formative assessment techniques with young learners – such as Butler and Lee's (2010) investigation into the effects of self-assessment among young learners – report that eight-year-olds are able to provide accurate self-assessments provided that the assessment items are coherently embedded in the classes. According to Butler (2009), at earlier stages of language education (i.e. at the primary level) policymakers encourage the adoption of teacher-based forms of assessment to the detriment of traditional paper-and-pencil tests, as source of both formative and summative data. Compared to their older peers, young pupils are less familiar with high-stakes examination processes (Hasselgreen, 2005), therefore language classrooms are a familiar setting for age-appropriate teacher-assessment procedures aimed at minimizing affective barriers (Hasselgreen, 2005). More specifically, YLLs tend to use their experiences to actively construct knowledge, and learn in a social environment, so teachers need to 'regard the

pupils socially, psychologically, cognitively and physically as a whole' (Sarigöz & Fişne, 2018: 381).

Another central node of the debate on early language learning, concerns the variability and the diversity characterising assessment in national programs across Europe and non-European contexts (Rea-Dickins, 2000). In Norway, for example, educational policies instruct against formally assessing young language learners and there is 'absolutely no tradition of testing in the primary school' (Hasselgreen, 2000: 262); in Italy, on the contrary, assessment is considered to be a positive aspect of primary education, and it covers non-linguistic dimensions of learning such as effort or interest, but the evaluation system is not standardized, so a considerable degree of variability is present across the country (Gattullo & Pallotti, 2000).

Through exploring more in detail the formative assessment practices conducted in primary schools in Italy, Gattullo (2000) claims that language teachers initiate some basic formative assessment cycles (i.e., questioning, negotiating, correcting, judging), but seem to be oblivious of the further step, namely making 'productive use of information they collect' so that student responses 'are not fully exploited for the potential insight that they may provide into the language learning process' (ibid, 283).

By way of conclusion, a special approach to assessment in early years learning contexts is strongly recommended in order to protect and sustain young pupils in their language learning endeavours, because 'assessment has the power to change children's lives' (McKay, 2006: 25).

1.2.6. Assessment in Italian primary schools: a new legislation

Primary (Elementary) Schools in Italy are part of the first cycle of education, together with Lower Secondary Education (Scuola Secondaria di Primo Grado). There are no state examinations between the two levels of the first cycle of education⁵.

⁵ Information retrieved from the European Commission's database on National Education Systems (Eurydice), accessible from: https://eacea.ec.europa.eu/national-policies/eurydice/content/italy_en

Starting from the school year 2020-2021, there has been a reform of the periodic and final assessment and evaluation procedures in the Italian Primary (Elementary) Schools. Before describing in detail the changes that have been made and the new requirements, it appears useful to provide an historical overview on assessment in Primary Education in Italy.

In Italy, the history of educational assessment is intrinsically linked to the history of pedagogy and education and to the evolution of schooling itself, and academic literature on the early years (from the unification of Italy until the second half of the 1970s) is relatively scarce (Tomasi, 2021). The first assessment system was based on a classist view of the Italian society and its aim was to alphabetize the masses and implement a rigid selection procedure thanks to the adoption of a ten-point grading scale ranging from 1 ('poorly', 'pessimamente') to 10 ('optimally', 'ottimamente') (Tomasi 2021, 67-68). The instrument of numerical grading was aimed at selecting students according to their performances in the classroom, with the consequence of preventing pupils who demonstrated poor skills from accessing higher grades of education (ibid, 2021). With the rise of Fascism, this system is retained and reinforced, and the reform of 1923 prescribes an assessment that is still summative in nature, with the adoption of adjectives instead of numbers and then again the re-introduction of the ten-point grading scale in an attempt to uniform the elementary level with the upper secondary level (Dal Passo, 2003). After the proclamation of the Republic, the emergence of new ideologies and disciplines such as docimology and the theories of Dewey and Freinet impacts the way education and assessment are interpreted, but these changes are limited to the academic spheres of the Italian society, and the classroom practices remain unaffected (Dal Passo, 2003). In the period 1950s-1970s, literacy and schooling rates are still rather low and the assessment mechanisms of the Italian education system are criticized by several parties, such as the Educational Cooperation Movement (Movimento di Cooperazione Educativa, MCE), Don Lorenzo Milani and several other student protest movements (Tomasi, 2021). In 1977, the Italian government issued a watershed law⁶ that, for the first time ever, stated the formative value of assessment and

⁶ Legge 4 Agosto 1977, n. 517. GU Serie Generale n. 224 del 18-08-1977. Law 517, issued on the 4th of August, 1977; Published in the Official Gazette of the Italian Republic on the 18th of August, 1977. Available at: <https://www.gazzettaufficiale.it/eli/id/1977/08/18/077U0517/sg>.

introduced a new report card that focussed on learning processes and on the pupil as an active member of a learning community (ibid, 2021). In the second half of the 1990s, the demands for a simplification of the report cards in the primary school led to the introduction, in 1966, of a new system based on individual summary assessments for each subject, expressed by a scale of adjectives (excellent, very good, good, sufficient, insufficient⁷). Another interesting development in the assessment regulations is the reform introduced in 2003 by the Minister for Education Letizia Moratti, that introduced a new element, the portfolio for the personal competences of the pupil, even though this was short-lived, in that the next Minister, Roberto Fioroni, abolished the portfolio in 2006 (Tomasi, 2021). Another sudden change in the history of assessment of the Italian education system was brought about by the reform introduced by the Minister for Education Mariastella Gelmini in 2008; this was a very controversial reform that reintroduced the ten-point grading scale and foresaw massive budget cuts that had negative consequences for the public schooling system (ibid, 2021).

The following reform, named 'Buona Scuola'⁸ (literally, 'Good School') was approved in 2015 and operationalized in 2017, with a decree law confirming the numerical system in tenths for periodical and final grading of primary school pupils, in addition to integrating a descriptive report about the pupils' learning process and about the global level of learning development (ibid, 2021).

Finally, due to the Covid-19 pandemic emergency that in March 2020 forced the general lockdown of the educational system, distance learning prompted the Italian Government to re-imagine the whole evaluation system, found to be utterly inadequate, especially for young learners (ibid, 2021). With the publishing of a new decree law⁹ in April and a law in June 2020, the Ministry for Education tackles the issues of learning in an online environment and introduces some innovations regarding assessment. Most recently, a Ministerial Order¹⁰

⁷ The original terms in Italian are: ottimo, distinto, buono, sufficiente, insufficiente.

⁸ Law n. 107, 13th July 2015.

⁹ Law decree n. 22, 8th April 2020.

¹⁰ M.O. n. 172, 4th December 2020.

published in December 2020 together with a document containing guidelines¹¹ on assessment, introduces a new report card to be adopted by primary schools, and to be used for the mid-term evaluation at the end of the first term, in February 2021 and for the final evaluation in June 2021. The reformed report card provides for the definition of a number of learning goals for each subject, the attainment of every one of whom will be assessed by adopting a four-level scale (1- in via di prima acquisizione/in the process of being acquired; 2- base/basic; 3- intermedio/intermediate; 4- avanzato/advanced). The levels are defined according to four dimensions associated with the learning process, namely (a) learner autonomy, (b) known or unknown context, (c) resources deployed by the learner and (d) consistency in manifesting learning progress. As for the subjects' learning objectives, these must be formulated in accordance to the National Guidelines for the Curriculum of Kindergarten and the First Cycle of Education¹², henceforth *National Guidelines* (2012). In the document containing guidelines on the new assessment procedures, henceforth *Guidelines* (2020), the minister highlights the fact that

la valutazione ha una funzione formativa fondamentale: è parte integrante della professionalità del docente, si configura come strumento insostituibile di costruzione delle strategie didattiche e del processo di insegnamento e apprendimento ed è lo strumento essenziale per attribuire valore alla progressiva costruzione di conoscenze realizzata dagli alunni, per sollecitare il dispiego delle potenzialità di ciascuno partendo dagli effettivi livelli di apprendimento raggiunti, per sostenere e potenziare la motivazione al continuo miglioramento a garanzia del successo formativo e scolastico (2020: 1).

The focus, therefore, is on teachers as agents of evaluation and on the pupils' learning progress accomplished through a formative kind of assessment that gives value to the

¹¹ Ministero dell'Istruzione, dell'Università e della Ricerca (2020). Linee Guida. La formulazione dei giudizi descrittivi nella valutazione periodica e finale della scuola primaria. Retrieved April 27, 2021, from: https://www.istruzione.it/valutazione-scuola-primaria/allegati/Linee_Guida.pdf.

¹² Ministero dell'Istruzione, dell'Università e della Ricerca (2012). *Indicazioni Nazionali per il curriculum della scuola d'infanzia e del primo ciclo di istruzione*, published in 2012 with the ministerial order n. 254, 16th November 2012.

progressive knowledge-building processes carried out by the pupils. In the following paragraphs, the text makes explicit reference to *assessment for learning* (*valutazione per l'apprendimento*, emphasis in original) as the underlying principle underpinning this new evaluative system based on formative assessment that aims at enhancing learning.

The text specifies that the learning objectives need to represent observable evidence of learning and be coherent with the goals for the development of skills. The learning objectives, therefore, must be formulated so as to include both the *action* executed by the pupils and the corresponding *disciplinary content* (2020: 3, emphasis in original)¹³. The document elaborates further on the four dimensions underlying the definition of the learning levels:

1. *l'autonomia* dell'alunno nel mostrare la manifestazione di apprendimento descritto in uno specifico obiettivo. L'attività dell'alunno si considera completamente autonoma quando non è riscontrabile alcun intervento diretto del docente;
2. la *tipologia della situazione* (*nota o non nota*) entro la quale l'alunno mostra di aver raggiunto l'obiettivo. Una situazione (o attività, compito) *nota* può essere quella che è già stata presentata dal docente come esempio o riproposta più volte in forme simili per lo svolgimento di esercizi o compiti di tipo esecutivo. Al contrario, una situazione *non nota* si presenta all'allievo come nuova, introdotta per la prima volta in quella forma e senza specifiche indicazioni rispetto al tipo di procedura da seguire;
3. le *risorse* mobilitate per portare a termine il compito. L'alunno usa risorse appositamente predisposte dal docente per accompagnare il processo di apprendimento o, in alternativa, ricorre a risorse reperite spontaneamente nel contesto di apprendimento o precedentemente acquisite in contesti

¹³ More in detail, the document specifies that the action refers to the cognitive process activated by the learner, and therefore must be expressed through verbs such as list, connect, generalise, distinguish, classify, exemplify, and so on. Furthermore, the disciplinary contents may be additionally categorised as factual, conceptual, procedural and metacognitive (2020: 3).

informali e formali;

4. la *continuità* nella manifestazione dell'apprendimento. Vi è continuità quando un apprendimento è messo in atto più volte o tutte le volte in cui è necessario oppure atteso. In alternativa, non vi è continuità quando l'apprendimento si manifesta solo sporadicamente o mai (2020: 3).

The descriptors of the learning levels are thus elaborated on the basis of the combination of the aforementioned dimensions (2020: 5):

Tabella 1 – I livelli di apprendimento.

Avanzato: l'alunno porta a termine compiti in situazioni note e non note, mobilitando una varietà di risorse sia fornite dal docente sia reperite altrove, in modo autonomo e con continuità.
Intermedio: l'alunno porta a termine compiti in situazioni note in modo autonomo e continuo; risolve compiti in situazioni non note utilizzando le risorse fornite dal docente o reperite altrove, anche se in modo discontinuo e non del tutto autonomo.
Base: l'alunno porta a termine compiti solo in situazioni note e utilizzando le risorse fornite dal docente, sia in modo autonomo ma discontinuo, sia in modo non autonomo, ma con continuità.
In via di prima acquisizione: l'alunno porta a termine compiti solo in situazioni note e unicamente con il supporto del docente e di risorse fornite appositamente.

Figure 9 – Learning Levels (Guidelines, 2020: 5).

For those children who are not able to reach some of the learning objectives or for those who appear to be 'in via di prima acquisizione', the minister indicates specific recovery measures and insists on the importance of creating individualised learning pathways, with an eye to inclusion and the right to diversity (2020: 5).

1.3 Reflective Teaching

1.3.1 Reflective Practice

The construct of reflective practice provides a strong theoretical framework which accounts for the choice of using a teacher diary as an instrument for data collection in the present study.

Throughout the years, scholars have offered a wide range of perspectives on the meaning(s) attributed to the construct of reflective practice (also called reflection or reflective teaching). Beauchamp published a comprehensive study providing an overview of the definitions of the concept and – although she was not able to identify a specific interpretation – introduced a useful framework for reviewing the literature (Beauchamp 2006, quoted in Collin et al. 2013). By conducting an in-depth analysis of the body of research on reflective practice (55 studies), Beauchamp pinpointed various defining elements tied to reflective practice – such as processes, object and motives – and some related items, as shown in Figure 10.

<i>Processes of reflection</i>	<i>Objects of the reflective process</i>	<i>Rationales for reflection</i>
Examine	Practice	Think differently or more clearly
Think and understand	Social knowledge	Justify one's stance
Problem-solve	Experience	Think about actions or decisions
Analyse	Information	Change thinking or knowledge
Evaluate	Theories	Take or improve action
Construct, develop, transform	Meaning	Improve student learning
	Beliefs	Alter self or society
	Self	
	Issues of concern	

Figure 10 - Beauchamp (2006)'s theorization of types of processes, objects and motives tied to reflective practice (Collin et al (2013)).

Far from being a comprehensive or accepted definition, Farrell's (2012) metaphor of reflection as a compass is particularly helpful to grasp the process of active self-evaluation (Burhan-Horasanli & Ortaçtepe 2016); and according to Farrell (2012: 7), indeed,

reflective practice [is] a compass of sorts to guide teachers when they might be seeking direction [...] as to what they are doing in their classrooms. The metaphor of reflection as a compass enables teachers to stop, look, and discover where they are at that moment and then decide where they want to go (professionally) in the future.

Some definitions of reflective teaching include references to the solitary aspect of the process, such as that provided by Cruickshank and Applegate, who characterize reflective teaching as 'the teacher's thinking about what happens in classroom lessons, and thinking about alternative means of achieving goals or aims' (1981: 4, as cited in Bailey et al., 2001:

36). Another similar definition is provided by Richards and Lockhart, who define reflective teaching as a practice where 'teachers and student teachers collect data about teaching, examine their attitudes, beliefs, assumptions, and teaching practices, and use the information obtained as a basis for critical reflection about teaching' (1994:1, as cited in Bailey et al., 2001: 36).

The temporal dimension is an interesting lens through which to view reflective practice. It allows to build an interpretive timeframe which includes three categories: reflection-on-action, reflection-in-action and reflection-for-action (Wilson, 2008; Farrell, 2012).

Wilson (2008) is particularly concerned with the chronological dimensions of reflection. His temporal model of reflective practice includes past, present and future, as shown in Figure 11 (Wilson 2008: 183).

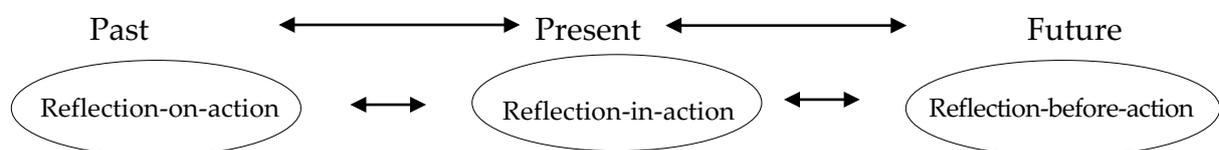


Figure 11 - A chronological perspective of reflective practice (reprinted from Wilson, 2008).

The first category, reflection-on-action, carries us back to the origins of reflective practice, that is to say to the seminal work of the American educator John Dewey on reflective inquiry (1933) and to the simple idea that 'reflection may be seen as an active and deliberative cognitive process, involving sequences of interconnected ideas which take account of underlying beliefs and knowledge' (Hatton & Smith 1994: 34).

Dewey defined 'reflective thought' as being an act of "active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which it tends" (1933: 6). He outlined what is, according to Farrell (2012: 10) 'the first real systematization of reflective inquiry into teaching', consisting of five non-sequential phases:

- suggestion, or the feeling of a problematic situation
- intellectualization, or the definition of the problem to be solved

- guiding idea, or the definition of working hypothesis for an initial observation or gathering of data
- reasoning, or the mental elaboration of the supposition
- hypothesis testing, or the exposition of a more refined idea and the subsequent testing of such idea in action (real or fictional)

This suggests that Dewey was encouraging educators to implement a form of evidence-based teaching (reflection-in-action) by incorporating methodical and analytical considerations into their teaching experiences so as to reach a deeper form of understanding and awareness that would lead to professional growth (Farrell, 2012).

Regarding the second category, reflection-in-action, an influential scholar was Donald Schön (1983), who focused on practitioner-generated intuitive practice and theorized the idea of 'reflection-in-action', that is to say the ability of professionals to think on their feet (Farrell 2012: 12) whenever new and problematic situations arise.

A key concept in Schön's position that is tightly related to reflection-in-action is that of "knowing-in-action", or the professional knowledge that teachers develop within action, that is to say the way they learn to teach by teaching (Munby, 1989). Schön (1983: 49-50) describes 'knowing-in-action' this way:

"[...] the workaday life of the professional depends upon tacit knowing-in-action. Every competent practitioner can recognise phenomena [...] for which he cannot give a reasonably accurate or complete description. In his day-to-day practice he makes innumerable judgements of quality for which he cannot state the rules and procedures. Even when he makes conscious use of research-based theories and techniques, he is dependent on tacit recognitions."

Altrichter et al. (2005) describe 'tacit-knowing-in-action' as a kind of professional action that 'flows smoothly and appears simple to an onlooker' (2005: 201) and is characterized by the following aspects:

- unity of action and thought
- unawareness of the origin of one's functional knowledge

- inability to verbalize one's functional knowledge

An example of 'tacit-knowing-in-action' – according to the authors – are routines, defined as "actions or mind-sets which have been built up through frequent repetition, [...] carried out comparatively quickly and [...] executed largely unconsciously" (Altrichter et al. 2005: 202).

Arguments in favour of routines come from studies on expert knowledge (Bromme 1985, quoted in Altrichter et al. 2005), where it has emerged that routines indicate the special ability of expert practitioners to synthetically organize their professional knowledge in order to solve issues in an effective and skilful manner.

According to Altrichter et al. (2005) reflection-in-action arises when professionals encounter obstacles to this usual flow of action (routine) and they are able to benefit from this type of reflective practice even without an explicit verbalization or formulation (2005: 203).

These positions apply to teachers as professional practitioners. In their everyday life teachers experience an abundance of swift exchanges and unforeseen interactions within the classroom, which often require an instantaneous intervention; furthermore, they generally tend to carry out an on-air evaluation of procedures which relies more on 'feeling, rather than a conscious application of principles' (Wallace 1991: 13). As reported by Wallace (1991), reflection may occur after a particularly good/bad professional performance or during the professional performance in order to foster a process of self-improvement for future action. For example, in order to carry out a reflection-on-action, teachers may think about an unplanned event that occurred in the classroom (a disciplinary issue, an unpredicted student error, an unpredicted comprehension difficulty, an unforeseen change in the lesson plan...) and answer to some questions to guide their reflecting process:

1. What was the problem or development, exactly?
2. How did you handle it?
3. Why did you handle it the way you did?
4. Would you handle it in the same way again? If not, why not?
5. Has the incident changed your general view of how to go about the practice of teaching? (Wallace 1991: 14).

The idea of knowing-in-action, together with the idea of reflection in the teaching practice is also useful in the area of professional education in teacher education programmes. Wallace (1991)'s reflective practice model for professional education development explains this relationship:

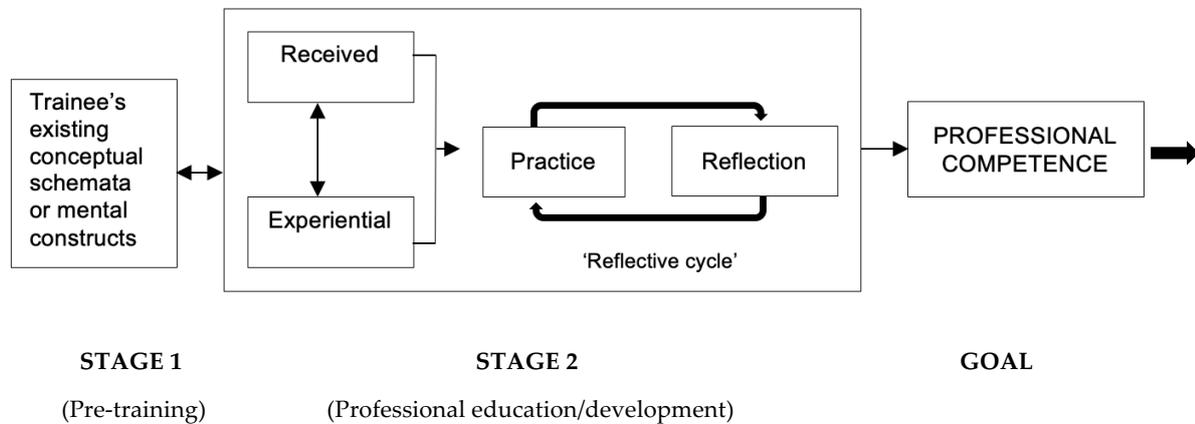


Figure 12 - Reflective practice model for professional education/development (Wallace, 1991).

At the base of Wallace's model is the concept that teachers possess what he calls 'received knowledge', consisting of theories, concepts and other forms of instruction that are part of the curriculum, and another type of knowledge, labelled 'experiential', arising from the educator's own direct experience of teaching in the field (Wallace, 1991). The key issues for practitioners are the way in which these two components of their know-how are associated to the interplay between them during the teaching action: while teachers often find it difficult to integrate the received knowledge into their lessons, they also resort to their experiential knowledge, often unconsciously (Bailey et al., 2001).

A challenging interpretation of Schön's stance, however, has been provided by Roth, Lawless and Masciotra (2001). In their paper "*Spielraum* and Teaching" the authors claim that the concept of reflection-in-action is only partially suitable to interpret the practice of teaching. Building upon the time constraints that affect teaching, and referring to Heidegger's concept of *Dasein*, they wield the idea of '*Spielraum*' ('room-to-maneuvre') to assert that when a 'surprising event' occurs, a time-out for immediate reflection would "lead to a loss of synchronicity with the unfolding conversation" (Roth et al. 2001: 26).

On the other hand, however, other scholars like for example Beck and Kosnik (2001) believe that reflection-in-action is a valuable and productive praxis and it enhances responsiveness and attentiveness in teaching:

there are many advantages to reflecting and making adjustments *while* we teach: there is less danger of forgetting what actually happened; we can be more realistic about the circumstances; we can get just the right shading in our solutions through immediate feedback; we can improve our teaching in *this* lesson rather than waiting until a future one; and the students can see us making on-the-spot adjustments and learn through modeling how to do the same in their practical endeavours (2001: 222).

Furthermore, regarding the 'surprising event', unlike suggested by Roth et al., these unpredictable variables trigger a state of 'heightened consciousness' in which the reflective activity takes place and this, according to the authors, is 'part of the normal, 'immediate' experience of practitioners and the major means whereby we enhance our practice' (Beck & Kosnik 2001: 224).

Finally, with reference to the third temporal category of reflection (reflection-for-action), scholars such as Farrell (2012), van Manen (1991) and Wilson (2008) offer some perspective into it.

Farrell (2012) maintains that this third category, that he calls reflection-for-action, emerges from the combination of the others. According to him, both reflection-on-action and reflection-in-action can promote a systematic and evidence-based investigation of problematic situations that has an impact on future endeavours.

A similar view is suggested by Wilson (2008), who is particularly concerned with the future dimension of reflection, what he calls 'reflection-on-the-future' or 'reflection-before-action'. In his opinion, thinking about possible future scenarios is a basic human instinct and 'one of the most powerful mental tools we have at our disposal' (Wilson 2008: 180). In addition, Wilson provides a clear definition of reflection-on-the-future as being "the act or process of reflecting on desirables and possible futures with the purpose of evaluating them as well as considering strategies intended to achieve the objective(s)" (Wilson 2008: 180).

Wilson (2008) argues that his view is also supported by other scholars such as van Manen, who observes that 'retrospective reflection on (past) experiences differs importantly from anticipatory reflection on (future) experiences' (van Manen 1991, quoted in van Manen 1995). Therefore, the concept of 'anticipatory reflection' seems to entail the teacher's ability to solve immediate problems while considering future problematic situations they may come across in the future thus showing a certain degree of anticipatory reasoning (van Manen, 1991).

More recently, Pollard lists seven key features of reflective teaching:

- Reflective teaching implies an active concern with aims and consequences, as well as means and technical efficiency.
- Reflective teaching is applied in a cyclical or spiralling process, in which teachers monitor, evaluate and revise their own practice continuously.
- Reflective teaching requires competence in methods of evidence-based classroom inquiry, to support the progressive development of higher standards of teaching.
- Reflective teaching requires attitudes of open-mindedness, responsibility and wholeheartedness.
- Reflective teaching is based on teacher judgement, informed by evidence-based enquiry and insights from other research.
- Reflective teaching, professional learning and personal fulfilment are enhanced through collaboration and dialogue with colleagues.
- Reflective teaching enables teachers to creatively mediate externally developed frameworks for teaching and learning (2002: 12-13).

The first feature relates to the role of teachers as translators and – more importantly – active contributors of educational policies from the society into the classroom (ibid, 2002). The second characteristic of reflective teaching involves the dynamic process of the reflective teaching action, that Pollard (2002, 16) summarizes in seven cyclical key stages: reflect, plan, make provision, act, collect data, analyse data and evaluate data. Regarding the third

feature, the crucial competences that teacher should possess are: reviewing relevant existing research, gathering new evidence (both objective and subjective data), analytical skills (data acquire meaning only when interpreted), evaluative skills and judgment (ibid, 2002). As for the attitudes mentioned in the fourth point, Pollard (2002) reports the same three attributes mentioned by Dewey (1933) and adds that in our contemporary society, such views are often questioned but for teachers “maintaining a constructive engagement, a willingness to imagine new futures, and a self-critical spirit” is of the utmost importance (ibid, 2002: 18). Regarding teacher judgment, Pollard (2002) refers back to the work of Schön (1983), reporting that the ‘swampy lowlands’ of ‘caring’ professions such as education implicate a preponderance of interpersonal and qualitative matters and entail a degree of judgment, but this judgment needs to be ‘informed’, that is to say it must stem from a profitable dialogue between teachers, researchers and policymakers (ibid, 2002). About collaboration with colleagues, Pollard (2002) refers back to Vygotsky (1986)’s socio-cultural theory by claiming that collective reflection is beneficial for the individual but also for the school community because it involves clearer interpretation of the goals, better analysis, reciprocal assessment and self-confidence in facing novelty. Finally, about the last feature, reflective teaching as creative mediation, it is intended as ‘the interpretation of external requirements in the light of a teacher’s understanding of a particular context and bearing in mind his or her values and educational principles’ (ibid, 2002: 22). There are different types of ‘creative mediation’: protective mediation (regarding existing principles that one wishes to maintain), innovative mediation (finding creative solutions within the new requirements), collaborative mediation (seeking support from other teachers), conspirational mediation (resistance to requirements that are deemed unsuitable) (ibid, 2002).

Zeichner and Liston (1996, quoted in Bailey et al. 2001: 40) examine five dimensions of reflection and the related behaviours, which are outlined in Figure 13.

1. RAPID REFLECTION	Immediate and automatic <i>Reflection-in-Action</i>
2. REPAIR	Thoughtful <i>Reflection-in-Action</i>
3. REVIEW	Less formal <i>Reflection-on-Action</i> at a particular point in time
4. RESEARCH	More systematic <i>Reflection-on-Action</i> over a period of time
5. RETHEORIZING AND REFORMULATING	Long-term <i>Reflection-on-Action</i> informed by public academic theories

Figure 13 – Zeichner and Liston (1996), *Dimensions of Reflection* (Bailey et al., 2001: 41).

As specified by Zeichner and Liston, the first dimension entails all the immediate decisions teachers made while they are teaching, and for their rapidity in execution, they may be perceived as ‘routine’. The second dimension, called ‘Repair’, refers to the adjustments to the schedule made by the teacher in response to students’ reactions. The other dimensions pertain to Reflection-on-Action, and range from an informal review system to a more complex and abstract process that ultimately lead to major changes in teachers’ own beliefs (1996, quoted in Bailey et al., 2001: 41-42).

Nagamine et al. (2018) focused on the role of emotions in reflective practices carried out in second language classrooms. The key construct in their research is that of L2 teachers’ ‘felt sense’, described as ‘a combination of emotion, awareness, intuitiveness, and embodiment’ (2018: 145). In their view, reflective teaching is an emotional activity aimed at exploring and critically reflect on one’s beliefs, and these are dynamic in nature, because they ‘can be transformed through different emotional experiences in and outside the classroom’ (ibid, 2018: 147). Affective factors might influence teachers’ cognition processes so that in some cases their stated beliefs do not correspond to their actual practices in the classroom (ibid., 2018).

One of the benefits of adopting a reflective teaching approach is that of initiating a journey of professional development for teachers. The concept of professional development has been explained by Lange (1990: 250, quoted in Bailey et al., 2001: 15) as 'a term used in the literature to describe a process of continual intellectual, experiential, and attitudinal growth of teachers', who claims that 'teachers continue to evolve in the use, adaptation, and application of their art and craft' (ibid, 2001).

1.3.2 Reflective Teacher Diary

In what has been considered the first comprehensive methodological guide to the use of diaries in social research (Bartlett and Milligan, 2015), Alaszewski provides a general definition of diaries and underscores some key defining features of this method:

A diary can be defined as a document created by an individual who has maintained a regular, personal and contemporaneous record (2006: 1).

As for the diary's application in the research of topics related to language teaching/learning, according to the definition provided by Bailey and Ochsner (1983, quoted in Bailey, 1991) in a pivotal study,

a diary study in second language learning, acquisition, or teaching is an account of a second language experience as recorded in a first-person journal. The diarist may be a language teacher or a language learner – but the central characteristic of the diary studies is that they are introspective: The diarist investigates his own teaching or learning. Thus he can report on affective factors, language learning strategies, and his own perceptions – facets of the language learning experience which are normally hidden or largely inaccessible to an external observer (1983: 189).

According to Bailey and Ochsner (1983, quoted in Bailey et al. 2001: 50), the process of a diary study consists of five subsequent stages, depicted in Figure 14:

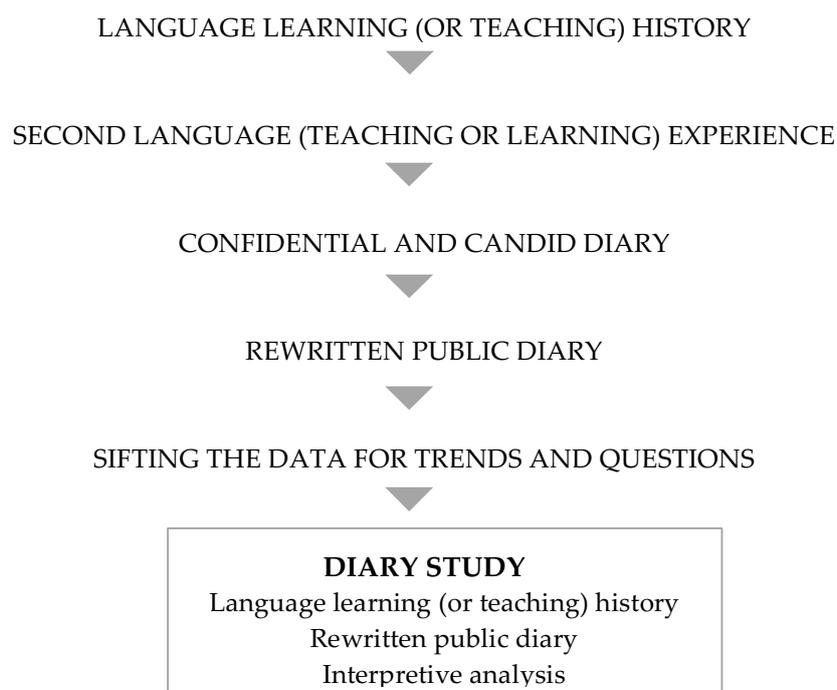


Figure 14 – Conducting a Language Learning or Teacher Diary Study (Bailey et al. 2001: 50)

As it emerges from the scheme, before systematically recording the teaching/learning experiences, the diarist should first present his/her own personal history of language teaching or learning, and after concluding the diary, it should be edited for publishing and – most importantly – it should be analysed in order to find recurrent themes or critical issues. Finally, the researcher or the diarist discuss the relevant factors pertaining the language teaching/learning experience and interpret the results in the final study in light of the current literature (Bailey et al., 2001).

McDonough and McDonough (1997: 121) found that diaries are “a pervasive narrative form” as they provide a tool to structure, formulate and react to the immediate experience which is recorded for a later reflection and inquiry. The authors highlight four key features of the ‘diary-generated text’ (1997: 124), which have a strong influence on the way this kind of data can be analysed:

- richness in terms of quality and quantity: different themes and perspectives are present concurrently;

- mediated or free introspection: the text can be strictly personal or shared with others (the researcher, for example);
- retrospective: the text might be subject to accuracy issues, although it is widely acknowledged that accuracy is not the main goal of this method;
- wide scope: even a small portion of text has the potential to open many possible avenues for analysis;

The use of the term 'journal' in literature refers primarily to the diary itself as a *product* of the study, but most researchers use the terms 'diary' and 'journal' or 'diary studies' and 'journal studies' interchangeably (Bailey & Curtis 2009; Numrich 1996). However, Wallace (1998) uses the term 'journal' to refer to a document that is very similar to a diary but is less candid, as it is intended for public reading and must therefore be "edited in the process of composition, like any other document that one knows will be read by others" (1998: 62).

Bailey and Curtis (2009: 69-70) provide a classification of diary studies according to the agent(s) of analysis, that is the person who carries out the analysis of the data collected in the journal. They identify two categories: primary (or direct, or introspective) analysis and secondary (or indirect, or non-introspective) analysis. In the first case, the person who wrote the journal's entries performs the subsequent content analysis. Conversely, in the second case, the analysis is performed by an external researcher that shall not coincide with the author (Bailey & Curtis, 2009: 69).

As Numrich (1996) points out, however, it is possible for the author (teacher) to conduct a primary analysis and then the study can undergo a secondary analysis – thus being incorporated with other similar studies – to compare the findings and find patterns, similarities and common categories. Such an analysis may serve as direction for teacher educators and managers of university programmes on how to interpret and understand the needs of new teachers (ibid, 1996).

Another possibility to categorize diary studies is to consider the recording mode, or the moment in which the author chooses to write on the journal. Bolger et al. (2003) suggest a two-fold level for the classification of diary studies: time-based studies and event-based studies. Diary studies with a time-based design are often focused on personal processes or

ongoing experiences, as they provide a systematic view of the selected variables. In this case, the researcher should determine the appropriate interval (fixed or variable schedule) and both can be subject to the use of a determined signal to prompt the recording of an entry (Bolger et al. 2003: 588). As for the event-based design, it requires the subject to provide an entry on the journal every time that a ‘triggering event’ occurs (ibid. 2003, 590).

Regarding the collection medium, diary studies’ design may involve the traditional pen and paper method – where the author has to fill in a questionnaire or provide a brief description or verbal narrative –, an electronic data collection method or even a combination of both, depending on the research design or on the diarist’s preferences (Dörnyei 2007; Bolger 2003). Bailey (1990: 215) provides a definition for the diary study as “a first person account of a language learning or teaching experience, documented through regular, candid entries in personal diaries and then analysed for recurring patterns or salient events”. According to Bailey (1990), such tools have been adopted to research three main areas: a) experiences of language learning, b) reactions of in-training teachers about their academic courses, and c) experiences of language teaching.

McDonough and McDonough (1997) identify different types of diaries, as synthetized in Figure 15:

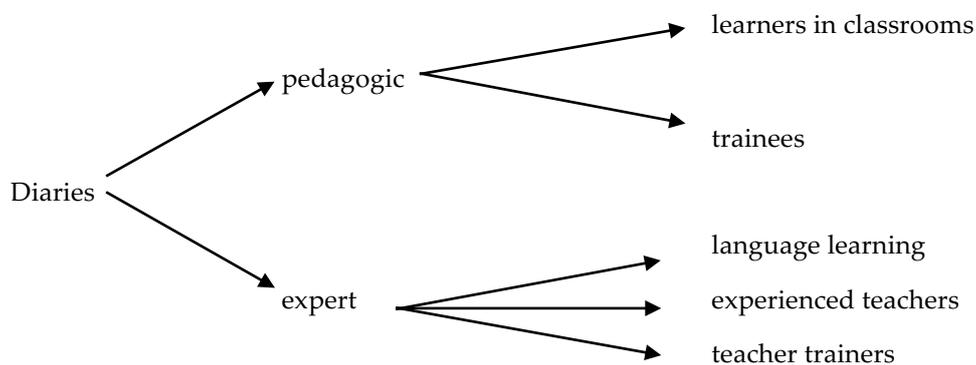


Figure 15 – Types of diary writing in language learning and teaching (McDonough & McDonough 1997, 133).

Concerning Teacher diaries, some possible general areas of exploration might be the teaching processes, student behaviours, collegial interactions and metacomment on keeping a diary (McDonough & McDonough, 1997). As stated by Richards and Lockhart, a teacher

journal 'is a teacher's or a student teacher's written response to teaching events', and another useful tool to help reflecting is a lesson report, that is to say 'a structured inventory or list which enables teachers to describe their recollections of the main features of a lesson' (2007: 9), different from a lesson plan because it describes how the lesson went with reference to the plan. The term 'teacher diary', is often used in connection with the reflective practice in teacher education or in the professional development of teachers. The study carried out by Jeffrey (2004) on his experience with keeping a teacher diary is a case in point. In the context of teaching English in Japan, Jeffrey chose this specific introspective method in order to undergo a process of professional development aimed at transforming his thinking about his attitudes towards his teaching and his students (ibid, 2004). Before diving into the main study, Jeffrey chose to undertake a 3-day pilot study, so as to test the feasibility of the study and tackle any possible issues with the procedure. To help him in the writing of the diary entries, he used his lesson plans and handouts, and he used the Wordsmith Tools 3 Program to analyse the entries, a tool he has defined 'essential' for the analysis part of his study (ibid, 2004). He claims that keeping a diary 'gave [him] insights into [his] teaching, from the brighter and the darker sides, as well as from specific and wider perspectives' and if he had not done the diary study most of those would have remained unexplored (2004: 16).

Possible limitations

As shown in the previous section, engaging in reflective practice oriented (RP-oriented) journaling offers important advantages, however, some concerns emerge from the literature about the use of diary methods.

According to Cao and Henderson (2021), engaging in a diary study may entail substantial challenges for researchers and participants alike. With regard to researchers, the effects of diary keeping on the participants' mental health or behaviours may become an unwanted side-effect if it has an unforeseen impact on the variables involved in the study (Cao & Henderson 2021, v). Another concern regards the quality and the quantity of collected data, as the researcher has limited control on the entries written by the participant and therefore cannot adjust the focus if the participant steers away from the requested task (ibid, 2021).

Ultimately, the ethical issues associated with the nature of the tool itself involve matters of anonymity and confidentiality as well as reliability (ibid, 2021).

Challenges for participants include for example commitment issues (as diaries are more time-consuming and demanding than other tools), fluctuation of quality and quantity of entries and the specificity of requested skills (literacy, information selection, critical thinking, IT skills for digital diaries...) (Cao & Henderson 2021, iv). The time commitment constraint is mentioned by Dörnyei (2007) and Bolger et al. (2003), who concur that the amount of dedication required by this instrument is particularly elevated. Rose (2020) suggests structuring the entries with guiding questions so as to target the relevant constructs and collect only relevant data that fit the purpose of the study and simplify the process of analysis.

As for the literacy constraint, if the participants are required to use the L2, there might be some issues with low and high proficiency users alike, as the diary methods are employed specifically to capture the peculiarities and the specifics of one's experience (Rose, 2020). An option to avoid this issue is to allow participants to use their first language, but in this case the researcher is required an extra effort for working on multiple languages during the analysis phase (Rose 2020). Another option suggested by Rose (2020) is to allow for multimodality, that is to say giving the participants the option to integrate their entries with drawings, audio recordings, pictures and so on.

Another major issue with diary methods is that of memory biases and unconscious editing (Fry, 1988; Nezlek 2012; Rose, 2020). If the recollection happens at a much later stage than the event itself, this becomes an "additional load on long-term memory" and the ability to accurately and truthfully report facts is compromised (Fry 1988: 160).

Nezlek (2012) explains the issue of timing in record-keeping very accurately:

when we remember, we create memories as such, and perhaps more, than we recall memories. In terms of providing descriptions of what has occurred, the longer the time between the event and the description, the more extraneous factors (factors not related to the event itself) can influence the description (2012: 4).

Furthermore, Nezlek (2012) mentions five factors that can affect the recollection of an event if too much time passes between the episode and the moment of writing it in the diary:

- recency: especially for same-type experience (e.g. interactions with the same person), recent events are recalled better, thus reducing the risk of alteration by intervening events that may have occurred in the meanwhile;
- salience: significant events are recalled better than ordinary ones and negative events tend to adversely influence the description of ordinary experiences;
- sense-making: personal implicit beliefs, theories and stereotypes intervene in the recollection of events and modify the distinctiveness of one's experiences;
- present state of mind: mood, emotions and attitudes may influence the recollection of positive and negative experiences or frame events in a pre-conditioned way;
- making distinctions: in immediate reports of experiences, different levels or aspects of an experience are separated and accounted for, while in single retrospective reports the hedonic dimensions (good-bad) tend to conglomerate (Nezlek 2012: 4-6).

Some other problems regarding diary/journal keeping by teachers emerge from the literature. For instance, according to Wallace (1998), two factors have a negative impact on the process: time constraints and mental fatigue. Firstly, teachers have such a tight schedule that finding the time to write extensively is a challenge and secondly, mentally revisiting the 'traumas of the teaching battlefield' is psychologically exhaustive (Wallace 1998: 64).

In short, Bailey notes that the doubts about diary studies emerging from the literature can be subject-related, data-collection related or data-analysis related, but most of these concerns regard issues of external validity and generalizability of results (1991). However, Bailey draws attention to the fact that 'achieving generalizability is neither the purpose nor the point of the diary study' (1991: 83). Rather, the point of the diary study is to explore a language learning/teaching phenomena from the point of view of the agents involved in the

process, and from this perspective, diaries can be a revealing source of knowledge (Bailey, 1991).

1.4 Conclusion

This chapter has offered an overview of theories and dynamics pertaining to reflective teaching and foreign language teacher-based assessment of young language learners.

The first part of the chapter focussed on the topic of assessment and featured a discussion about educational assessment theories and factors involved in foreign language assessment. Moreover, the role of teachers in assessment was analysed, by referring to constructs such as diagnostic competence, assessment literacy, and emotional challenges connected to assessment. Another key aspect of assessment is the context in which it takes place, therefore attention was given to the classroom as the 'theatre' where learning happens and the different actors come together to create meaningful assessment opportunities to sustain learning. A brief outline about evaluation in the Italian educational system was provided in order to introduce the newest reform for the evaluation of pupils in primary schools, whose implementation was directly experienced by the author, and formed the context for the present research.

In the second part of this chapter, space was given to exploring the dynamics and the factors involved in the reflective teaching approach to assessment. The concept of reflective practice was investigated, with particular attention to the use of introspective methods – such as the teacher diary – to enhance professional development and assist the teacher in implementing a learning-oriented assessment. In particular, both the characteristics of teacher diaries and the advantages of the diary method were explored. Finally, some concerns related to diary methods were addressed.

2. RESEARCH PROJECT

2.1 Introduction

The current chapter is aimed at outlining the methodology of the present research project. First, a brief contextual introduction will be provided, together with the definition of the questions that guide this research (paragraph 2.2) Then, the paradigm and method of the research will be discussed and the research design will be presented, so as to explain choice of the data collection tool and its fundamental role in the research project (paragraph 2.3). More specifically, it will be ascertained that the present study belongs to the Interpretive paradigm and it can be defined as an action research project in the wider field of classroom research (paragraph 2.3.1). In paragraph 2.3.2, some information will be given with regard to the research participants, in the form of a brief language teaching biography. It will also be highlighted that the current research employs a qualitative introspective method through the use of diary as a data collection tool (paragraph 2.3.3). Thus, the chosen method of data analysis, namely the qualitative content analysis, will be explained and discussed in detail (paragraph 2.3.4).

2.2 Context and Research Questions

The current research project focuses on the author's teaching experience at a public primary school in the school year 2020-2021. In order to gain a better understanding of the aim of the study and the methodologies adopted in order to pursue the objectives, an overall description of the instructional context will be provided.

The author (henceforth teacher-researcher) took up service under a fixed-term contract at a primary school located in the mainland area of Venice, Italy, in October 2020, in order to substitute an absent tenured teacher who taught English to a 3rd grade class and Italian, English, History, Geography, Arts and Music to a 5th grade class. The 5th grade class was composed of 20 pupils, 13 females and 7 males between the ages of 10 and 11 years old. The 3rd grade class was composed of 18 pupils, 9 females and 9 males between the ages of 8 and

9 years old. The school is located in a medium-sized municipality within the metropolitan city of Venice, populated by roughly forty-two thousand residents. The school is part of a larger school (which comprises several primary schools as well as one middle school) and the primary school in question is situated in a suburban area which is fairly off-centred in relation to the city centre. The school is housed in a large building and hosts twelve classrooms with a total of circa two-hundred pupils and thirty practitioners (teachers, special education teachers and educators).

The teacher-researcher began teaching the aforementioned classes in the second part of the first period (December 2020), and spent the first month assessing the pupils' learning needs by administering diagnostic tests and by informally assessing their proficiency in English. At the same time, it was necessary to establish an affective bond with the pupils, who manifested feelings of distrust and disorientation due to the fact that the previous teacher left the job unexpectedly and they had to adapt to yet another alteration of their 'classroom habitat'. In order to create an atmosphere of trust and well-being, during the first few weeks the teacher-researcher provided the pupils with ice-breaking activities and playful tasks designed to introduce herself to the class while, at the same time, getting to know the pupils, their interests and attitudes. Once a climate of confidence was established, the teaching activities began focusing on the programs of each subject and on the various curricular activities and projects organised by the institution.

The two questions guiding this study are:

- How can reflective teacher diaries be used for the assessment of EFL primary pupils? (RQ1)
- What are the advantages and disadvantages of reflective teacher diaries for the professional development of novice EFL teachers? (RQ2)

The first question aims at exploring the use of the teacher diary in order to implement a reflective approach to the assessment of young EFL learners in the research context. This

question will be addressed by interpreting the introspective data through a qualitative content analysis (QCA) methodology, that will be examined in detail in paragraph 2.3.4.

The second research question focuses on the aspect of professional development in relation to the reflective approach adopted for this study. The aim is to investigate the advantages and disadvantages of journaling as a tool for reflecting teaching, in relation to how this praxis might encourage novice teachers through their professional development. An answer to this question will be provided by both the qualitative content analysis and the personal experience of the teacher-researcher.

2.3 Method

2.3.1 Research Methodology

The choice of an epistemological paradigm is an important step in every qualitative research project, as this choice has implication in every stage of the research, and most scholars¹⁴ agree that being aware of the philosophical frameworks is of pivotal importance for every researcher (Mertens, 2014). In other words, in order to undertake a solid and replicable investigation of a given social phenomenon *'it is necessary to conduct research with eyes philosophically open'* (Blaikie and Priest 2017: 3, emphasis in original). Given that researchers work from their unique standpoint, 'good research requires making [their] assumptions, paradigms, and frameworks explicit in the writing of a study, and, at a minimum, to be aware that they influence the conduct of inquiry' (Creswell & Poth, 2016: 15).

In fact, paradigms are 'a way of looking at the world' and are 'composed of certain philosophical assumptions that guide and direct thinking and action' (Mertens 2014, 7). The present research belongs to the paradigm of Constructivism, often combined with or described as Interpretivism (Creswell & Poth, 2016; Mertens, 2014). This paradigm posits that 'knowledge is socially constructed by people active in the research process' and 'research is a product of the values of researchers and cannot be independent of them' (Mertens 2014, 16). Furthermore, this paradigm stresses the importance of contextualising

¹⁴ Mertens (2014, 6-7) reports some controversial views on this matter; in particular, Patton (2002) claims that referring to paradigms is unnecessary and somewhat damaging to the research process.

the analytical process by filtering the results according to the researcher's interpretation, hence the term 'interpretive research' (Creswell & Poth, 2016).

The approach adopted in the current research is that of an action research, carried out in the wider contextual field of classroom-based research. In the field of language education, several definitions of action research have been proposed. Cohen et al. (2000: 297), for example, argue that action research 'is a powerful tool for change and improvement at the local level' that can be adopted by a single teacher, a group of teachers or by a teacher/teachers working with one or more researchers in order to investigate many different areas, such as 'evaluative procedures' in order to 'improving one's methods of continuous assessment (Cohen et al., 2000: 297).

Burns (2005: 57) writes that action research 'is seen as a means towards creating meaning and understanding in problematic social situations and improving the quality of human interactions and practices within those situations'. In other words, as opposed to quantitative, experimental approaches which aim at generalizing, in action research teachers-researchers are interested in finding tailor-made, practical solutions to what Allwright and Bailey (1991) refer to 'puzzles', instead of using the term 'problem' which, according to their view, may imply a negative meaning.

According to Kemmis and McTaggart (1988, quoted in Nunan, 1992), three key features of action research are (a) that it is conducted by practitioners, (b) that it is collaborative and (c) that it has a revolutionary intent inside its context. Such characteristics are shared by Cohen and Manion (1994, quoted in Cohen et al., 2000: 297), who define action research as 'a small scale intervention in the functioning of the real world and a close examination of the effects of such an intervention'. In relation to these defining features, Nunan (1992) argues that collaboration and the strive for change are not crucial components of action research projects, as teachers may choose or have to work independently and descriptive studies are equally relevant if carried out properly.

The definition that has been adopted as a reference for the present research project, however, is the one provided by Wallace, who rooted it in professional development and wrote that action research is a process based on 'systematically collecting data on your

everyday practice and analysing it in order to come to some decisions about what your future practice should be' (1998: 4). This definition is consistent with Wallace's (1991) model for reflective teaching and professional development, already discussed in paragraph 1.3.1. This approach fitted the research objectives best, as it is intended to implement active and purposeful changes in the teacher-researcher's practice; in fact, these changes need to be rooted in an inquisitive framework that relies on qualitative data gained from everyday classroom practice.

2.3.2 Participants and Language Teaching History

The subject in this diary study is the teacher-researcher (also referred to as 'diarist'), as the data collected refer to her teaching experience in the aforementioned educational context. The diarist has Italian as her first language, and she acquired English, German and French as foreign languages in formal instruction contexts (primary, middle, high school and university courses). The diarist is a novice teacher, this being her first teaching experience in a formal education institution. She had some previous experience as a private foreign language tutor, for pupils ranging from 7 years old to 14 years old. As a private FL tutor, she taught English, German and French to pupils of various ages. As for her academic path, the diarist holds a BA degree in Languages, Civilization and Sciences of Languages and is currently seeking an MA degree in Linguistics, specialising in language teaching.

2.3.3 Instruments and Procedures

In the current research project, the teacher-researcher is the participant and the action research consists of gaining useful insights into pupils' learning progress in order to perform an ongoing formative assessment of their English Language learning progresses, as required by the new legislation on evaluation concerning Italian Primary Schools. In order to achieve this objective, it was chosen to adopt an introspective method and therefore the teacher diary was selected as an introspective tool for the data collection phase of the study. Introspection is defined by Nunan (1992: 115) as 'the process of observing and reflecting on one's thoughts, feelings, motives, reasoning processes, and mental states with

a view to determine the ways in which these processes and states determine our behaviour'; he also uses the term introspection techniques to refer to techniques adopted in order to investigate such aspects of our cognitive processes. As noted by Nunan (1992: 116), however, such techniques may also be labelled 'retrospective', because 'there will always be a gap, however slight, between the event and the report'.

The teacher-researcher chose to adopt an unstructured approach to diary design, as the key interest for this research project was to explore the diarist's 'unspoken and often personal actions or experiences' (Bartlett & Milligan, 2015: 32). Therefore, the teacher-researcher chose a thematic area linked to the research aims (classroom-based assessment) and decided to record her experiences, activities and feelings related to that specific area, with the freedom to include any other aspect of her practice that she felt were important in relation to the study overall goals. An advantage of this free format diary technique is that the data can be recoded and analysed in many different manners, and since this study includes only one diary, the relatively small sample allows for a deeper analysis without putting too much strain on the researcher's time and resources (Bartlett & Milligan, 2015).

Following the advice provided by the diary studies completed by Sá (2002) and Jeffrey (2004), the teacher-researcher chose to implement a three-day pilot study before starting with the main investigation. The pilot study was conducted from Friday February 19th ending on Wednesday February 24th 2021. The purpose of the pilot study, as mentioned in Jeffrey's (2004) own diary study, was to ensure the feasibility of the main study, especially regarding any issues or difficulties that might arise during the procedure in terms of participant effort, time constraints or memory biases (see paragraph 1.3.2).

The diary entries were recorded by following an event-contingent design, with every entry being written after each English class (both 5th grade and 3rd grade). Each English class lasted approximately one hour, but in some cases two classes were consecutive, therefore the lesson lasted two hours. The pilot study produced a total of four diary entries and involved the following classes:

- Friday, 19th February 2021: 5th grade English class (1hr) + 3rd grade English class (2hrs)
- Monday, 22nd February 2021: 3rd grade English class (1hr)

- Wednesday, 24th February 2021: 5th grade English class (2hrs)

The entries consist of a verbal account of varying length and are written in the native language of the teacher-researcher (Italian). The chosen medium is a blend of pen-and-paper and electronic collection. The narrative system adopted in the present diary study is based on Erikson's (1986, quoted in Sá, 2002: 152) guidelines, namely:

- a) the recording of the diary entries follows an unstructured approach, meaning that no pre-defined categories have been identified;
- b) data stems from the teacher-researcher's selection and sampling;
- c) entries are written in a simple and colloquial language;

Following Bailey's (1991) suggestion, in order to obtain truthful and valuable data, the diarist's privacy needs to be taken into careful consideration. Therefore, prior to undertaking the pilot study, the teacher-researcher made the decision to consider the diary as private as possible, and therefore adopted a very personal and confidential style and tone in recording her feelings and thoughts. As a consequence, the excerpts that are published in this study are a polished version of the original diary, which has been redacted in order to protect the teacher's and the pupils' privacy.

The pilot study was carried out easily, but some issues emerged during this phase. In particular, on the first day, during the two classes (5th and 3rd grade), no notes were taken during the teaching phase and the entries were written after the work day was over, before dinnertime. The teacher-researcher found the recollection to be faulty and imprecise, because many of the 'mental notes' that she made during the classes were gone by the time she could write the entries down. In order to solve these issues, the following day (Monday, one entry, 3rd grade) the teacher-researcher took some rough notes during class, even though this procedure was found to be quite challenging, because of the quick-pace of the classes. On the final day of the pilot study, the teacher-researcher implemented yet another change in the procedure. Right after the end of the lesson, she used the transition time in between two classes to write a rough entry (pencil-and-paper) featuring the most salient episodes that happened during the lesson, with the help of the notes taken during teaching.

This procedure was also adopted in the diary study carried out by Sá, who registered ‘strong inference critical incidents in class [...] as memory aids to be expanded later in diary form’ (2002: 153). The final entry was written in a password-protected Word document right after the school day was over, with the help of the lesson plans and the notes taken during and right after the classes.

Due to an outbreak of the covid-19 pandemic that involved the school and resulted in months of distance learning, the main study was delayed. It was conducted from Wednesday April 7th ending on Monday May 3rd 2021, and features a total of 16 diary entries. The study involved the following classes:

- Wednesday, 7th April: 5th grade English class (2hrs)
- Friday, 9th April: 5th grade English class (1hr) + 3rd grade English class (2hrs)
- Monday, 12th April: 3rd grade English class (1hr)
- Wednesday, 14th April: 5th grade English class (2hrs)
- Friday, 16th April: 5th grade English class (1hr) + 3rd grade English class (2hrs)
- Monday, 19th April: 3rd grade English class (1hr)
- Wednesday, 21st April:
- Friday, 23rd April: 5th grade English class (1hr) + 3rd grade English class (2hrs)
- Monday, 26th April: 3rd grade English class (1hr)
- Wednesday, 28th April:
- Friday, 30th April: 5th grade English class (1hr) + 3rd grade English class (2hrs)
- Monday, 3rd May: 3rd grade English class (1hr)

In diary design, the duration of data collection is a key methodological factor that needs to be carefully considered, and it depends on the type of variables that the researcher wishes to examine and on the overall objectives of the study (Bartlett & Milligan, 2015). For the current project, in order to minimize issues of completion such as respondent fatigue, and in order for it to be a useful instrument for the teaching action, the teacher-researcher chose to complete the diary over the fairly short timeframe of one month.

The procedure for the collection of the data during the main study followed the structure implemented during the pilot study, with all the necessary modifications. Notes were taken during all the above mentioned classes, by taking advantage of pupils' independent work time during classes, like for example group work, pair work or individual completion of tasks. Then, right after the classes were over, a quick outline of the entry was written down on a notebook. These notes, together with the lesson plans were used to write the final entries, which were recorded digitally on a password-protected Word document on the teacher-researcher's personal computer.

The research design used for the main diary study was adapted from Bailey et al. (2001) and involves the following steps:

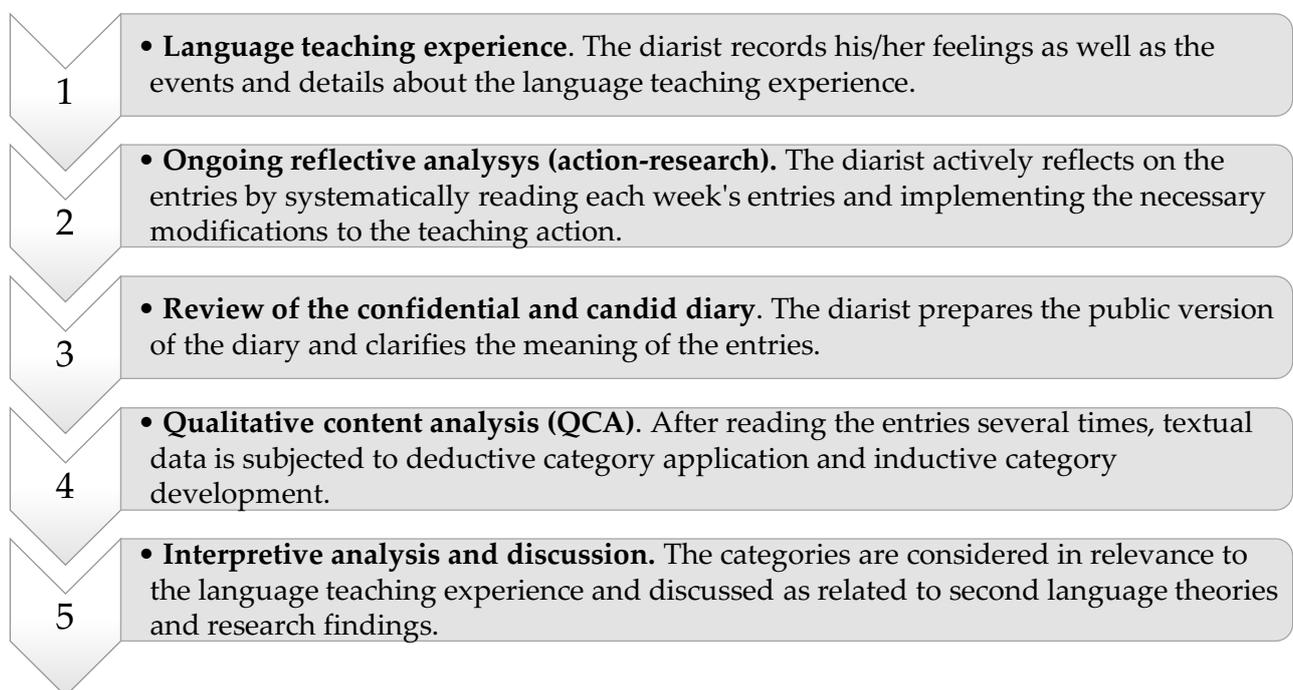


Figure 17 – Research Design.

2.3.4 Method of data analysis

During the second phase of this study, diary entries were re-read on a weekly basis, in order to reflect on the events that shaped teaching, learning and assessing, and in order to adjust the practice whenever needed. This 'contextual' and raw analysis of the data represents an important step within the reflective cycle, as it is by reasoning and reflecting on the

'experiential knowledge' (Wallace, 1991, see paragraph 1.3.1) that teachers can develop and grow as practitioners. This first analysis did not include any formal method, as the diary was used merely as a tool for the implementation of a reflective methodology. This exploratory phase helped with continuous assessment of pupils' skills and progressive development towards the learning goals, and guided the teacher-researcher in the whole process. Finally, during the fourth phase of the research project, the whole diary will be analysed by using an interpretive approach.

The data that was collected in the diary entries of the main study will be analysed by implementing a qualitative content analysis (QCA) procedure. This 'systematic, rule-bound procedure' (Mayring, 2015: 369) involves a deep examination of linguistic data with the aim to classify extensive chunks of text into more manageable categories and implies a subjective interpretation of contents (Hsieh & Shannon, 2005). This method of data analysis was developed by Philipp Mayring (1983, quoted in Glaser-Zikuda et al., 2020) and consists of:

- embedding data in the communication context
- adopting a rule-guided and systematic procedure of analysis
- developing or applying categories
- quantifying the categories if theoretically reasonable, and
- adopting some quality criteria (e.g., inter-rater reliability) (Glaser-Zikuda, 2020: 4).

Deductive category application and inductive category development are two strategies of analysis that can be applied at different stages of the qualitative content analysis procedure (Glaser-Zikuda, 2020). The analysis of the textual data occurred during the fourth phase of the current research project (see Fig. 17) and involved the 'uncensored' version of the diary, in order to maintain a close grip with the information provided in the data.

Within this phase, two distinct steps of data analysis can be identified. The first part of data evaluation will be carried out by performing a 'directed content analysis', a procedure that consists of creating initial coding categories based on existing theories or prior research and its goal is to 'validate or extend conceptually a theoretical framework or theory (Hsieh & Shannon, 2005). This process is what has been defined by Mayring as 'deductive category assignment' or 'structuring' (2015: 376). First of all, a deductive coding framework (or a

'category system') is identified and elaborated on the basis of the theoretical background supporting the research project, and subsequently the raw data is analysed based on some specific coding rules. This process aims at identifying smaller meaning units and labelling them with a code, that will be chosen according to the coding framework previously established. According to Bengtsson, a meaning unit is 'the smallest unit that contains some of the insights the researcher needs, and it is the constellation of sentences or paragraphs containing aspects related to each other, answering the question set out in the aim' (2016: 11).

As suggested by Mayring (2000), in order to create a transparent coding framework, it is necessary to follow a specific structure; in this study, for each category, an appropriate definition, an 'anchor sample' and a coding rule have been established in order to provide a rigorous and straightforward procedure. According to Mayring, definitions help to determine 'which text components belong in a given category', anchor samples are 'concrete passages [...] cited as typical examples to illustrate the character of those categories' and coding rules are established 'for the purpose of unambiguous assignment to a particular category' in case of delineation problems amongst the categories (2015: 377).

For the present research project, some practical arrangements suggested by Mayring (2015) will be applied and adapted to the research design. First of all, the diary will be edited and printed out in an analysis-friendly layout: a high readability font will be chosen, and a wide margin will be created in the right side of the page, in order to accommodate any annotations or remarks. For the procedure of deductive category assignments, a colour will be assigned to each category, and then all the fitting text passages will be underlined by using pencils of different colours. Then, the sub-categories will be numbered and marked by writing numbers in the margin of the text.

The second step of data evaluation will consist of analysing the parts of source materials not included in the previous analysis in order to create inductive categories that might help interpret additional or innovative aspects not examined in the theory (Glaser-Zikuda et al., 2020). This process has been defined by Mayring (2000) as 'inductive category development', and is a reductive process that analyses the text on the basis of a criterion of

definition derived from the theory, which basically guides the researcher in the selection of the aspects to take into consideration within the text. The gradual reduction of the source material into key information will result in the formation of categories, that will be progressively revised and reduced. On the practical side, the process will be similar to that described above. Through the use of highlighters of different colours, the material will be coded line-by-line. Preliminary codes or notes will be marked next to the meaning units. Halfway through the material, the categories will be revised, adapted and defined; ultimately the whole source material will be analysed and coded.

Finally, additional information will be provided by performing a quantitative analysis of the categories and sub-categories, of both deductive and inductive nature. This procedure will involve counting the occurrences of each category, putting these quantitative data in an Excel worksheet, and visually presenting the results through the use of comprehensive graphs. The reason for this integration of quantitative steps of analysis, is to provide useful meta-data about some aspects of the assessment behaviours registered by the diarist that might be useful for the discussion about the quality of her reflective practice.

3. ANALYSIS OF DATA

The aim of this chapter is twofold: (a) to illustrate the coding framework adopted to perform the deductive qualitative content analysis (QCA) and (b) to present the results of the qualitative content analysis performed on the source material, including the inductive (paragraph 3.1.), deductive (paragraph 3.2) and quantification phases (paragraph 3.4).

The deductive coding framework that has been developed and adopted in the current study is illustrated in the following table and is derived from the theory underlying the construct of diagnostic competence (Edelenbos & Kubanek-German, 2004: 278) – see page 26 above:

QCA INDUCTIVE CODING FRAMEWORK		
CATEGORY	DEFINITIONS (SUB-THEMES)	CODING RULES
1st level of diagnostic competence	(S1) Intuitive observations, isolated observations; (S2) observation geared to the whole class; (S3) observation geared towards the dominant pupils	include observations geared towards individual pupils (non-dominant) and specific groups of pupils ('some students...', 'other students...')
	<i>ANCHOR SAMPLE</i>	
	"i ragazzi erano molto entusiasti, la maggior parte di loro, molto competitive, volevano assolutamente scrivere le parole in modo corretto...alcuni invece non erano per nulla contenti, anzi sembravano molto in ansia" ¹⁵	
2nd level of diagnostic competence	(S4) Common knowledge about children as language learners; (S5) comments based on own learning history; (S6) first realizations that more contextual clues would be needed to interpret the situation	exclude observations geared towards whole class/dominant pupils; include reference to past teaching experience or own learning experiences (past language teachers); probing questions 'I wonder if...'; need to talk to colleagues/parents etc.
	<i>ANCHOR SAMPLE</i>	
	"questo alza il filtro affettivo (soprattutto per alcuni), quindi devo cercare di lavorare su questo aspetto, perché è fondamentale che si sentano a proprio agio nel chiedere spiegazioni e porre domande" ¹⁶	
3rd level of diagnostic competence	(S7) Ability to administer and correct ready-made test sheets in an appendix of a textbook, to classify pupils as good, average or weak learners; (S8) ability to bring together various observations (e.g., those gained through exchanges with colleagues), (S9) seeing language growth with a pedagogical	include comments on class tests (including INVALSI); include information from parent-teacher conferences and teacher meetings; pupil progresses and identification of

¹⁵ Translation: "the pupils were very excited, the majority of them, they were very competitive, they wanted to write the words correctly...but some others were not happy at all, they seemed very anxious".

¹⁶ Translation: "that raises the affective filter (especially for some pupils), so I should try to work on this aspect, because it is of fundamental importance that they feel at ease in asking for explanations or asking questions".

	eye.	pedagogical reasons that justify such progress;
	<i>ANCHOR SAMPLE</i>	
	"l'inglese sicuramente non è il punto forte di Holly ¹⁷ , parlando con la collega mi ha riferito che lo scorso anno il suo atteggiamento era anche peggiore, perché non sopportava l'insegnante di inglese" ¹⁸	
4th level of diagnostic competence	(S10) Ability to write a report with many prefabricated formulas; (S11) emerging ability of the teacher to create a narrative about an individual learner; (S12) ability to understand categories of diagnosis; (S13) ability to compare children using these categories; (S14) ability to judge a child's silence; (S15) ability to situate in terms of her/his own language acquisition trajectory; (S16) mental ongoing diagnosis leading to expansion, code-switching, etc.	include deeper reflections/remarks/concerns about any pupil or comparisons among pupils
	<i>ANCHOR SAMPLE</i>	
	"Hanno un livello di competenza simile in realtà, ma due approcci alla produzione linguistica totalmente differenti" ¹⁹	
5th level of diagnostic competence	(S17) Teacher becomes aware of how his/her categories and diagnosis are influenced by teaching style and professional biography (reflective practitioner); (S18) ability to understand basic features of simple tests, and the limitations of tests; (S19) ability to select and adapt such assessment instruments as contained in the various teacher handbooks; (S20) ability to correct and explain his/her expectations	Include any spontaneous changes or modifications to the lessons fuelled by observations and reflections based on diagnosing pupils' learning behaviours
	<i>ANCHOR SAMPLE</i>	
	"mentre osservavo i ragazzi ho realizzato che dovevo lavorare esplicitamente sulle strategie di comprensione e quindi ho deciso di fare la correzione insieme a loro subito dopo la fine della prova" ²⁰	
6th level of diagnostic competence	(S21) 'Pädagogischer Takt': ability to promote optimum student learning, i.e., to further autonomy, language awareness, self-assertiveness; (S22) to be able to give a rich hermeneutic interpretation of a language learning	Include only 'fruitful moments, where teachers know their class profited to the maximum' (Edelenbos & Kubanek-German, 2004: 279); include observations about whole class and individual pupils;

¹⁷ In compliance with applicable laws on data protection and in order to respect the pupils' anonymity, all the first names that appear in this paper are fictional. Likewise, other sensitive information that might lead to the identification of individual pupils has been redacted through the use of asterisks.

¹⁸ Translation: "Surely Holly isn't big on learning English, I spoke with my colleague and she told me that last year her attitude was even worse, as she despised her English teacher".

¹⁹ They have a similar level of competence, actually, but their approach to language production is utterly different".

²⁰ Translation: "while observing the pupils, I realized I had to work explicitly on comprehension strategies, and so I decided to check the test right after they complete it"

	situation, which becomes even more complex through dialogue, as a ongoing process	
	<i>ANCHOR SAMPLE</i>	
	“Notavo che si aiutavano tra loro, suggerendo strutture linguistiche corrette, cosa che succede raramente in classe [...] e volevano davvero fare l’attività, per una volta non erano annoiati o spaventati, volevano usare la lingua per FARE qualcosa, e si divertivano tantissimo, tanto che mi hanno chiesto di rifarlo!” ²¹	

The definitions are those provided by Edelenbos and Kubanek-German (2004: 278) as descriptors to the levels of diagnostic competence, and for the purpose of the present study, they function as sub-categories to the main categories (represented by the single levels of diagnostic competence, one to six). In the table above, the sub-categories are labelled as (S#), with a total of 22 sub-categories emerging from the theoretical framework in question.

3.1 Deductive Category Application

FIRST LEVEL OF DIAGNOSTIC COMPETENCE

This category was the most frequent one throughout the whole diary. The results of the inductive content analysis revealed that the diarist recorded this kind of assessing behaviour multiple times in every diary entrance. Within this category, some occurrences are compatible with the sub-category of ‘observations geared towards the whole class’ (S2), as illustrated by the following excerpts from the diary:

“Inizialmente non capivano, alla domanda ‘what’s the weather like today?’, ho ottenuto in risposta occhi fuori dalle orbite e sguardi perplessi...nessuno sembrava aver capito di cosa stessi parlando”.

“Devo dire che grazie a questa tecnica, i ragazzi erano insolitamente molto più attenti, forse si sentivano responsabili della correzione per il compagno...interessante”.

²¹ Translation: “I noticed they were helping each other out by suggesting the right linguistic structures, something that rarely happens inside the classroom [...] and they actually wanted to do the task, for once they weren’t bored or afraid, they wanted to use the language to DO something, and they were having so much fun that they asked me to do it all over again!”.

“Quasi tutti hanno ripetuto tra sé la parola, allora l’ho scritta alla lavagna e ripetuta [...] Sembravano molto interessati e intrigati dall’attività”.

“I bambini sono riusciti a rispondere a tutte le domande, anche se non sempre hanno usato le strutture linguistiche corrette”.

“Inizialmente erano confusi, non sapevano come fare per riflettere sulla tipologia di errori”

“[...] i ragazzi sembravano sollevati e colpiti da questa rivelazione. Sono abituati a dare molta importanza al voto, ma pian piano si dovrà andare oltre”.

These exemplars show that the observations entail a general evaluation from the teacher of the reactions to activities or formal/informal assessment tasks, especially regarding their oral comprehension skills or their oral production skills. The kind of diagnosis of learning behaviours that the teacher has shown, however, is fairly superficial, it aims at recording ‘a general sense of’ rather than a set of specific diagnostic episodes.

Within this category, it was possible to observe that the majority of occurrences fall within the sub-category of ‘observations geared towards the dominant pupils’ (S3), a label that in the coding process was expanded to include observations geared towards individual pupils or observations regarding some pupils in particular or part of the class. The following excerpts are representatives of this sub-category:

“alcuni sbirciavano nelle pagine precedenti [...] Daisy, Jenny, Paul and Hannah sembravano particolarmente preoccupati perché non sembravano trovare le informazioni che cercavano”

“Purtroppo solo Wendy, Jenny e Sharon hanno individuate qualche parola (sun, rainbow, cloud). Gli altri si sono limitati a osservare, senza fare nulla”

“evidentemente non aveva capito la consegna. Ho rispiegato che bisognava scegliere le frasi vere. Allora ha capito e ha risposto correttamente, anche se non ricordava il significato della parola ‘hungry’, vista durante la didattica a distanza”

“Inizialmente non capivano, ma poi Hope ha notato che c'erano delle doppie in alcuni verbi...allora hanno iniziato a focalizzarsi sullo spelling. Daisy ha notato l'assenza di una e in ‘writing’.”

“e qualche bambino non ricordava le preposizioni [...] ho notato che Louis e Jacob confondono in e on”.

The results that emerge from these data indicate that the majority of these assessment events indicate issues in oral comprehension, language production or gaps in pupils' knowledge. The tone of these observations is predominantly negative; the teacher-diarist shows an inclination to register either unsuccessful outcomes, or particularly outstanding performances. Other text passages focus on specific learning behaviours put in place by pupils, such as noticing or showing signs of language awareness, but the diarist simply describes these events, without deepening their causes, purposes or implications.

SECOND LEVEL OF DIAGNOSTIC COMPETENCE

The results of the inductive content analysis showed that the diarist recorded this kind of assessing behaviour less often compared to the first category. Within this category, the majority of occurrences are compatible with the sub-category (S4) common knowledge about children as language learners, as illustrated by the following excerpts from the diary:

“ho scelto David [...] e Jack, dato che so che sono molto amici, quindi David si sarebbe sentito a proprio agio. Ho notato che fare così ha molti benefici, quando si tratta di esercizi orali, le dinamiche socio-relazionali della classe vanno sempre tenute in considerazione, soprattutto in quinta, sono pre-adolescenti! Farli esercitare con me come partner, per esempio, è controproducente, perché si sentono troppo a disagio, sotto esame...uno stress eccessivo”

“la melodia effettivamente era molto accattivante, e [il video] proponeva la ripetizione della canzone cantando ‘sotto voce’, ‘a voce normale’ o ‘ad alta voce’. Ai bambini piacciono sempre questo genere di cose”

“adorano quando porto qualcosa di concreto in classe [realia]! Dovrei farlo più spesso effettivamente, mi rendo conto che sono sempre interessati e attentissimi”

“Come titolo ho fatto scrivere ‘survey’ e ho spiegato che ci saremmo trasformati in tanti piccoli statistici. A loro piace sempre quando si ‘cambia scenario’ e si interpretano ruoli nuovi”

“Così ho deciso di proporre un’attività di TPR [...]. Credo sia una delle tecniche migliori per questa classe...li aiuta a calmarsi e a mantenere la concentrazione, sono molto attivi e fisici nel loro modo di apprendere quindi penso che dovrei usare di più queste metodologie attive in futuro”

These text passages reveal that the teacher pays great attention to the target group she is teaching. She considers the characteristics of young language learners in terms of their preferences, their attitudes, and their specific skillset, in order to use the methodology or the methodologies that will contribute best to her pupils’ learning path. For example, it emerged that when teaching English to her younger pupils (third grade), bringing real objects in the classroom helps with their attention span, besides boosting their engagement and dedication to the learning objective(s). While, for example, when dealing with her older pupils (fifth grade), she needs to consider their physical-emotional developmental stage (pre-adolescence), and therefore she has to tackle different issues, such as low self-esteem peaks, quest for identity, fear of scrutiny from other classmates and other affective variables. No text passage has been coded as (S5) ‘comments based on own learning history’.

Regarding the sub-category (S6) ‘First realizations that more contextual clues would be needed to interpret the situation’, a coding rule was established so as to include all episodes in which the diarist refers to the need to delve into an episode and ask contextual

information about the pupil to her colleagues or to the pupil's parents. Furthermore, text passages in which the diarist sets out to monitor her pupils' progress have been coded using this sub-category. The following are representative examples from this sub-category:

“mi chiedo se abbiano seguito bene durante la didattica a distanza oppure se si siano persi qualche pezzo....dovrò verificare nei prossimi giorni”

“devo capire il perché Charlie si vergogni così tanto di parlare in inglese, non capisco...magari la prossima volta posso chiedere ai genitori se hanno qualche considerazione in merito”

“Dovrei confrontarmi con le colleghe di classe per vedere se è così anche nelle altre discipline”

These episodes indicate that the teacher could not explain the reason for certain behaviours and therefore she was not able to perform a proper assessment. The solution was that of referring to other sources (colleagues, parents) for obtaining more information that might help with diagnosing a pupil's behaviour.

THIRD LEVEL OF DIAGNOSTIC COMPETENCE

The results of the inductive content analysis indicate that the diarist recorded this kind of assessment behaviour quite rarely throughout the whole diary. Within this category, the analysis found only a few meaning units for each sub-category:

“ma d'altronde anche nel test somministrato in DAD Harry non aveva ottenuto un punteggio molto alto nell'attività di word-picture matching, ha delle difficoltà in ambito lessicale”

“Parlando con la madre di Charlie durante i colloqui, ha detto che è sempre così anche quando è a casa, fatica a concentrarsi e a stare fermo, in alcuni giorni più che in altri, [...] mentre in altri non c'è verso di farlo lavorare”

“Isla, Victoria e Hazel hanno chiesto di potersi esibire davanti alla classe, [...] si vedeva che si stavano divertendo molto e mentre ballavano cantavano le parole correttamente. Si vede che questo ambito è il loro forte, e i progressi linguistici arrivano di conseguenza”

“L’inglese sicuramente non è il punto forte di Holly, parlando con la collega mi ha riferito che lo scorso anno il suo atteggiamento era anche peggiore, perché non sopportava l’insegnante di inglese”

“Nel pomeriggio ho corretto i test [dell’unità 2, tratti dal libro di testo], e alcuni mi hanno sorpreso, per esempio Jacob e Lily hanno svolto tutti gli esercizi correttamente, ascolto incluso. Altri invece sono andati decisamente male (Adam, Ella, Noah...), ma mi aspettavo risultati sotto la media”

The diagnostic behaviours recorded by the diarist mainly concern episodes of ‘traditional testing’ such as administering and evaluating end-of-unit test sheets from the teacher guidebook included in the teacher pack that came with the class books, and therefore have been coded as (S7). These tests are calibrated to measure the extent to which the pupils have learnt the contents of the relative unit, but the teacher tried to use them in a formative rather than summative way (by providing individualized feedbacks to each pupil and by assigning them tailor-made make-up activity sheets). Nevertheless, she recorded some comments in her diary, which refers to the overall results of some pupils. In the excerpts above, some comments refer to particularly poor or under-average performances, as well as some surprisingly outstanding ones. The teacher thus demonstrates the ability to classify pupils as above-average or under-average, but this is a rather summative-like type of assessment, because it is based on the pupils’ performances on a single test.

As for the other sub-categories, the content analysis revealed some instances of discussion with colleagues or pupils’ parents in order to gather more information that might be useful for the assessment of problematic behaviours or under-average performances in formal or informal situations. Finally, the last sub-category, (S9) is also present in a couple of instances, but is quite rare. The teacher’s comment involves both an assessment of a student’s progress (language growth) and the hypothesis of a learning behaviour that

supports and promotes this kind of progress, such as the passion for dancing and singing in the example reported above. In other words, this kind of diagnostic behaviours suggests that the teacher reflects more deeply upon a certain evidence of progress and attempts to explain its causes more deeply than simply acknowledging and appreciating it.

FOURTH LEVEL OF DIAGNOSTIC COMPETENCE

This category, together with the following, was the second most frequent one throughout the whole diary, after the first one. The results of the inductive content analysis revealed that the diarist recorded this kind of assessment behaviour multiple times throughout the diary, but many sub-categories were not used during the coding procedure. No meaning unit has been identified for the sub-categories 10 (ability to write a report with many prefabricated formulas), 12 (ability to understand categories of diagnosis) and 15 (ability to situate in terms of her/his language acquisition trajectory). Sub-category (S11), 'emerging ability of the teacher to create a narrative about an individual learner' was most frequent sub-category. Here are some selected excerpts:

“Theo è più abile nel comprendere, ma fa più fatica ad esprimersi correttamente, anche se si butta, e produce lingua anche se non è sicuro al 100% della correttezza, fa delle ipotesi che vengono poi indirizzate da me. Hazel invece ha un approccio più conservatore, se così si può dire. Quando non sa come rispondere, sta in silenzio, e aspetta le mie domande-guida per procedere sul terreno dove ha qualche sicurezza, qualche appiglio. Hanno un livello simile di competenze in realtà, ma due approcci alla produzione orale completamente differenti”

“[...] tranne Jane che come sempre ha avuto bisogno di dieci minuti in più rispetto agli altri. La lentezza è una sua caratteristica intrinseca, è così anche nelle altre discipline, molto di più in quelle di ambito linguistico rispetto all'ambito logico-matematico, dove è leggermente più veloce e ricettiva”

“Louis mi ha chiesto se poteva tornare alla pagina precedente e guardare i vocaboli, ho risposto di sì, certamente. Lui ha sempre bisogno di un supporto, non gli piace quando non ricorda le parole o non sa dove guardare. Predilige attività dove ha un certo controllo sui materiali”

*“Kate in particolare sembrava molto scossa, lei reagisce molto male quando sbaglia, soprattutto in inglese, forse perché in italiano è cosciente di non sapere bene la lingua, mentre in inglese è molto più avanzata degli altri (in ***** fanno molte più ore di inglese), quindi si rimprovera sempre molto severamente quando sbaglia, è capitato in diverse occasioni che piangesse o si demoralizzasse”*

These passages present some detailed and specific narratives involving individual pupils. Some of them, like the first one, includes also the sub-category (S13), ‘ability to compare children by using categories of diagnosis’, because the teacher refers to two pupils, and explains that they have different approaches to expressing themselves in the foreign language. This text passage, therefore, incorporates two different sets of meaning units, but they have been presented together in order to facilitate the reader’s comprehension. As for the other excerpts, they contain additional information gathered by the teacher during the course of the school year that are used in order to explain or expand the narrative about that individual pupil. This information refers to pupil’s intrinsic characteristics as generic learners or as language learners, but they might also refer to their attitude towards language learning, as can be observed in the last excerpts.

Another sub-category within the fourth level of diagnostic competence, is (S14), ‘ability to judge a child’s silence’. This sub-category was present in the source material, but not as often as the previous one. Again, this category occurred in association with another sub-category, (S16), ‘mental ongoing diagnosis leading to expansion, code-switching etc.’, as it is possible to observe in the examples below.

“Ho chiesto ‘what do you see in this picture?’, indicando l’immagine sul libro. Ma Arthur non ha risposto. Era distratto da qualcosa che aveva sussurrato Adam. Ho ripetuto la domanda, ma Arthur

continuava a guardarmi in silenzio. Allora ho indicato il cane e ho chiesto 'what's this?' 'che cos'è questo qui?', allora si è illuminato ed ha urlato 'DOOOO!'

"Ho chiesto ad Ella 'what colour is the ball?' e lei ha risposto 'yellow', ho continuato suggerendo 'yes, right, it is yellow', mimando al contempo i mattoncini...allora ha capito e ha ripetuto 'it is yellow'"

"Quasi tutti hanno risposto bene, ma poi ho chiesto a Theo della provenienza di Rob, e lui non ha risposto, dal suo sguardo ho capito che non aveva capito la domanda. Allora ho suggerito 'I come from Italy, you come from Italy!' e ho indicato il disegno del pianeta alla lavagna.... a quel punto ha capito quello che intendevo dire, e ha risposto correttamente"

In all these episodes recorded by the diarist, it is possible to observe that while the teaching action progressed, she was able to tell that her pupils had some comprehension issues, and she could identify the reason (distraction, poor comprehension). She also chose the best strategy to cope with it, be it expanding her reasoning in order to give more contextual clues or switching to the pupil's native language in order to facilitate comprehension. The teacher also adopted non-verbal communication strategies, such as mimicking. In excerpt two, the teacher realizes that the pupil has understood the question, but she does not use the full structured answer she was hoping for; consequently, she illustrates the full answer by reminding the pupil of an agreed upon strategy (mattoncini, 'bricks' of different colours were used in order to teach grammar structures). Or again, in the last excerpt, the teacher points to an iconic aid (a drawing on the blackboard) in order to guide and support her pupil's comprehension issue, after ascertaining that the reason for his silence was that he was having trouble grappling with the meaning of the key-words in the question.

FIFTH LEVEL OF DIAGNOSTIC COMPETENCE

This category encompasses a deeper level of reflection involving the teacher's teaching style, level of expertise and expectations, besides his/her ability to manage testing in an optimal

way based on the classroom level and features. The data emerging from the content analysis performed in the source material revealed that only a minor portion of the identified meaning units belong to this category. Only one text passage was coded as (S17) 'teacher becomes aware of how his/her categories and diagnosis are influenced by teaching style and professional biography':

“poi ho proposto un’attività di rinforzo lessicale sul quaderno. Io tendo a farli parlare molto, a concentrarmi sulla competenza orale, ma nel tempo mi sono resa conto che se non scrivono sul quaderno, molto spesso dimenticano quello che facciamo, o non lo fissano bene. Lo ritengo un metodo un po’ ‘tradizionale’ (mi ricorda i miei anni alle elementari!), ma anche guardando i quaderni operativi delle mie colleghe, loro si basano molto sul quaderno ed essendo questo il mio primo anno, mi sento di fare affidamento su di loro, per ora...”

This passage refers to both the lack of experience and classroom practice of the diarist, who is a novice teacher at her first experience, and an explicit reference to her teaching style, which is based on the communicative approach. She focuses on teaching her pupils to actively use the foreign language, all the activities she presents are embedded in a situation or context and have a clear purpose. She also believes in the power of a constant interaction and exposure to the foreign language and in the assessment process she focuses on fluency rather than accuracy. In this episode, however, she calls into question the effectiveness of her style, because she feels like her pupils are not fully benefiting from her vision. She believes they also need to write things down in order to remember them more easily and more readily, and also as a future reference. She points out that her elder (and therefore more experienced) colleagues rely heavily on workbooks during their lessons, so she assumes it is a 'safe' choice, for the moment, to model her approach on theirs, in order to cover the basics, even if she concedes that this method is a fairly traditional one. The teacher's inexperience is key here in order to interpret this reflection. She questions herself as a practitioner because she sees herself as a 'trainee' that needs guidance and counselling from her more knowledgeable peers, something that might happen during weekly teacher

meetings or during inter-team meetings (meetings involving teachers of the same subject area).

Concerning the following sub-category – (S18) ‘ability to understand basic features of simple tests, and the limitations of tests’ – during the analysis it emerged that the diarist recorded only a few examples of this:

“Ho volute ripassare il tempo atmosferico perché la settimana scorsa tra le prove di simulazione INVALSI che avevo proposto in classe, c’era una prova di ascolto sulle previsioni meteo, e anche se la struttura della prova era piuttosto intuitiva, c’erano solo delle icone, quindi la mancanza delle parole non aiutava i ragazzi, anzi...e mi sono accorta che quasi nessuno ricordava quei contenuti, anche se sono sicura che li avessero già affrontati gli scorsi anni”

By observing this text passage, it is possible to conclude that the diarist recorded her opinion about a test (a mock-test for the INVALSI test), which she considered to be too difficult for her pupils’ level, as it was lacking written words, which, together with the icons that were already there, would have supported the pupils’ oral comprehension. This realization brought her to redefine the learning path and introduce new learning objectives.

Another excerpt has been coded as pertaining to the sub-code (S19) ‘ability to select and adapt such assessment instruments as contained in the various teacher handbooks’:

“Durante l’ultima mezz’ora di lezione ho somministrato una prova di verifica relativa alla scorsa UdA. Ho modificato le prove differenziandole per livelli, dato che ormai mi è piuttosto chiara la situazione della classe. Ho anche aggiunto una task di comprensione orale per rendere la prova più completa. Osservando i bambini mentre svolgevano la prova, mi è sembrato che quelli della fascia ‘debole’ fossero più sereni, e quelli della fascia ‘alta’ più stimolati”.

Here, the diarist refers to the act of evaluating and manipulating tests in order to create a tailor-made version that can be adapted to each and every student in her classroom. She was able to divide her pupils according to their proficiency in English, so that they could

have an instrument that would capture their abilities more faithfully. Furthermore, the diarist recorded the impressions and feelings she got by observing while the pupils completed their tests, and she was reassured that she had managed to improve their learning experience.

Finally, regarding the last sub-category, (S20) 'ability to correct and explain his/her expectations', this was the most frequent one amongst the meaning units coded within this main category, and here are some representative meaning units:

"Kelly mi ha poi chiesto come mai biblioteca si dicesse 'library' e non 'bibliotec', per esempio. E io le ho risposto che è un false friend! Facce confuse, nessuna reazione. Non avevano idea di cosa fossero i false friends! Incredibile. Ero convinta che gli scorsi anni li avessero fatti, almeno i più comuni. Quindi ho deciso di dedicare alcuni minuti a questo argomento, anche perché i ragazzi sembravano divertiti da questo fenomeno...continuavano a chiedermi altri esempi"

"[...] ho approfittato per verificare anche le altre wh-questions, e con mio grande disappunto ho notato che fanno ancora fatica a distinguerle! Quasi tutti ricordavano who, però...curioso. Allora abbiamo ripassato questo argomento usando la tecnica della mano, che avevo presentato loro alcune settimane fa"

"Avevano già fatto gli ordinali con la supplente precedente, a novembre, ma purtroppo ho constatato che si ricordavano ben poco. A questo punto ho dovuto ridimensionare le mie aspettative, quindi abbiamo ripassato tutti i numeri, compresi gli ordinali fino al 10"

In these episodes, the diarist records her (mostly bewildered) reactions when realizing her pupils lacked certain key competences or basic knowledge that she thought they had mastered by then. She talks explicitly about lowering her expectations or deviating from the planned lesson in order to fill in the gaps in the pupil's learning path. Throughout the diary, the episodes recorded by the teacher suggest that the pupils need frequent revisions of basic contents, in order to make up for the discontinuous learning path they experienced.

SIXTH LEVEL OF DIAGNOSTIC COMPETENCE

Based on the definition provided by the theoretical framework, no meaning units from the source material fit perfectly into this category. However, the theory indicates that there might be some instances of 'fruitful moments', in which the teacher is particularly pleased with the outcome of a lesson, and feel like they have created an ideal learning environment. The occurrences that emerge from the data are very rare, but nonetheless, some instances of what comes close to the idea of 'fruitful moment' can be identified:

"Notavo che si aiutavano tra loro, suggerendo strutture linguistiche corrette, cosa che succede raramente in classe [...] e volevano davvero fare l'attività, per una volta non erano annoiati o spaventati, volevano usare la lingua per FARE qualcosa, e si divertivano tantissimo, tanto che mi hanno chiesto di rifarlo!"

"e poi Hope è intervenuta dicendo che questa parola la uso sempre io quando devo dirgli che è ora di uscire dopo la mensa. Luke le ha dato ragione. Questo commento di Hope mi ha reso particolarmente felice, perché significa che è stata in grado di collegare un momento di comunicazione autentica ad un momento di apprendimento didattico. Sono queste le cose che mi dicono più di mille verifiche scritte...la capacità dei ragazzi di usare la lingua e comprenderne gli usi 'nella vita reale' vale molto di più di una verifica eseguita alla perfezione, secondo la mia opinione"

"Victoria ha notato che nella storia ho detto 'he' per parlare di Rod, ma lui è una rana quindi bisognerebbe usare it, come avevo detto loro la scorsa volta. Mi ha fatto molto piacere questa sua osservazione, significa che ha assimilato il ragionamento corretto e ha trasferito le sue conoscenze da una situazione nota ad una non nota"

In the first excerpt, the teacher has documented a particularly profitable and productive activity she was able to organise in the school garden. She was surprised to see her pupils cooperate and make an active use of the foreign language, something that, as she refers, rarely happens during indoors lessons. This passage was coded in this category because it

reflects critical aspects of learning behaviours that were being diagnosed by the teacher: autonomy in the use of the target language, use of learning strategies and attitude towards the target language. The second and the third excerpts, instead, have a different quality. In these two episodes, the diarist recounts particularly brilliant remarks made by some pupils. She highlights that the interesting part of these considerations is the reasoning behind them, namely the fact that the pupils were able to make connections between the language used in an authentic context and that used during a lesson, and this kind of assessment information is more powerful than any other summative information that may stem from a test or any form of formal assessment.

3.2 Inductive Category Development

In this second phase, the aim was to create a new category system in order to include any other aspects that were not covered in the previous step of the content analysis. First of all, an overarching theme for the development of categories was defined deductively on the grounds of theoretical studies examined in the previous paragraph. The researcher chose to focus on aspects connected to learning oriented assessment practices, as theorized by Turner and Purpura (2015) (see paragraph 1.2.3.) and on Wallace (1991)'s reflective cycle (see paragraph 1.3.1.).

The initial set of categories identified within the source material was subsequently refined and reorganized, in order to create clearly defined and separated codes. The following categories were created: (a) teacher's emotional challenges linked to assessment episodes; (b) teacher's reflections about future action; (c) changes implemented by the teacher based on reflections about past action. An example of the open coding analysis process:

<i>TEXT PASSAGE(S)</i>	<i>DEFINITIONS / WORKING CATEGORY</i>	<i>SUB-CATEGORY</i>	<i>MAIN CATEGORY</i>
<p>“è stato molto divertente assistere alle loro rappresentazioni, alcuni sono particolarmente teatrali...è un modo molto piacevole per valutarli in modo informale”</p> <p>“in mensa finalmente Jenny ha usato la struttura corretta per chiedermi il formaggio grana! Che soddisfazione!”</p>	<p>The teacher reports positive emotions linked to assessment episodes</p> <p>pleasure/wellbeing satisfaction/pride</p>	Positive emotions	Emotional challenges of assessment
<p>“è stato frustrante dover intervenire sempre per correggerlo”</p> <p>“Hanno avuto tantissime difficoltà con le ultime due task, perché si basavano sulla produzione scritta e il livello era decisamente troppo alto...inizio ad essere seriamente preoccupata da questa prova”</p>	<p>The teacher reports negative emotions linked to assessment episodes</p> <p>frustration concern/anxiety</p>	Negative emotions	

3.2.1 Emotional Challenges of Assessment

The inductive phase of the qualitative content analysis performed on the data revealed the presence of many meaning units pertaining to the emotional/affective dimension of assessment, one of the components of learning-oriented assessment. The diarist reported her feelings and mental state with reference to specific aspects of her teaching action. Some aspects, such as emotional reactions triggered by student behaviour, interactions with families or colleagues, and institutional engagements for example, were discarded from the analysis, as they do not relate with the overarching theme that was selected. In line with the purpose of the present research project, and considering the research questions, only the diarist’s feelings and emotions directly associated with assessment episodes or diagnostic behaviour were included in the coding process.

More specifically, the results of the coding process showed the presence of emotions, feelings and mental states that may be termed ‘negative’, such as frustration, concern, disappointment, sadness or anxiety, as evidenced by the following excerpts from the diary:

“è stato frustrante dover intervenire sempre per correggerlo”

“Hanno avuto tantissime difficoltà con le ultime due task, perché si basavano sulla produzione scritta e il livello era decisamente troppo alto...inizio ad essere seriamente preoccupata da questa prova”

“Stranamente Victoria mi ha chiesto il significato di isn’t [...] Questo mi ha un po’ scoraggiato perché solitamente lei acquisisce molto bene gli argomenti che affrontiamo insieme, quindi credo di non aver dedicato abbastanza tempo a questa forma...o forse ho proposto le attività sbagliate....sono un po’ persa...”

“Avevo programmato di ritirare le prove e correggerle a casa, per poi fare la restituzione e il feedback collettivo, ma poi ho deciso che sarebbe stato più produttivo ed efficace fare la peer review subito dopo la prova. Non l’avevo mai fatto prima, quindi ero molto nervosa, ero preoccupata di non riuscire a gestire questa nuova pratica...”

“Holly e Paul ancora non avevano capito che erano due forme diverse dello stesso verbo, quindi ho ripreso la coniugazione per l’ennesima volta. Molto frustrante che ancora non abbiano appreso questi concetti basilari.”

“Aaron ci è rimasto molto male ovviamente [...] mi spiace sempre molto quando devo penalizzare studenti in gamba, vedo che si impegnano, ma non riescono bene in attività di questo tipo...vorrei non dover fare queste prove...”

“Ella e Adam ancora non usano la frase corretta per chiedere di uscire [...] non capisco come mai a questo punto dell’anno non l’abbiano interiorizzata, non nascondo di essere parecchio preoccupata”

As for the kind of diagnostic activity that the diarist was carrying out while experiencing these negative feelings, they include both formal and informal assessment moments.

For example, in excerpts 1, 3, 5, and 7 the teacher is performing an in-action assessment, she is evaluating her pupils’ abilities and competences while performing instructional activities such as revising, completing activities and exercises, or during semi-formal moments. She is mainly disappointed in the outcome of these assessments: she realizes that the pupils still haven’t reached a determined learning objective, for example.

In other instances, instead, she is expressing negative judgements about formal assessments such as testing, as evidenced by excerpt 6, where the teacher feels sorry about giving a bad mark to a student that she considers to be generally capable and skilled, so she is criticising an assessment tool that does not fully reflect that child's actual competence in English. Excerpt 2 is also concerned with a formal kind of assessment (in that case it was an INVALSI test), so the diarist expressed a state of worry about this test in relation to her pupils' level of preparation. In excerpt 4, instead, the diarist relates to her fear of not being able to guide her class through a new kind of formative assessment procedure (i.e.: peer review), because she is a novice teacher with very little classroom experience, so it will be a 'trial and error' kind of experience.

On the other hand, the analysis also resulted in the identification of an array of positive emotions linked to classroom assessment episodes, such as pleasure, wellbeing, satisfaction, pride and happiness, as it is possible to observe in the following diary excerpts:

"è stato molto divertente assistere alle loro rappresentazioni, alcuni sono particolarmente teatrali...è un modo molto piacevole per valutarli in modo informale"

"in mensa finalmente Jenny ha usato la struttura corretta per chiedermi il formaggio grana! Che soddisfazione!"

"questo commento di Hope mi ha reso particolarmente felice, perché significa che è stata in grado di collegare un momento di comunicazione autentica ad un momento di apprendimento didattico"

"Sono contenta che stiano facendo dei progressi. Inizialmente avevano delle difficoltà, ma ora sembrano più sicuri nel produrre frasi in lingua. Sono molto soddisfatta di vedere questi miglioramenti"

"Hannah ha [...] mobilitato delle risorse reperite spontaneamente in un contesto informale. Mi sono sentita davvero molto fiera di lei, e le ho fatto i complimenti in modo esplicito"

As it emerges from examining these text passages, all of these instances occur during informal moments, when the diarist is talking about classroom diagnostic activities, like for example in the first excerpt. In the whole diary, no meaning unit coded as 'positive emotions' is attributable to summative-like testing or formal assessment episodes. The diarist recorded especially strong positive feelings especially when she noticed her pupils were showing evidence of being autonomous or self-directed learners, as in excerpts 3 and 5. She also expressed emotions of joy and pride when recording episodes in which her pupils were making progress or were showing evidence of good quality language growth.

3.2.2 Reflective Cycle

The second group of categories that has been inductively identified within the diary, deals with some components of the reflective cycle, such as reflection-before action (category labelled as 'future actions') and reflection-on-action (category labelled as 'changes based on past actions'). The majority of the meaning units pertaining to this theme have been coded as 'teacher's reflections about future action'. Here are some excerpts from the diary:

"ma come fare per stimolare anche gli altri? Dovrei organizzare qualche attività per saperne di più sui loro gusti, sulle loro passioni [...], nei prossimi giorni penserò a qualcosa da fare, magari anche nel dopo-mensa, quella mezz'ora di tempo potrebbe essere davvero preziosa"

"Dovrò lavorare di più la prossima settimana, per organizzare delle attività migliori, più complete. Nei prossimi giorni però potrei inserire questa domanda sul tempo atmosferico nella nostra morning routine, così posso verificare in itinere l'apprendimento di queste strutture comunicative"

"Dovrei predisporre delle attività mirate in questo senso, prendo nota. [...] Forse dovrei proporre delle attività ogni giorno, magari facendo loro delle domande. Mi viene in mente un dado, potrei lanciarlo ogni mattina e ad ogni numero corrisponde una domanda..."

“Devo ricordarmi di fare sempre così anche per le prossime lezioni, è stato un errore pensare che ricordassero tutto. [...] La prossima lezione potrei integrare usando una sedia e dando dei comandi che includano le preposizioni”

“Venerdì dovrò dedicare almeno un’ora a questo tema, penso sia davvero importante che almeno i numeri siano ben acquisiti. Approfitterò per introdurre anche quelli dopo il 20.”

In these excerpts, it is possible to observe how the diarist reflects on a future teaching action by hypothesizing activities to be presented to the class and learning objectives to be achieved in the future. Some of these reflections are based on negative outcomes experienced during class, such as in excerpt 2 and 4. Others, instead, are based on the teacher’s desire to further explore and look into her pupils’ learning needs, like in excerpts 1. Finally, some other examples involve the need to revise and solidify the pupils’ abilities, like in excerpts 3 and 5, where the teacher sets out to achieving some specific learning goals or decides to introduce new methods to assess her pupils in an informal manner.

In addition to this category, another sub-theme that emerged from the diary was that of ‘changes implemented by the teacher based on reflections about past action’. Only a few instances were coded in the source material, and they are temporally located at the end of the study, and they involve just one class, the third grade. Here are the excerpts:

“Dato che rileggendo il diario ho notato che i bambini erano sempre un po’ agitati durante la lezione introduttiva di una nuova UdaA, ho pensato di presentare loro i numeri attraverso un approccio più pratico”

“Anche per questa UdA, ho deciso di iniziare da un’attività pratica invece che affidarmi solamente al libro. Ho quindi proposto loro di creare un robot utilizzando cartoncini e sagome stampate... attività ideale per una classe di bambini attivi e particolarmente pratici come loro!”

These instances reveal that the teacher openly talks about the decision to modify her teaching method in order to better cater for the specific learning needs of her pupils, who are described as a particularly active and ‘hands-on’ group. The decision to change her approach to the introduction of a new learning unit to the class resulted from the reflections she recorded during the previous lessons, and from reading about a problematic situation she had reflected upon in previous entries.

3.3 Quantification

Finally, a quantification was performed by counting the deductive and inductive categories occurrences within the source material in order to provide supplementary information. The results are presented in the following four visual diagrams. The pie chart in Figure 18 illustrates the percentages regarding the categories of assessment behaviours identified in the whole sample. The bar graph in Figure 19 illustrates the temporal evolution of the occurrences of the deductive categories in the study, while the other diagrams (Fig. 20 and Fig. 21) illustrate the inductive categories that emerged from the second part of the content analysis.

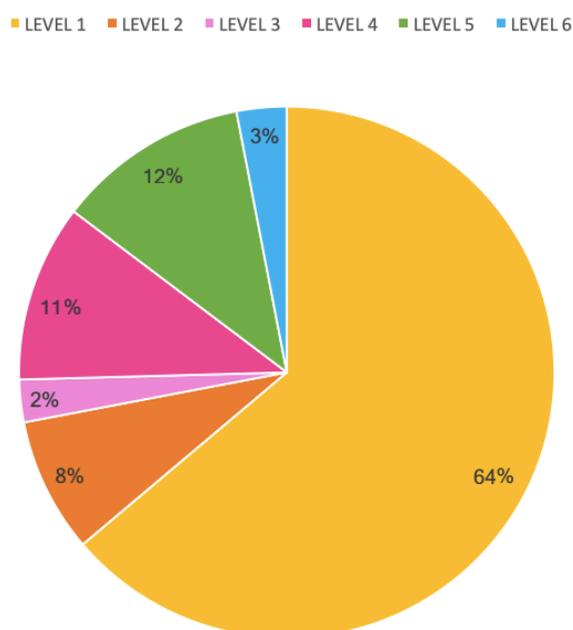


Figure 18. Percentages of categories within the whole sample.

The results from the whole source material (Figure 18) show that, in general, the majority of the text passages (64%) was coded as ‘1st level of diagnostic competence’ (LEVEL 1: intuitive

and isolated observations, or observations geared towards the dominant pupils), while all the other levels were coded with significantly lower frequencies. Categories '2nd level of diagnostic competence' (LEVEL 2: common knowledge about children as language learners; comments based on own learning history; first realizations that more contextual clues would be needed to interpret the situation), '4th level of diagnostic competence' (LEVEL 4: emerging ability of the teacher to create a narrative about an individual learner, ability to compare children) and '5th level of diagnostic competence' (LEVEL 5: teacher becomes aware of how his/her diagnoses are influenced by teaching style; ability to correct/explain his/her expectations) were all coded with similar frequencies, ranging from 8 to 12%. The least frequent categories within the diary are '6th level of diagnostic competence' (LEVEL 6) and '3rd level of diagnostic competence' (LEVEL 3: ability to administer and correct ready-made test sheets, to classify pupils as good, average or weak learners, ability to bring together various observations, seeing language growth with a pedagogical eye), which appeared much less than the other categories (3% and 2%, respectively).

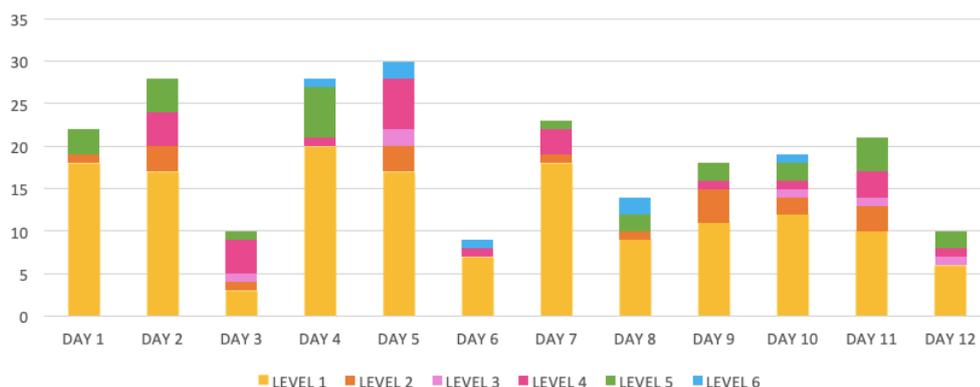


Figure 19. Temporal distribution of deductive categories occurrences within the whole sample.

Considering the same categories in terms of frequency during the study time framework, the results (Figure 19) reveal that, throughout the study, the frequency of LEVEL 1 diagnostic competence is higher during the first half of the study, and then it tends to decrease during the second half of the study, from day 8 onwards. As for LEVEL 5 diagnostic competence, the data show that this kind of occurrence fluctuates more at the beginning of the study, while it becomes more stable during the second half of the study,

even if the frequency in this phase is lower. With respect to LEVEL 6, the data show that this occurrence seems to emerge in the midway section of the study (from day 4 to day 10). Examining the graph, however, it is clear that the other levels of diagnostic competence fluctuate throughout the study, with no apparent trend.

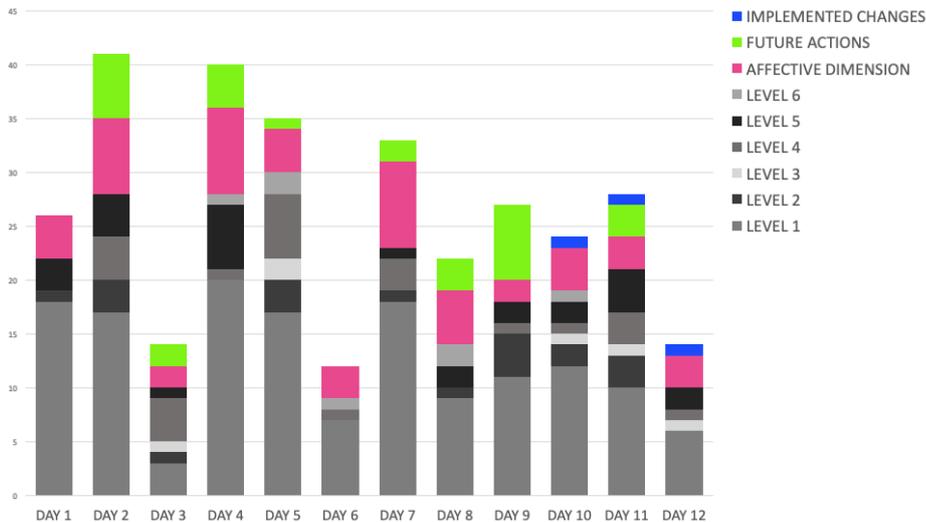


Figure 20. Temporal distribution of inductive categories occurrences within the whole sample.

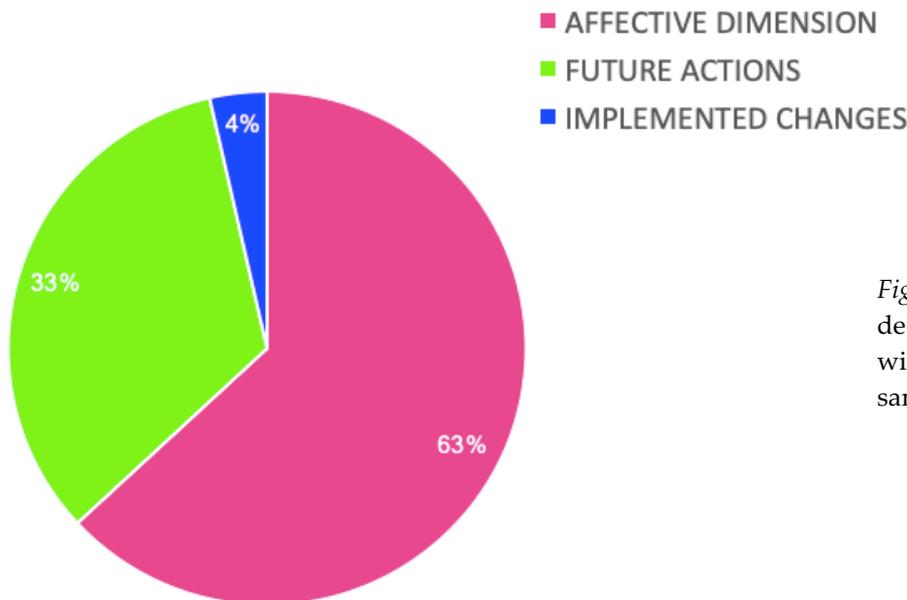


Figure 21. Percentages of deductive categories within the whole sample.

Considering instead the categories that have been inductively extracted from the source material, namely (i) affective dimension, (ii) future actions, and (iii) implemented changes, it emerges from the data, the bar graph (Fig. 20) shows that this last category is present only

in the last period of the study (days 10-12), while the affective dimension and future actions categories are distributed quite evenly throughout the diary study. As shown in the following pie chart (Fig. 21), the dominant category among the inductive ones, based on the frequency of appearance in the text, is the 'affective dimension' (63%), which represents more than half of the text passages coded during the inductive phase of the QCA. As for the 'future actions' category, the data show that 33% of the text passages has been coded under this category, while only 4% of the occurrences refers to the 'implemented changes' category.

4. PRESENTATION OF RESULTS

In the current chapter, the data analysed in the previous section will be discussed. In the first section (paragraph 4.1) the data will be discussed in relation to the research questions and the general aim of the study. In paragraph 4.2, some remarks will be made about the limitations of the current study, and some implications concerning future research will be outlined.

4.1 Discussion

This diary study presents the experience of a EFL primary school teacher (the author of this study) who set out to implement a reflective methodology in order to comply with new national regulations on assessment and evaluation for the primary level. When analysing the data regarding the assessment of young EFL learners, the themes that emerged from the qualitative content analysis of the teacher diary ranged from the different levels of diagnostic competence attributed to the teacher on the basis of her narrative, to other aspects of classroom assessment such as emotional challenges or the implementation of a reflective cycle on the part of the teacher-diarist. The discussion of the results reported in the previous paragraph will be structured around the research questions, as formulated in paragraph 2.2.

FIRST RESEARCH QUESTION: How can reflective teacher diaries be used for the assessment of EFL primary pupils?

Taking into consideration the results of the inductive qualitative content analysis, it is possible to notice that they are consistent with the ongoing assessment phase of the diary study (phase 2, as described in Fig. 17, see paragraph 2.3.3.). In this phase of the diary study, the diarist reflected actively on the entries as they were recorded in the diary, by systematically reading and re-reading them each week. As supported by the data presented in the previous paragraph, the teacher implemented some changes during the study, and these changes originated from the reflective practice that she adopted. The answer to the first research question, 'How can reflective teacher diaries be used for the assessment of EFL primary pupils?' needs to take into consideration two aspects.

Firstly, the findings indicate that the diarist's accounts of the many informal assessment episodes that happened during her English lessons triggered a reflective routine in which the teacher would verbalize issues or obstacles experienced during classes, together with exposing her 'experiential knowledge' and would try to make sense of them on the basis of her knowledge of the target learning group. For example, as seen in paragraph 2.4.2.2, the teacher would notice that the pupils were restless or distracted while she introduced new content from a learning unit, and by reflecting on this problematic aspect of her practice, she realized that her pupils' specific learning needs required a new, more 'hands on' approach. This finding is consistent with the notion of 'reflective cycle', as theorized by Wallace (1991), who describes this type of experience and claims precisely that it arises from the practitioner's direct teaching experience, and notes that there is a constant flow of information exchange between a teacher's experiential knowledge and his/her received knowledge, which stems from his/her previous educational background.

A second aspect is that the ongoing informal analysis carried out by the teacher-diarist while writing her diary served as an ongoing informal assessment of her pupils that was based on empirical data from observation, which is a critical component of classroom-based assessment. These results also reaffirm the established research in classroom-based approaches carried out by Rea-Dickins (2001), who claims that this kind of assessment is 'routinely embedded' within daily teaching actions and guides teachers in many different decision-making processes that range from what is being taught next, what needs to be revised and how it is going to be taught. In particular, the data gathered from the diary was used by the teacher-diarist at the end of the school year, when she had to formulate an evaluation of the overall learning levels reached by her pupils in English. This evaluation had to follow the new normative requirements specified in the Ministerial Order n. 174 (4/12/2020), to be applied starting from the 2020/2021 school year. As already described in paragraph 1.2.6, this has required a new approach to assessment and evaluation, in the perspective of assessment *for* learning. By abandoning the idea of assessment *of* learning, the ministry indicates that assessment should value the gradual process of knowledge-building put in place by the pupils. The diary proved to be an extremely useful tool in this

sense, as it allowed the teacher to observe how her pupils' language proficiency was evolving over time. The informal analysis carried out by the teacher has focused on the various learning dimensions highlighted by the Ministry in the *Guidelines* (2020), namely (a) degree of autonomy in learning, (b) type of learning situation (known or unknown), (c) type of resources employed by the student, and (d) continuity in the manifestation of learning. There is a considerable body of research which recognizes the fundamental role of introspective methods for the implementation of 'assessment for learning' strategies. For instance, Gardner (2012) claims that in practical terms, assessment for learning is implemented in classroom through 'pedagogic style, student-teacher interaction, self-reflection (teacher and student), motivation and a variety of assessment processes (2012: 3). The teacher diary can be considered a tool for teacher self-reflection aimed at promoting better teacher self-awareness and therefore a better learning environment. The data that emerged from the diary, indeed, supports the teacher's self-awareness also in terms of her emotional sphere. The emotional components that emerged in the various entries were reported in paragraph 2.4.1, where both positive and negative emotions were manifested by the teacher in connection with various types of classroom assessments, and this strongly improved her level of awareness in terms of feelings and emotions linked to assessing her pupils.

Furthermore, in order to answer the first research question, the teacher diary was also analysed by taking into consideration the construct of diagnostic competence, as theorized by Edelenbos and Kubanek-German (2004), who outlined different levels of diagnostic competence that can be attributed to foreign language teachers (see paragraph 1.2.4.). As it emerges from the results presented in paragraph 2.4.1, the main use of the teacher diary was that of creating a sort of 'interactive reflective repository', a collection of observations and reflections based on these observations noted down by the teacher-diarist. The interactive aspect of this tool is that during the writing stage the teacher would consult the various entries and look for recurrent patterns, unforeseen developments or unexpected dynamics within the classrooms. More specifically, the results suggest that the teacher's diagnostic behaviours that have been recorded in the diary fall predominantly within the first level of

diagnostic competence, while only a few instances might be characterized as indicators of the other levels of competence. This result might be explained in light of the paucity of experience possessed by the teacher, who is a novice practitioner at her first full-time teaching experience. Nevertheless, the diary also reports some episodes which have been inductively categorized as 'teacher's reflections about future action' ('future actions' in short). In her diary, the teacher would refer to some unforeseen developments or incidents that hindered her planned action, she would speculate about the reasons behind these happenings and, ultimately, she would make alternative plans for the next lesson, in order to find a way to tackle and actively overcome the issue. There is a considerable body of research which examines the mechanisms and components of this kind of 'reflection-in-action'. So, as Altrichter et al. (2005) assert, reflection-in-action arises when teachers encounter some kind of obstacle to the usual flow of action. Wallace (1991), explains that this kind of reflection may indeed occur after a particularly bad professional performance and it may foster a process of self-improvement for future action. This finding is also consistent with Wilson (2008)'s view of the future dimension of reflection (which he calls 'reflection-before-action'), that implies an evaluation, on the part of the teacher, of possible future scenarios, thus demonstrating 'anticipatory reasoning' abilities (van Manen, 1991). An interesting finding that emerges from the data is that the frequency of 'first level observations' tends to decrease towards the end of the study, suggesting that there might be a qualitative change in the last block of diary entries, where the teacher registers less intuitive/collective observations about the whole class and focuses more on individual pupils or on other aspects of her teaching, thus deepening her reflections by including elements from the other levels of diagnostic competence. A possible explanation for this might be that the teacher's ability to reflect on her practice was gradually improving because she was becoming more familiar with the diary method.

To answer the research question, the data provided by the diary entries represent an authentic form of formative assessment, in that the feedback that the teacher extracted from her daily journaling was then used to transform her actions in the classroom in a way that would sustain and scaffold her pupils' learning. In fact, according to Black et al. (2004),

assessment for learning (formative assessment) is a type of self-reflection that actively employs feedback as a way to modify the teaching and learning activities in order to meet learning needs.

SECOND RESEARCH QUESTION: What are the advantages and disadvantages of reflective teacher diaries for the professional development of novice EFL teachers?

The second question in this research is ‘What are the advantages and disadvantages of reflective teacher diaries for the professional development of novice EFL teachers?’.

In order to answer to this question, it is possible to interpret the quantitative findings presented in paragraph 2.4.3. in a positive way by considering the progressive decrease of meaning units coded as ‘first level of diagnostic competence’ as a sign of development and growth in the teacher’s ability to reflect critically on her practice.

On the basis of the data presented in paragraph 3.2.2, it appears that the teacher was actually able to actively use the feedback she extracted from the diary in order to modify and adapt her practices, like for instance when she noticed that the pupils of her younger class (3rd grade) needed a different approach to the introduction of a new learning unit. She noted on her diary that she managed to change the old approach and adopt an active methodology in order to better suit her class’ learning needs. This represents a key advantage of journaling in a reflective way: because she was critically reflecting on the reason why her pupils were so restless and distracted, she was able to make hypothesis and test them, thus discovering that she needed to use a different approach. These findings also reaffirm the established research in FL teaching which had found similar results (Jeffrey, 2004; Sá, 2002). Nevertheless, based on the personal experience of the teacher-researcher, a disadvantage of the diary method is linked to the so called respondent fatigue, a phenomenon that has been widely cited in the literature (Dörnyei, 2007; Bailey, 1991; Bolger et al., 2003). According to the teacher-researcher, keeping a diary has been a time-consuming activity, and on occasion it has proven to be challenging to write down the entries on the same day of the lesson, because of the teacher’s many institutional commitments, such as staff meetings, teacher body collegiate gatherings or interviews with children’s parents. In days when the teacher

was particularly busy with planning or performing her institutional duties, it was hard to find the time and the energy to mentally revise the happenings of the lesson and carefully examining them in a critical way. Furthermore, in those days in which the entries were written later in the evening – due to the aforementioned institutional commitments – issues of memory caused the recollection to be less precise and thorough. This disadvantage has largely been addressed in the literature, and it is one of the main reasons why diaries are not very popular amongst researchers (Fry, 1988; Nezlek, 2012).

Another major disadvantage of journaling is the lack of focus on individual pupils. The teacher was able to describe and reflect on the class as a general unit, but she struggled with focusing on individual pupils, except for some specific cases. The data confirm this aspect, in that the occurrences that belonged to the fourth level of diagnostic competence (namely sub-category S11-emerging ability of the teacher to create a narrative about an individual learner) were far less present in the diary. Observation grids might prove more useful if the objective is that of collecting data for every pupil, while the diary in this study served more as a ‘helm’ that helped the teacher to navigate her daily practice.

4.2 Limitations and Further research

Regarding the limitations of the present study, the first one involves the limited sample, namely the fact that the data were collected from only one subject, so that the results are limited to the specific case involved and cannot be generalized to a larger population. Even though this study follows the design of an action research, it would have benefitted from the participation of more teachers, thus becoming a collaborative action research. Gathering data from different teachers would have allowed for a comparative analysis of different journaling experiences, and including a phase of collaborative reflection with a group of teachers would have enhanced this research project.

Another limitation concerns the data collection instrument itself, even if it represents the main core of this research study. Some limitations of diary methods have been already discussed in paragraph 1.3.2. This methodology allows for a deep and unique plunge into a subject’s consciousness, thus allowing the researcher to access some of the hidden

perspective provided by an individual's innermost thoughts. At the same time, however, these data are highly subjective and therefore they might provide a biased, nonrepresentative view of the classroom reality, which might have been different if captured by one or more external observers.

The timeframe of the study could also be improved. The duration of the data collection phase, in particular, could be longer, so as to deliver a fuller and more thorough perspective on the teacher's reflective journey towards professional development. The limited duration of the collection phase of this study (one month), in fact, could be considered a flaw of this study's design, as it did not allow for a complete examination of long-time effects of journaling, as it could have been the case with a longitudinal approach.

Regarding the main objective of this research, namely exploring the efficiency of reflective practice in foreign language teaching, several studies could be brought forward in order to deepen the knowledge of aspects concerning reflective teaching applied to assessment in foreign and second language education. It would be interesting to carry out a diary study on a larger scale, by involving teachers of different areas of Italy in order to compare results from the analysis of their journals. Another fruitful avenue to explore would be that of collaborative reflective practice, in which teams of teachers working in the same school or district would come together and discuss about their journey towards professional development.

Further quantitative research in the field of reflective teaching and journaling would also be necessary, in order to study the effects of implementing this practice within the context of primary schools, both in Italy and abroad.

CONCLUSION

The current MA research project was designed around a precise practical urge: finding an efficient way to implement the new regulations on assessment and evaluation imposed by the Italian Ministry of Education with the publication of the ministerial order n. 172/2020. This reform foresees a paradigm shift from a summative to a formative way of assessing pupils learning levels. This represents a challenging adjustment that will be completed in the years to come, through the hard work and continuous strive for training and professional development of teachers and educators. It is precisely this strong desire to improve oneself that brought the researcher to investigate, through an action research, the implementation of a teacher diary method to assess her young pupils of English as a foreign language in a learning-oriented way.

The data collected during the course of one month by the diarist was analysed both during the writing phase and after the conclusion of the school year. While the former was an explorative phase, aimed at practicing a continuous assessment of the pupils' progressive development towards the learning goals, the latter was carried out by applying a rigorous procedure of qualitative content analysis (QCA).

The findings reflected both the advantages and the disadvantages mentioned in the literature about conducting a research through the use of a diary method.

Issues were identified and discussed in relation to integrating a journaling experience into the busy schedule of primary school teachers, who need to fulfil many institutional commitments such as meetings with their colleagues or conferences with parents.

Overall, this study was a precious tool for the continuous and final assessment of the pupils progresses and difficulties in learning English as a foreign language. In conclusion, hopefully this research will encourage other FL teachers and educators to re-examine their practices and embark on a reflective journey that will let them be the best version of themselves, for the benefit of their pupils.

REFERENCES

- Alaszewski, A. (2006). *Using diaries for social research*. London: Sage Publishers. <https://dx.doi.org/10.4135/9780857020215>.
- Alderson, J. C., Brunfaut, T., & Harding, L. (2017). Bridging assessment and learning: a view from second and foreign language assessment. *Assessment in Education: Principles, Policy & Practice*, 24(3), 379-387. <https://doi.org/10.1080/0969594X.2017.1331201>.
- Allwright, D., & Bailey, K. M. (1991). *Focus on the language classroom: An introduction to classroom research for language teachers*. Cambridge University Press.
- Altrichter, H., Feldman, A., Posch, P., & Somekh, B. (2005). *Teachers investigate their work: An introduction to action research* (2nd Edition). Routledge.
- Assessment Reform Group (1999). *Assessment for Learning: Beyond the Black Box*. Cambridge: School of Education, Cambridge University.
- Assessment Reform Group (2002). *Assessment for Learning: 10 Principles. Research-based principles to guide classroom practice*. Retrieved June 23, 2021 from: https://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSE/Eng_DVD/doc/Afl_principles.pdf.
- Bailey, K. M. (1990). The use of diary studies in teacher education programmes, in Richards J. C., & Nunan D. (eds.) *Second Language Teacher Education*, Cambridge University Press, 215-226.
- Bailey, K. M. (1991). Diary studies of classroom language learning: The doubting game and the believing game. In Sadtona, E. (Ed.), *Language acquisition and the second/foreign language classroom* (Anthology Series 28) (pp. 60– 102). Singapore: SEAMEO Regional Language Center.
- Bailey, K. M., Curtis, A., & Nunan, D. (2001). *Pursuing Professional Development: the self as source*. Boston: Heinle & Heinle.
- Bailey, K., & Curtis, A. (2009). Diary Studies. *OnCue Journal*, 3 (1), 67-85.
- Bartlett, R., & Milligan, C. (2015). *What is Diary Method?*. Bloomsbury Academic. Retrieved May 10, 2021, from <http://dx.doi.org/10.5040/9781472572578>.
- Beck, C., & Kosnik, C. (2001). Reflection-in-action: In defence of thoughtful teaching. *Curriculum Inquiry*, 31(2), 217-227. <https://doi.org/10.1111/0362-6784.00193>.

- Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open*, 2, 8-14. <https://doi.org/10.1016/j.npls.2016.01.001>
- 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., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi delta kappan*, 86(1), 8-21. <https://doi.org/10.1177/003172170408600105>.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31. <http://dx.doi.org/10.1007/s11092-008-9068-5>.
- Blaikie, N., & Priest, J. (2017). *Social research: Paradigms in action*. John Wiley & Sons.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual review of psychology*, 54(1), 579-616. <http://dx.doi.org/10.1146/annurev.psych.54.101601.145030>.
- Brown, G. T. L. (2018). *Assessment of Student Achievement*. Routledge. <https://doi.org/10.4324/9781315162058>.
- Brown, G. T. L., Gebril, A., Michaelides, M. P., & Remesal, A. (2018). Assessment as an emotional practice: Emotional challenges faced by L2 teachers within assessment. In Martinez Agudo, J. (ed.) *Emotions in second language teaching*, Springer, 205-222. http://dx.doi.org/10.1007/978-3-319-75438-3_12.
- Burhan-Horasanlı, E., & Ortaçtepe, D. (2016). Reflective practice-oriented online discussions: A study on EFL teachers' reflection-on, in and for-action. *Teaching and Teacher Education*, 59, 372-382. <http://dx.doi.org/10.1016/j.tate.2016.07.002>.
- Burns, A. (2005). Action research: An evolving paradigm?. *Language teaching*, 38(2), 57-74. <http://dx.doi.org/10.1017/S0261444805002661>.
- Butler, Y. G. (2009). How do teachers observe and evaluate elementary school students' foreign language performance? A case study from South Korea. *TESOL quarterly*, 43(3), 417-444. <http://dx.doi.org/10.1002/j.1545-7249.2009.tb00243.x>.

- Butler, G. Y., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language testing*, 27(1), 5-31. <http://dx.doi.org/10.1177/0265532209346370>.
- Carless, D. (2007). Learning-oriented assessment: conceptual bases and practical implications. *Innovations in education and teaching international*, 44(1), 57-66. <http://dx.doi.org/10.1080/14703290601081332>.
- Cao, X., & Henderson, E. F. (2021). *Exploring Diary Methods in Higher Education Research: Opportunities, Choices and Challenges*. Routledge.
- Cohen, L., Manion, L., & Morrison, K. (2000). Action Research. In *Research methods in education*. Routledge. <https://doi.org/10.4324/9780203224342>.
- Collin, S., Karsenti, T., & Komis, V. (2013). Reflective practice in initial teacher training: Critiques and perspectives. *Reflective practice*, 14(1), 104-117. <http://dx.doi.org/10.1080/14623943.2012.732935>.
- Council of Europe (2000). European Language Portfolio (ELP): Principles and Guidelines. Strasbourg: Council of Europe. (Document DGIV/EDU/LANG (2000) 33)
- Council of Europe (2001). A Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge: Cambridge University Press.
- Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches* (4th Edition). London: Sage Publications.
- Dal Passo, F. (2003). Storia della Scuola Italiana. Semestrale di Studi e Ricerche di Geografia (1), 3-147.
- Day, R. R. (1990). Teacher observation in second language teacher education. In Richards, J. C., & Nunan, D. (Eds.), *Second language teacher education* (pp. 43-61). Cambridge University Press
- Dann, R. (2014). Assessment as learning: blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, 21(2), 149-166. <https://doi.org/10.1080/0969594X.2014.898128>.
- DeLuca, C., Lapointe-Mcewan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28 (3), 251-272. <http://dx.doi.org/10.1007/s11092-015-9233-6>.

- Dewey, J. (1933). *How we think. A restatement of the relation of reflective thinking to the educative process*. Boston, MA: Houghton Mifflin.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford university press.
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessments: The limited scientific evidence of the impact of formative assessments in education. *Practical Assessment, Research, and Evaluation*, 14(1), 7.
- Earl, L. M. (2003) *Assessment as Learning: Using Classroom Assessment to Maximise Student Learning*. Thousand Oaks, CA, Corwin Press.
- Earl, L. M. (2013). Assessment for learning; Assessment as learning: Changing practices means changing beliefs. In Hong Kong Education Bureau (Ed.). *Assessment and learning* (2), 1–5. Hong Kong: The Hong Kong Government Printer.
- Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: The concept of ‘diagnostic competence’. *Language testing*, 21(3), 259-283. <http://dx.doi.org/10.1191/0265532204lt284oa>.
- Endsley, M. R., (2013). Situation awareness. *The Oxford handbook of cognitive engineering*, 1, 88-108. <http://dx.doi.org/10.1093/oxfordhb/9780199757183.013.0006>.
- Engin, M. (2011). Research diary: A tool for scaffolding. *International Journal of Qualitative Methods*, 10(3), 296-306. <http://dx.doi.org/10.1177/160940691101000308>.
- European Commission (2002). Education and training in Europe: different systems, mutual aims for 2010. Luxembourg: Bureau for Official Publications of the European Union.
- Farrell, T. S. (2012). Reflecting on Reflective Practice: (Re) Visiting Dewey and Schon. *Tesol Journal*, 3(1), 7-16. <http://dx.doi.org/10.1002/tesj.10>.
- Firoozi, T., Razavipour, K., & Ahmadi, A. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing in Asia*, 9(1), 1-14. <http://dx.doi.org/10.1186/s40468-019-0078-7>.
- Freeman, D. (1989). Teacher training, development, and decision making: A model of teaching and related strategies for language teacher education. *TESOL quarterly*, 23(1), 27-45. <http://dx.doi.org/10.2307/3587506>.

- Fry, J. (1988). Diary studies in classroom SLA research: Problems and prospects. *JALT Journal*, 9(2), 158-167.
- Gardner, J. (2012). Assessment and Learning. Introduction. In Gardner, J. (Ed), *Assessment and Learning*. London: Sage Publications, 1-8.
<http://dx.doi.org/10.4135/9781446250808.n1>.
- Gattullo, F. (2000). Formative assessment in ELT primary (elementary) classrooms: an Italian case study. *Language Testing*, 17(2), 278-288.
<http://dx.doi.org/10.1177/026553220001700210>.
- Gattullo F., & Pallotti G. (2000). Foreign language teaching to young learners in Italy. In Nikolov, M., Curtain, H. (eds.) *An early start: young learners and modern languages in Europe and beyond*. European Council for Modern Languages: Graz, Austria.
- Gerstl-Pepin, C., & Patrizio, K. (2009). Learning from Dumbledore's pensieve: Metaphor as an aid in teaching reflexivity in qualitative research. *Qualitative Research*, 9(3), 299-308.
<http://dx.doi.org/10.1177/1468794109105029>.
- Gibbons, P. & Hammond, J. (2001). What is scaffolding? In Hammond J. (Ed), *Scaffolding: Teaching and learning in language and literacy education*. Primary English Teaching Association, Newtown, Australia, 13-26.
- Gläser-Zikuda, M., Hagenauer, G., & Stephan, M. (2020). The potential of qualitative content analysis for empirical educational research. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 21(1). <https://doi.org/10.17169/fqs-21.1.3443>.
- Goodson, I., & Sikes, P. J. (2001). *Life history research in educational settings: Learning from lives*. Open University Press.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning—tensions and synergies. *The Curriculum Journal*, 16(2), 207-223.
<http://dx.doi.org/10.1080/09585170500136093>.
- Hasselgren, A. (2000). The assessment of the English ability of young learners in Norwegian schools: an innovative approach. *Language testing*, 17(2), 261-277.
<http://dx.doi.org/10.1177/026553220001700209>.

- Hasselgreen, A. (2005). Assessing the language of young learners. *Language testing*, 22(3), 337-354. <http://dx.doi.org/10.1191/0265532205lt312oa>.
- Hatton, N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and teacher education*, 11(1), 33-49. [http://dx.doi.org/10.1016/0742-051X\(94\)00012-U](http://dx.doi.org/10.1016/0742-051X(94)00012-U).
- Henry, A., & Apelgren, B. M. (2008). Young learners and multilingualism: A study of learner attitudes before and after the introduction of a second foreign language to the curriculum. *System*, 36(4), 607-623. <http://dx.doi.org/10.1016/j.system.2008.03.004>.
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, 15(9), 1277-1288.
- Jeffrey, D. (2004). A teacher diary experience. *Asian EFL Journal*, 6(2), 1-19.
- Johnson, K. E., & Golombek, P. R. (2011). The transformative power of narrative in second language teacher education. *Tesol Quarterly*, 45(3), 486-509. <http://dx.doi.org/10.5054/tq.2011.256797>.
- Knight, P. T. (2002). Summative assessment in higher education: practices in disarray. *Studies in higher Education*, 27(3), 275-286. <http://dx.doi.org/10.1080/03075070220000662>.
- Lee, I., Mak, P., & Yuan, R. E. (2019). Assessment as learning in primary writing classrooms: An exploratory study. *Studies in Educational Evaluation*, 62, 72-81. <http://dx.doi.org/10.1016/j.stueduc.2019.04.012>.
- Lim, G. S. (2013). Assessing English in Europe. In Kunnan, A. J. (Ed.) *The companion to language assessment*, vol. 4. <https://doi.org/10.1002/9781118411360.wbcla100>.
- Lischetzke, T. (2014). Daily diary methodology. In Michalos, A.C. (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 1413-1419). Springer. http://dx.doi.org/10.1007/978-94-007-0753-5_657.
- Loh, C. E., & Liew, W. M. (2016). Voices from the ground: The emotional labour of English teachers' work. *Teaching and Teacher Education*, 55, 267-278. <http://dx.doi.org/10.1016/j.tate.2016.01.016>.

- Luft, J. (1961). The Johari Window: a graphic model of awareness in interpersonal relations. *Human relations training news*, 5(1), 6-7.
- Mayring, P. (2000). Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(2), Art. 20.
- Mayring P. (2015) Qualitative Content Analysis: Theoretical Background and Procedures. In: Bikner-Ahsbahs A., Knipping C., Presmeg N. (eds) Approaches to Qualitative Research in Mathematics Education. Advances in Mathematics Education. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-9181-6_13
- Marshall, B., & Drummond, M. J. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research papers in education*, 21(2), 133-149. <http://dx.doi.org/10.1080/02671520600615638>.
- McKay, P. (2006). Assessing Young Language Learners. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511733093>
- Mertens, D. M. (2014). An Introduction to Research. In *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. London: Sage Publications.
- McDonough, J., & McDonough, S. (1997). *Research methods for English language teachers*. Routledge. <http://dx.doi.org/10.4324/9781315832548>.
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational measurement: Issues and practice*, 22(4), 34-43. <http://dx.doi.org/10.1111/j.1745-3992.2003.tb00142.x>.
- Munby, H. (1989). Reflection-in-action and reflection-on-action. *Current issues in education*, 9(1), 31-42.
- Nagamine, T., Fujieda, Y., & Iida, A. (2018). The role of emotions in reflective teaching in second language classrooms: Felt sense, emotionality, and practical knowledge acquisition. In Martinez-Agudo (Ed.), *Emotions in second language teaching*, Springer, 145-163. http://dx.doi.org/10.1007/978-3-319-75438-3_9.

- Newstrom J. W., & Rubenfeld S. A. (1983). The Johari Window. A Reconceptualization. *Developments in Business Simulation & Experiential Exercises. Proceedings of the Annual ABSEL Conference*, Volume 10, 117-120.
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in education*, 14(2), 149-170. <http://dx.doi.org/10.1080/09695940701478321>.
- Nezlek, J. B. (2012). Introduction. In Nezlek, J.B., *Diary methods for Social and Personality Psychology*. London: Sage Publications. <http://dx.doi.org/10.4135/9781446287903.n1>.
- Nikolov, M. (2016). Trends, Issues, and Challenges in Assessing Young Language Learners. In Nikolov, M. (ed.) *Assessing Young Learners of English: Global and Local Perspectives*. Educational Linguistics, vol 25. Springer, Cham. https://doi.org/10.1007/978-3-319-22422-0_1.
- Nikolov, M., & Mihaljević Djigunović, J. (2011). All shades of every color: An overview of early teaching and learning of foreign languages. *Annual Review of Applied Linguistics*, 31, 95–119. <http://dx.doi.org/10.1017/S0267190511000183>.
- Numrich, C. (1996). On becoming a language teacher: Insights from diary studies. *TESOL Quarterly*, 30(1), 131-151. <http://dx.doi.org/10.2307/3587610>.
- Nunan, D. (1992). *Research methods in language learning*. Cambridge University Press.
- Piccardo, E. (2012). Multidimensionality of assessment in the Common European Framework of Reference for languages (CEFR). *OLBI Journal*, 4, 37-54. <https://doi.org/10.18192/olbiwp.v4i0.1106>
- Piccardo, E. (2019). The Common European Framework of Reference (CEFR) in language education: Past, present, and future. TIRF & Laureate International Universities. Retrieved from: <https://bit.ly/2AwDuKZ>.
- Pinter, A. (2011). *Children learning second languages*. Springer. <http://dx.doi.org/10.1057/9780230302297>.
- Pollard, A. (2002). *Reflective Teaching. Effective and Evidence-informed Professional Practice*. London: Continuum.
- Purpura, J. E. (2004). *Assessing grammar*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511733086>.

- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100(S1), 190-208. <http://dx.doi.org/10.1111/modl.12308>.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral science*, 28(1), 4-13. <http://dx.doi.org/10.1002/bs.3830280103>.
- Rea-Dickins, P. & Germaine, K. (1992). *Evaluation*. Oxford: Oxford University Press.
- Rea-Dickins, P. (2000). Assessment in early years language learning contexts. *Language Testing*, 17(2), 115–122. <http://dx.doi.org/10.1177/026553220001700201>.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language testing*, 18(4), 429-462. <http://dx.doi.org/10.1177/026553220101800407>.
- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249–258. <https://doi.org/10.1191/0265532204lt283ed>
- Richards, J. C., & Lockhart, C. (2007). *Reflective teaching in second language classrooms*. Cambridge university press.
- Roth, W. M., Lawless, D. V., & Masciotra, D. (2001). Spielraum and teaching. *Curriculum Inquiry*, 31 (2), 183-207. <https://doi.org/10.1111/0362-6784.00191>.
- Rose, H. (2020). Diaries and journals: Collecting insider perspectives in second language research. In McKinley, J. & Rose, H. (Eds.) *The Routledge handbook of research methods in applied linguistics*. Abingdon: Routledge, 348-359.
- Sá, J. (2002). Diary writing: An interpretative research method of teaching and learning. *Educational Research and Evaluation*, 8(2), 149-168.
- Sarigöz, İ. H., & Fişne, F. N. (2018). English language assessment and evaluation practices in the 4th grade classes at main stream schools. *Journal of Language and Linguistic Studies*, 14(3), 380-395.
- Schön, D.A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.
- Scriven, M. (1991). *Evaluation thesaurus*. London: Sage Publications.
- Serragiotto, G. (2016). *La valutazione degli apprendimenti linguistici*. Bonacci.
- Stiggins, R. J. (1999). Assessment, student confidence, and school success. *Phi Delta*

- Kaplan, 81 (3), 191-198.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83 (10), 758-765. <http://dx.doi.org/10.1177/003172170208301010>.
- Taras, M. (2001). The Use of Tutor Feedback and Student Self-assessment in Summative Assessment Tasks: Towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education*, 26 (6), 605-614. [10.1080/02602930120093922](https://doi.org/10.1080/02602930120093922)
- Taras, M. (2002). Using assessment for learning and learning from assessment. *Assessment & Evaluation in Higher Education*, 27(6), 501-510. <http://dx.doi.org/10.1080/0260293022000020273>.
- Taras, M. (2005). Assessment–summative and formative–some theoretical reflections. *British journal of educational studies*, 53(4), 466-478. <http://dx.doi.org/10.1111/j.1467-8527.2005.00307.x>.
- Taras, M. (2009). Summative assessment: The missing link for formative assessment. *Journal of further and Higher Education*, 33(1), 57-69. <http://dx.doi.org/10.1080/03098770802638671>.
- Timpe-Laughlin, V., & Cho, Y. (2021). Reflecting on assessing young foreign language learners. *Language Testing*, 38(3), 343-355. [10.1177/0265532221998264](https://doi.org/10.1177/0265532221998264).
- Tomasi, A. (2021). "Voti a perdere". Il tema della valutazione nel Movimento di Cooperazione Educativa e nella storia della scuola italiana. (Unpublished master's thesis, Università degli Studi di Modena e Reggio Emilia, Modena/Reggio Emilia, Italy). Retrieved from <https://morethesis.unimore.it/theses/available/etd-03202021-200821/>
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Buckingham: Open University Press.
- Turner, C. E., & Purpura, J. E. (2015). Learning-oriented assessment in second and foreign language classrooms. In Tsagari, T. & Banerjee, J. (eds.) *Handbook of second language assessment* (pp. 255-274). Boston: De Gruyter Mouton. <https://doi.org/10.1515/9781614513827>.
- van Manen, M. (1991). *The tact of teaching: The meaning of pedagogical thoughtfulness*.

- Albany, NY: SUNY Press. <https://doi.org/10.4324/9781315417134>.
- van Manen, M. (1995). On the epistemology of reflective practice. *Teachers and teaching*, 1(1), 33-50. <http://dx.doi.org/10.1080/1354060950010104>.
- Vogt, K., & Tzagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11 (4), 374-402. <http://dx.doi.org/10.1080/15434303.2014.960046>.
- Vygotsky, L.S. (1986). *Thought and language*. Cambridge, MA: MIT Press.
- Wallace, M. J. (1991). *Training foreign language teachers: A reflective approach*. Cambridge University Press.
- Wallace, M. J. (1998). *Action research for language teachers*. Cambridge University Press.
- Webb, D.C. (2004). Classroom assessment as a research context: variations on a theme of pedagogical decision making. Paper presented at the 10th Conference of International Congress of Mathematics Education, Copenhagen.
- William, D., & Thompson, M. (2008). Integrating Assessment with Learning: What Will It Take to Make It Work? In: Dwyer, C. (Ed.), *The Future of Assessment: Shaping Teaching and Learning*. Routledge: New York, NY, USA, 53-82. <http://dx.doi.org/10.4324/9781315086545-3>.
- Wilson, J. P. (2008). Reflecting-on-the-future: A chronological consideration of reflective practice. *Reflective Practice*, 9 (2), 177-184. <http://dx.doi.org/10.1080/14623940802005525>.
- Wininger, S. R. (2005). Using your tests to teach: Formative summative assessment. *Teaching of Psychology*, 32 (3), 164-166. http://dx.doi.org/10.1207/s15328023top3203_7.
- Zembylas, M. (2005). Beyond teacher cognition and teacher beliefs: The value of the ethnography of emotions in teaching. *International journal of qualitative studies in education*, 18 (4), 465-487. <https://doi.org/10.1080/09518390500137642>.
- Zeng, W., Huang, F., Yu, L., & Chen, S. (2018). Towards a learning-oriented assessment to improve students' learning—a critical review of literature. *Educational Assessment, Evaluation and Accountability*, 30 (3), 211-250. <http://dx.doi.org/10.1007/s11092-018-9281-9>.