UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Magistrale in

Lingue e Letterature Europee e Americane

Classe LM-37

Tesi di Laurea

Teachers' and students' perceptions of peer assessment and peer feedback in university English courses

Relatrice
Prof. Katherine Ackerley

Laureanda
Chiara Colombo
n° matr.1207690/LMLLA

Anno Accademico 2019/2020

Table of contents

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Katherine Ackerley, who has patiently guided and assisted me throughout the whole process of writing my dissertation. Thank you for your valuable advice and contributions, which have made this important journey all the more pleasant.

Many thanks to my co-supervisor, Professor Fiona Dalziel, who has taken the time to meet me on-line to discuss the topic of my dissertation. Thank you for your insights, which have proved very helpful and relevant, especially in the writing of Chapter Four.

Devo ringraziare anche la mia famiglia: i miei genitori, Francesca e Giovanni, che mi hanno sostenuta economicamente nei miei studi e che mi hanno sempre dimostrato sostegno ed entusiasmo durante tutto il mio percorso universitario e i miei quattro fratelli, Stefano, Daniele, Giacomo e Michele che, nonostante la lontananza fisica, sono sempre presenti nel mio cuore e nei miei pensieri. Ringrazio anche i miei nonni, Paola e Bernardo, Silvia e Alberto che mi hanno fatta e che mi fanno sentire sempre speciale. Grazie anche ad Anna, mia cugina, che mi ha sempre stimata moltissimo e con la sua dolcezza mi ha incoraggiata a seguire questa traiettoria.

E per finire non posso non ringraziare Pietro, che con la sua presenza e con il suo affetto mi accompagna quotidianamente da ormai due anni, tirando fuori i miei lati migliori. Sei stata una bellissima sorpresa. Spero di saperti rendere sempre orgoglioso.

Introduction

Conventional assessment methods, first and foremost teacher-assessment, have recently started to welcome new forms of assessment. These so-called 'non-traditional/alternative assessments' are meant to face new demands of education, such as the mastering of the communicative competence and the accomplishment of authentic tasks. In fact, alternative assessment is also known under the name of 'authentic assessment' (Huerta-Macías 2002).

Among these alternative assessments we can find, for instance, portfolios, learner diaries, rubrics, self-assessment and peer assessment. Some scholars claim that alternative forms of assessment hold "ethical potential" (Lynch 2001: 228 in Brown 2004) in the promotion of fairness and objectivity and have welcomed these new assessment procedures with great enthusiasm and commitment, claiming that these can also foster learner autonomy. Others, on the contrary, are firmly against introducing such assessment methods as ordinary classroom practices for a number of reasons.

Especially peer assessment has received much attention in recent years due to a growing interest in the promotion of learner independence and autonomy. Indeed, whereas in the past it was believed that valuable and significant learning experiences depended solely on the teachers, now a part of the responsibilities belongs to the learners, who are asked to be the masters of their own learning. Indeed, peer assessment seems to involve a changing role for students, "one in which they are brought into the heart of teaching and learning processes and decision-making" (James and Pedder 2006: 28 in White 2009).

The focus of my dissertation will be on peer assessment. More specifically, I will seek to find out teachers' and students' perceptions of peer assessment and peer feedback, and especially the opinions of second-year master student of languages who are aspiring schoolteachers. Since peer assessment can be an exceptional pedagogical tool to actively engage students in their learning process and to promote learner autonomy, I am interested to find out whether aspiring schoolteachers would ever propose a peer assessment activity to their future students.

This paper will be structured as follows: Chapter One will be the literature review, in which I will introduce important terminology, such as the difference between

assessment, *feedback* and *testing*, and between *formative* and *summative assessment*, the three fundamentals of testing, namely *validity, reliability,* and *objectivity* and the concept of *learner autonomy*.

Assessment and testing are often used as synonyms, although they carry slightly different meanings: assessment is a "systematic process" (APA Dictionary of Psychology) whereby important information about a student's learning process is acquired, whereas a test is a "formal procedure" (Brown, 2004: 251), which provides information about the learner's performance and learning process at a given moment in time. Instead, while formative assessment is integrated with the teaching and, thus, can guide future teaching and help teachers to form future plans, summative assessment is an assessment that occurs after a course has ended, thus not really helping teachers to improve their teaching. As to validity, reliability, and objectivity, they are respectively defined as a criterion which refers to the extent to which a test measures what it purports to measure, a criterion which refers to the consistency of a test, and the degree to which a test (or an assessment) is unbiased, neutral, impartial and independent of external influence. Last but not least, learner autonomy has been defined as the "ability to take charge of one's own learning" (Holec 1981: 3 in Novello 2014: 20). With this definition in mind, peer assessment and self-assessment seem to be exceptional methods to foster learner autonomy. Indeed, self-assessment encourages students' self-reflection and self-improvement. Instead, peer assessment can also raise students' awareness, because by judging the work of others, students are likely to become more aware of their own skills.

Chapter One will also contain a brief overview of the wide variety of alternative or non-traditional assessment available, namely portfolios, learner diaries, and self-assessment. Portfolios are collections of written samples of language (or, spoken samples of language in the case of e-portfolios) produced by students throughout a course. They are extremely useful, especially in that they can show how much a student has progressed. Learner diaries (or, journals) consist in entries written by students, with which they can record their learning goals and the strategies used to achieve those goals. Diary writing is an important pedagogical tool, in that it encourages self-reflection and autonomous learning. Lastly, self-assessment, as its name suggests, is a "self-administered assessment" (Garner, 1999: 50). Self-assessment is an exceptional tool to raise students' language awareness and self-reflection, in that they are encouraged to identify both their

strengths and their weaknesses (i.e. those areas of language that will need particular improvement).

Chapter Two will be devoted entirely to peer assessment, its advantages and disadvantages and some of the possible solutions to the problems that peer assessment entails, which have been identified by different scholars. More specifically, I will examine peer assessment in conjunction with the key issues which are often raised as objections toward this evaluating procedure (i.e. validity, reliability and objectivity), as well as the potential advantages that this assessment method could have on the students and the learning process.

Peer assessment is a procedure, through which students assess their classmates' written or spoken performance and provide constructive feedback on the performance. It can entail students providing a grade to each other (peer grading/peer assessment), or students simply provide feedback to their peers (peer review/peer feedback). Peer assessment also has positive repercussions on self-assessment, since students, by assessing the work of others, are expected to be able to identify their own weaknesses as well. For the purpose of my dissertation, I will use the term 'peer assessment' to also include peer feedback.

Then, eight studies on peer assessment will also be reviewed in Chapter Two, with an explanation as to how they have served as models for my own study of peer assessment presented in the following chapter. These studies took different skills into consideration, namely writing skills (the studies by Mendonça and Johnson 1994; Azarnoosh 2013; Jung 2016), and speaking skills (White 2009; and Peng 2010). Furthermore, two meta-analysis have also been reviewed, namely Falchinov and Goldfinch (2000) and Double et al. (2019).

Chapter Three will be devoted to the discussion of questionnaire results about peer assessment, which was distributed among a cohort of second-year master students of languages, and to the analysis of my own study of peer assessment with two first-year bachelor students of languages. Both groups of students have been investigated in different ways (i.e. through a questionnaire, and through a peer assessment activity on writing skills) to identify differences and similarities in their perceptions of peer assessment. More specifically, my own study of peer assessment was partly based on some of the studies reviewed in Chapter Two, namely Mendonça and Johnson (1994) for

the investigation of peer feedback interactions, Peng (2010) for the pre-activity and post-activity questionnaires, and Azarnoosh (2013) for the investigation of perceived friendship biases during peer assessment.

Last but not least, Chapter Four will be a discussion of questionnaire findings about peer assessment and peer feedback, which was distributed among a group of university teachers of English (both professors and CELs). This chapter will provide an insight into teachers' perceptions of peer assessment based on their personal experience, and especially discuss teachers' opinions about the benefits of peer assessment.

Chapter 1: Assessment

"A teacher is one who makes himself progressively unnecessary." — T. Carruthers

This chapter provides an overview of school assessment, its various facets and its main features. The traditional teacher assessment in school seems to have a negative connotation between students, since many students and their parents only see it as a grading procedure with an end in itself. However, assessment entails much more than just grading. In fact, as I will try to demonstrate with this dissertation, grading is just the final [and minor] step of the whole process.

According to Novello (2014: 9)[1], the ultimate goal of school assessment should be to

> "open a channel of communication between teachers and learners, so that the latter can express their concerns, doubts and needs to their teachers. On the contrary, teachers will receive important information on the degree of satisfaction of their students with the evaluation procedure and with the course overall, thus enabling them to make significant changes and adjustments to their evaluation methods, should these be needed."

To begin with, it seems important to devote the initial part of this chapter to the clarification of specific terminology, which will recur frequently throughout this dissertation. The differences between 'test', 'assessment' and 'feedback', formative and summative assessment, holistic and analytic scoring will be provided. Then, the concept of 'alternative assessment' will be clarified, and some of the key issues that are often raised against alternative assessment practices, namely validity, reliability, and objectivity, will be pointed out. To follow, an overview of the various alternatives in assessment will be offered. Last but not least, the elucidation of some of the theoretical foundations of language acquisition, namely Krashen's *Second Language Acquisition Theory* (SLAT), also known as 'natural approach', Chomsky's *Language Acquisition Device* (LAD), Bruner's *Language Acquisition Support System* (LASS) and his notion of 'scaffolding', and Vygotsky's *Zone of Proximal Development* (ZPD) will be given. These theories are essential to foreign language teaching and assessment, since they emphasize

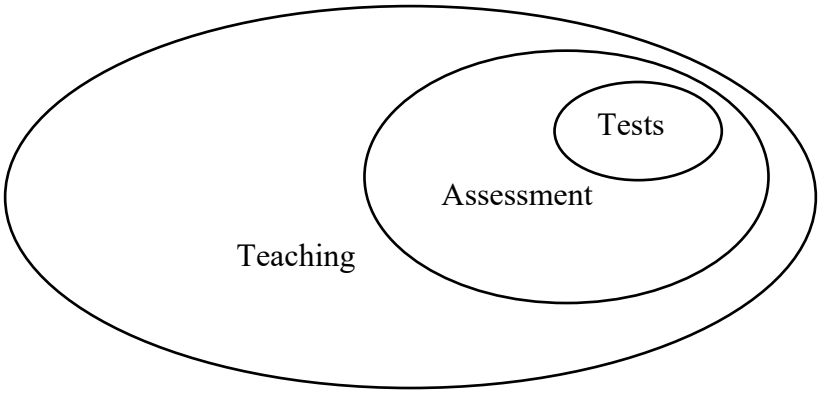
[1] All citations from Novello were translated from Italian by me for the purpose of this dissertation. Only in a few cases did I decide to also leave the original citation in Italian in footnote.

the link between evaluation and teaching. Indeed, arguably, knowing what and how to teach, means knowing what and how to assess. Therefore, it seems of primary importance that teachers are made more aware of how students learn foreign languages to help them to plan their courses and, more importantly, to achieve an effective classroom assessment.

1.1 Test, Assessment and Feedback

According to Brown (2004), the relationship between teaching, assessment and tests can be represented through a diagram, which I report below.

Figure 1: Diagram of the relationship among testing, assessment and teaching (Brown 2004).



As shown in *Figure 1*, teaching includes assessment, which, in turn, includes tests. Or, it could be also said that tests are embedded within assessment, which is embedded within teaching.

First of all, the notion of 'testing' should be clarified and, more specifically, the distinction between 'test' and 'assessment' should be pointed out. Very often, in fact, these two terms are inappropriately used as synonyms, even though they represent two separate things.

1.1.1 Testing: Features and Typologies

Testing has been defined by Brown (2004: 251) as "a formal procedure, usually administered within strict time limitations, to sample the performance of a test-taker in a specified domain". More importantly, tests provide information about the learner's

performance and learning process at a given moment in time: teachers will not be able to judge what a learner can actually do (in terms of language mastery) with just a single test.

In this respect, Brown (2004: 117) argues that "one important principle for assessing a learner's competence is to consider the fallibility of the results of a single performance, such as that produced in a test". Indeed, as will become clear later in this dissertation, many factors can affect the performance of learners.

Novello (2014: 37-38) expands the definition provided by Brown by stating that tests involve a "comparison between the results of the performance and the pre-determined objectives of the school curriculum". These objectives can be specifically chosen by the teachers (and sometimes agreed on with the students), but they must also respect specific curriculum requirements.

Novello (2014) provides a brief overview of the different types of language tests available. Among these, she mentions aptitude tests, which aim to make predictions about the ability of students to learn a new language, diagnostic tests, which are "designed to diagnose specified aspects of a language" (Brown, 2004: 46), progress tests, which are meant to monitor the learning process (i.e. during a course), achievement tests, whose objective is to verify the achievement of the objectives in the curriculum and which may come at the end of a course, or during a course, and proficiency tests that show what students can do with a language (often related to the 'can do' statements from the CEFR[2]). More specifically, Brown (2004: 44) claims that proficiency tests are almost always summative (summative assessment will be dealt with in *§ 1.2*). Moreover, there is another type of test, which is often adopted in the Italian education system: the entry test. This type of test, however, is often misused by teachers, in that they mark it as if it was an 'ordinary' test. However, entry tests should only help the teacher to plan the course, based on the level that his/her students reach, according to the test. Therefore, entry tests should not be assigned grades.

In the context of proficiency tests, whose aim is to verify what learners can do with a certain language, it seems important to make a distinction between what Balboni (2015) calls *sapere la lingua* (knowing the language), *saper fare lingua* (knowing how to "make language"), and *saper fare con la lingua* (knowing "what to make" with a

[2] Common European Framework of Reference for Languages.

language). According to Balboni, knowing a language means knowing and mastering linguistic competence (i.e. phonology, syntax, morphology, lexical and textual aspects of the target language), extra-linguistic competence, such as kinesics (body language), proxemics (physical/interpersonal distance), chronemics (the role of time in a society) and socio-pragmatic competence. Instead, knowing how to "make language" means mastering those integrated skills within the target language, such as translating, taking notes, paraphrasing and summarizing. Last but not least, knowing "what to make" with a language means knowing the social, pragmatic and cultural rules that govern every communicative event.

Novello (2012) offers a model for teachers when it comes to designing and writing a test. According to her, four criteria should be met: first, teachers need to share criteria for evaluation with their students (*condivisione dei criteri di valutazione*) before the students take the test. Then, students need to be trained to take tests (*preparazione alle prove*). Next, familiarity with tests should also be taken into consideration (*familiarità con le prove di verifica*). This particular aspect is also mentioned by Hughes (2003: 47-48): "every effort must be made to ensure that all candidates have the opportunity to learn just what will be required of them". Last, students need to have a chance to make up for the tests that did not go well (*possibilità di recupero delle verifiche con esito negativo*). This chance to make up for unsuccessful tests can be ensured, for instance, also through different testing methods, which suit different students. This last criterium also links in with what Gipps (1994 in Novello 2012) calls "equity", or "fairness": in accordance with the principle of fairness, indeed, students should be granted enough opportunities to prove their linguistic abilities through different testing methods.

Put more simply, tests need to be appropriate, fair, and accurately scored. Marking accuracy and consistency will be addressed in *§1.3* with rating scales. Moreover, it is good practice to prepare students to a test to achieve valid and reliable results. The issues of validity and reliability will be the focus of *§1.5*.

Having clarified these points, it seems important to emphasize that a test is, therefore, only a small fraction of the whole assessment process. Indeed, assessment "connotes a much broader concept in that most of the time when teachers are teaching, they are also assessing" (Brown, 2004: 252).

1.1.2 Assessment

Several definitions of assessment have been provided. According to the *APA Dictionary of Psychology*, assessment can be considered a "systematic process of obtaining information from participants and using it to make inferences or judgments about them". The term systematic is especially relevant, since it indicates that assessment should not be carried out without a system, a method, but rather it is a thorough and detailed evaluation, which should be based on very precise criteria. Clear and precise criteria, indeed, help teachers to give focused and specific feedback that will actually be helpful to the learner.

Brown (2004: 4) adds to this definition that assessment is an "ongoing process", through which important information about the learners can be gained, such as the students' preferred learning modes, their difficulties and their needs. Similarly, Peñaflorida (2002: 346) also argues that assessment "involves the means of obtaining information about students' abilities, knowledge, understanding, attainments, or attitudes". Last but not least, Weiss (1972 in Dam and Legenhausen 2010: 120) defines assessment as "the systematic gathering of information for the purpose of making decisions". Put more simply, Dam and Legenhausen (2010: 121) explain that assessment implies that both "learners and teachers reflect on the experiences gained in language learning and teaching, which will lead to awareness raising and prepare the ground for decision making".

Assessment can either be informal or formal: informal assessment can take numerous forms, such as "incidental, unplanned comments and responses, coaching and other impromptu feedback to the student" (Brown 2004: 5). On the contrary, formal assessment is systematic and, more importantly, carefully planned. Indeed, many scholars and experts emphasize that assessment criteria must be carefully decided upon beforehand and shared with the learners:

> "assessing without having pre-determined criteria leads to an unjustified evaluation. In addition, sharing the criteria for assessment with learners ensures transparency of assessment, thus also enhancing a feeling of trustworthiness toward the teacher on the part of the learner" (Novello 2012: 39).

The information which teachers acquire through assessing their students should not be overlooked, but rather qualitative and/or quantitative value should be assigned to

it (Novello 2014: 37). This quantitative and qualitative value is known under the name 'feedback', which will be discussed in the next section (*§1.1.3*).

In the context of assessment, it seems important to briefly mention the concept of 'learning style'. Learning styles differ across individuals and could be defined as "cognitive, affective, and physiological traits that are relatively stable indicators of how learners perceive, interact with, and respond to the learning environment" (Keefe, 1979: 4 in Brown, 2000: 114). Given the variations in learning styles, it is essential that teachers differentiate their teaching as well as their assessment methods as much as possible to ensure an optimal "soil" for Meaningful Learning[3].

It is well-known that mixed-ability classes, that is classes in which students have different learning styles and, more importantly, different paces of learning, entail numerous issues. To name a few, if teachers do not individualize their teaching and their assessment method to suit each student, then students might feel unmotivated, unengaged and uninterested, because they might feel that their teacher is unfair and biased. As suggested by Favrot[4] (2004 in Caon and Tonioli, 2016: 139), teachers should always try to differentiate as much as possible the objectives and contents of the course, their teaching resources, and, more importantly, the criteria for assessing their students. If all these measures are taken, students will greatly benefit from them, and meaningful and significant learning experiences are more likely to occur.

Novello (2012) mentions another issue, which is deeply involved in assessment: that of motivating learners towards evaluation. It is well-known that motivation plays an essential role in the learning process, and more so in the language learning process, since language learning requires special effort, dedication, practice and perseverance. Likewise, Maslovaty and Kuzy (2002: 201) also argue that "student's motivational goals are significant components of the learning process". Having said this, motivation is also closely linked with assessment. Indeed, if learners perceive that they can improve their

[3] According to Novak (1984), Meaningful Learning (*apprendimento significativo*) occurs when learners establish a connection between new information/knowledge and pre-existing information/knowledge. Its opposite counterpart is Rote Learning (*apprendimento meccanico*), which occurs when students memorize new information without linking it to pre-existing knowledge.

[4] "Favrot (2004) […] indica come grazie a questa differenziazione degli obiettivi e dei contenuti, emerge così la possibilità per ogni studente di procedere nell'apprendimento in base al proprio ritmo; gli studenti più in difficoltà avranno la possibilità di avanzare in maniera più lenta approfondendo alcuni aspetti mentre gli studenti eccellenti potranno arricchire le proprie conoscenze continuando a progredire nell'acquisizione di nuovi contenuti."

learning thanks to assessment, then they will be more willing to undergo their teacher's (or, their peers') judgements and listen to the suggestions that these might give to them.

From a cognitive perspective, motivation has to do with an individual's decisions, the "choices people make as to what experiences or goals they will approach or avoid, and the degree of effort they will exert in that respect" (Keller 1983: 389 in Brown 2000: 160). Novello (2012: 96) notes the importance of making students understand the "functionality of assessment", thus making them more motivated toward it and, overall, more engaged in their learning. As Novello (2012) argues, students will be more likely to trust their teacher, in that they will perceive that he/she wants to ensure that they experience successful learning.

According to Balboni (1994 in Novello, 2012: 92) there are three kinds of motivation for learning a new language: duty, need and pleasure. It goes without saying that motivation related to duty, that is imposed from the outside, is the weakest among the three: "those who pursue a goal only to receive an external award from someone else are extrinsically motivated to succeed in a task" (Brown 2000: 162). In this respect, it is important to emphasize that grades are also among the typical extrinsic rewards. Then, motivation related to need is also weak, since it only lasts until our need is fulfilled. An example of motivation related to a need would be, for instance, a specific language certification for a job: if an employee wants a job that requires a language certification, then he/she will be motivated to obtain this certification, but he/she will probably not carry on studying the language, once his/her need (i.e. obtaining the language certification) has been fulfilled. Instead, motivation related to pleasure (i.e. driven by curiosity, passion, interest) is the strongest, because it can regenerate itself. In this respect, Brown (2004: 270) claims that "developing intrinsic motivation that comes from a self-propelled desire to excel is at the top of the list of successful acquisition of any set of skills".

Balboni (2006 in Novello, 2012: 92) further suggests the importance of supporting the student's pleasure of learning (*piacere di apprendere*), pleasure of exploring a variety of materials, classroom activities, and pedagogical tools (*piacere della varietà*), pleasure of challenges (*piacere della sfida*), and pleasure of considering oneself an 'autonomous learner' (*piacere di pensarsi autonomi*), all of which contribute to more positive language learning experiences.

1.1.3 Feedback: Quantitative and Qualitative

According to Hattie (2012: 129), the aim of feedback is to "reduce the gap between where the student 'is' and where he or she is 'meant to be'", meaning that teachers must have a good understanding of where their students are (in terms of their learning). In his view, feedback addresses three very important questions: where am I going? how am I going there? and where to next? The first question regards the learning intentions: "teachers need to know and communicate to their students the goals of the lesson" (Hattie 2012: 131). The second question is concerned with the individual steps that students need to take in order to achieve a goal. Last, feedback answering the third question should "assist students in choosing the most appropriate challenges and developing more self-regulation over the learning process" (Hattie 2012: 132). The scholar then argues that feedback is very powerful, but it needs to be optimized:

> "for feedback to be received and have a positive effect, we need transparent and challenging goals (learning intentions), an understanding of current status relative to these goals (knowledge of prior achievement), transparent and understood criteria of success, and commitment and skills by both teachers and students in investing and implementing strategies and understandings relative to these goals and success criteria" (Hattie 2012: 151).

Moreover, feedback should be as detailed and as tailored as possible. Indeed, feedback that is addressed to the whole class is unlikely to be effective, because students will not understand that the comment pertains to them: "the aim is to provide feedback that is 'just in time', 'just for me', 'just for where I am in my learning process', and 'just what I need to help me move forward'" (Hattie 2012: 137).

As briefly mentioned above, assessment results in quantitative or qualitative feedback. These two types of feedback are both very important, but they play different roles. According to Novello (2012: 101), qualitative feedback is the best solution both for teachers and students, because through qualitative feedback students "can become more aware of the critical aspects of their performances, and teachers can have a better picture of the actual abilities of the learners."

Likewise, Cameron (2001 in Novello 2014: 56) maintains that, through qualitative feedback, learners "understand the target performance, compare target and current performance, and close the gap between target and current performance". Cameron (2001 in Novello 2014) indicates three types of qualitative feedback: corrective feedback,

evaluative feedback, and strategic feedback. The former simply provides the correct input. The second "includes a judgement about the performance" (Novello, 2014: 56). Instead, the latter offers helpful suggestions for future performances.

On the contrary, quantitative feedback consists in assigning grades (or, points) to a performance. This procedure, however, is less useful for teachers and, especially, for learners, since it does not inform the teacher of what the learner can actually do, nor does it help the teacher provide beneficial and useful advice to his/her learners on how to improve their performance.

Evidently, a combination of quantitative and qualitative feedback seems to be the best solution: teachers can assign a grade to their students' performance, but they can also offer a large amount of useful information that justifies those grades.

In this respect, it seems important to stress that not only teachers, but also students can give feedback to their peers; in this case, we are dealing with peer feedback (which will be addressed in Chapter Two). Furthermore, students can also be responsible for their own assessment; in this case, we are dealing with self-assessment (*§1.4.6*).

Keller and Westfall-Greiter (2014) provide helpful advice on how to give effective feedback and emphasize that feedback should focus different aspects of a performance, namely: on content, design, language and mechanics. Hattie (2012) also argues that feedback works at four levels, namely: task (also known as 'corrective feedback' because it is information-focused, such as correct/incorrect), process, self-regulation (i.e. encouraging students to monitor and self-regulate their learning), and self (feedback at this level usually includes praise).

As stated at the beginning of this section, according to Hattie (2012), feedback helps to reduce the gap between the actual learning stage of students and where students should be with their learning. Hattie (2012) then mentions various ways, in which the feedback provided can help to reduce this gap, namely: increase motivation in students, direct the student's attention towards the processes that can help him or her achieve a task, clarify possible misunderstandings of task instructions, and provide cues that help students succeed in the task. Hattie (2012: 130) also emphasizes that, although feedback

> "thrives on error, […] acknowledging errors allows for opportunities. Error is the difference between what we know and can do, and what we aim to know and do. […] Knowing this error is fundamental to moving towards success. This is the purpose of feedback."

1.1.4 Errors vs. Mistakes

In this regard, it seems important to make another distinction between what error analysis, a branch of applied linguistics which studies and analyses errors made by second language learners, defines as 'error' and 'mistake' (it. 'errore' vs. 'sbaglio'). As Richards and Schmidt (2002: 199) claim, errors consist in:

> "[…] the use of a linguistic item (e.g. a word, a grammatical item, a speech act, etc.) in a way which a fluent or native speaker of the language regards as showing faulty or incomplete learning."

Therefore, errors seem to result from incomplete or partial knowledge and/or from a lack of competence. Instead, the two scholars regard 'mistakes' as simply "caused by lack of attention, fatigue, carelessness, or some other aspect of *performance*" (2002: 199). Mistakes are merely "failures to utilize a known system correctly" (Brown 2000: 217). Therefore, mistakes do not necessarily need to be corrected and learners should not be too penalized for making them: in fact, very often learners can detect their own mistakes and correct themselves (almost) immediately. On the contrary, errors need prompt corrective actions on the part of the teacher so as not to become fossilized. However, as Novello (2012: 105) points out, teachers should "welcome errors, using them as opportunities for linguistic reflections". In fact, "in some cases, the valorisation of errors proves very useful, since through them a new 'linguistic mechanism' has been discovered" (106). In this respect, Brown (2000: 218) claims that "errors indeed reveal a system at work". This is, for instance, the case of exceptions: *I go*, *you go*, *he goes*, but *I can*, *you can*, *he can* (instead of he 'cans'), or *one cat*, *two cats*, but *one sheep*, *two sheep* (instead of two 'sheeps').

However, having said this, as Novello (2012) further explains, "errors can also reveal objectives that have not yet been met by the learner, as well as fossilizations and difficulties" (106). In this case, corrective actions are needed to cover the lacking competence.

One question may now arise. How can one acknowledge whether an incorrect term, form or sentence is simply a (distraction) mistake, rather than an error? According to Balboni (2015), frequency is one of the most accurate indicators to detect whether such deviations should be considered errors or mistakes: by observing their frequency of recurrence, teachers can decide whether they are mistakes or errors and, thus, carry out corrective actions.

Errors have been thoroughly studied in the field of foreign and second language learning "to discover the processes learners make use of in learning and using a language" (Richards and Schmidt, 2002: 199). The two scholars classify errors into *lexical error* (vocabulary), *phonological error* (pronunciation), *syntactic error* (grammar), *interpretive error* (misunderstanding of a speaker's intention or meaning), and *pragmatic error*. Learners of foreign languages (and cultures) must pay attention to all the above-mentioned error categories, but especially to the last one. Indeed, this latter consists in the "production of the wrong communicative effect, e.g. through the faulty use of a speech act or one of the rules of speaking" (Richards and Schmidt, 2002: 199) and, therefore, it can cause the greatest harm to a communicative event, since the listener might be offended.

1.2 Formative and Summative Assessment

A further distinction is now required between formative assessment and summative assessment.

The former is a procedure that takes place continuously throughout a course, through which teachers monitor the development of the learning process and obtain useful information on the progress of their learners, their weaknesses, and the effectiveness (or, ineffectiveness) of the teaching and assessment methods. According to Çelik and Türkan (2014: 421) formative assessment is often employed by teachers "in order to make decisions about the direction of their teaching". By monitoring their students' learning process, indeed, teachers can adjust their teaching and make substantial changes to suit their learners' needs while the learning is still taking place. Therefore, formative assessment "accompanies the educational project and provides frequent data for its improvement" (Novello, 2012: 103). Brown (2004: 267) includes observation and teacher-student conference among formative assessment practices, whose use "greatly enhances teacher's intuitive impressions by offering a tangible corroboration of conclusions". Indeed, it appears to be very important to create a classroom atmosphere, in which teachers and students can freely exchange opinions, doubts and suggestions with each other. Observation will be dealt with more in depth in *§1.4.3*.

On the contrary, summative assessment, as the term denotes, "aims to measure, or summarize, what a student has grasped, and typically occurs at the end of a course or unit

of instruction" (Brown, 2004: 6). It usually entails a process, whereby a final grade is decided on a combination of assignments or a final test. Summative assessment is, therefore, generally an "assessment of profit rather than competence" (Novello, 2014: 39), and does not provide useful information to the teacher.

1.3 Holistic and Analytic Scoring: Rating Scales

When it comes to rating students' performances, using rating (or, evaluation) scales may come in handy. These consist in items or descriptors, to which scores are assigned, which describe potentially every aspect, or the overall impression, of the student language ability. There are two main approaches to scoring; one is holistic (or, impressionistic) scoring, and the other is analytic scoring.

The former consists in the assignment of a single score to a performance (i.e. a piece of writing, an oral test…) based on "an overall impression of it" (Hughes, 2003: 94-95). Evidently, this approach is very rapid and, if executed properly, that is scored by more than one scorer, it can result in high scorer (or, interrater) reliability. Scorer (or, rater) reliability will be addressed further in *§1.5* In this respect, Hughes (2003: 95) claims that high scorer reliability can be obtained especially

> "if each student's work is scored by four different trained scorers. […] There is nothing magical about the number 'four'; it is simply that research has quite consistently shown acceptably high scorer reliability when writing is scored four times".

As maintained by Hughes (2003: 95), holistic scoring needs to be well-conceived and well-organized. Indeed, "not every scoring system will give equally valid and reliable results […]. The system needs to be appropriate to the level of the candidates and the purpose of the test". Hughes (2003: 96), therefore suggests that teachers decide which descriptors of an already existing holistic scale to use, depending on the purpose of the assessment: "testers have to be prepared to modify existing scales to suit their purposes". Indeed, these tend to be very generic and imprecise, since they are conceived as a "one-size-fit-all" kind of scale, but sometimes they can also become too detailed. For instance, Hughes (2003) mentions the *ACTFL* (American Council for the Teaching of Foreign

Languages) proficiency guidelines for writing, which consist of ten different levels[5] (from Novice-Low to Superior), in which foreign language learners can be placed according to their written performance. As shown in *Table 1* below, all the levels are described very thoroughly. However, it is highly questionable that every student at a certain level has the same "amount" of abilities: "the descriptors imply a pattern of development common to all language learners. They assume that a particular level of grammatical ability will always be associated with a particular level of lexical ability" (Hughes, 2003: 100). This, however, is highly improbable.

Table 2: Descriptors for writing, level 'Advanced'. Taken from the ACTFL website[6].

Advanced Writers at the Advanced level are characterized by the ability to write routine informal and some formal correspondence, as well as narratives, descriptions, and summaries of a factual nature. They can narrate and describe in the major time frames of past, present, and future, using paraphrasing and elaboration to provide clarity. Advanced-level writers produce connected discourse of paragraph length and structure. At this level, writers show good control of the most frequently used structures and generic vocabulary, allowing them to be understood by those unaccustomed to the writing of non-natives.
Advanced High Writers at the Advanced High sublevel are able to write about a variety of topics with significant precision and detail. They can handle informal and formal correspondence according to appropriate conventions. They can write summaries and reports of a factual nature. They can also write extensively about topics relating to particular interests and special areas of competence, although their writing tends to emphasize the concrete aspects of such topics. Advanced High writers can narrate and describe in the major time frames, with solid control of aspect. In addition, they are able to demonstrate the ability to handle writing tasks associated with the Superior level, such as developing arguments and constructing hypotheses, but are not able to do this all of the time; they cannot produce Superior level writing consistently across a variety of topics treated abstractly or generally. They have good control of a range of grammatical structures and a fairly wide general vocabulary. When writing at the Advanced level, they often show remarkable ease of expression, but under the demands of Superior-level writing tasks, patterns of error appear. The linguistic limitations of Advanced High writing may occasionally distract the native reader from the message.

[5] The updated guideline for writing (2012) consists of eleven levels: Distinguished, Superior, Advanced, Intermediate, and Novice. The major levels Advanced, Intermediate, and Novice are divided into High, Mid, and Low sublevels (taken from the *ACTFL* website: https://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012/english/writing).

[6] American Council for the Teaching of Foreign Languages. DOI: https://www.actfl.org/sites/default/files/pdfs/ACTFLPerformance_Descriptors.pdf

Advanced Mid
Writers at the Advanced Mid sublevel are able to meet a range of work and/or academic writing needs. They demonstrate the ability to narrate and describe with detail in all major time frames with good control of aspect. They are able to write straightforward summaries on topics of general interest. Their writing exhibits a variety of cohesive devices in texts up to several paragraphs in length. There is good control of the most frequently used target-language syntactic structures and a range of general vocabulary. Most often, thoughts are expressed clearly and supported by some elaboration. This writing incorporates organizational features both of the target language and the writer's first language and may at times resemble oral discourse. Writing at the Advanced Mid sublevel is understood readily by natives not used to the writing of non-natives. When called on to perform functions or to treat issues at the Superior level, Advanced-Mid writers will manifest a decline in the quality and/or quantity of their writing.
Advanced Low
Writers at the Advanced Low sublevel are able to meet basic work and/or academic writing needs. They demonstrate the ability to narrate and describe in major time frames with some control of aspect. They are able to compose simple summaries on familiar topics. Advanced Low writers are able to combine and link sentences into texts of paragraph length and structure. Their writing, while adequate to satisfy the criteria of the Advanced level, may not be substantive. Writers at the Advanced Low sublevel demonstrate the ability to incorporate a limited number of cohesive devices and may resort to some redundancy and awkward repetition. They rely on patterns of oral discourse and the writing style of their first language. These writers demonstrate minimal control of common structures and vocabulary associated with the Advanced level. Their writing is understood by natives not accustomed to the writing of non-natives, although some additional effort may be required in the reading of the text. When attempting to perform functions at the Superior level, their writing will deteriorate significantly.

Moreover, as Hughes (2003) points out, another issue may arise when we are trying to rate someone based, for instance, on the above-mentioned scale: what should teachers do when the learner fits partly in one level and partly in another level (i.e. if the descriptors that are used to describe the learner's language abilities belong partly to one level and partly to another level)? According to Hughes (2003: 100), this decision "must depend in part on the purpose of the assessment". Holistic scales require extensively trained raters, since "one score masks differences across the subskills within each score" (Brown, 2004: 242).

Let us now shift our focus on analytic scoring and analytic scales. By the term 'analytic scoring' a method of scoring which entails a separate score for each aspect of a task is meant. As suggested by Harris (1968 in Hughes 2003), the different aspects to be taken into consideration when rating oral skills are, for example, grammar, vocabulary,

mechanics (structure), fluency and form. The total score will then be given by the sum of all the scores relating to each of the above-mentioned aspects. In this particular scale by Harris, every dimension allows for from one to six points, one point being the worst and six points being the best (see *Table 2* below).

Table 2: Part of the analytic scale designed by Harris (1968 in Hughes 2003).

Grammar 6 Few (if any) noticeable errors of grammar or word order. 5 Some errors of grammar or word order which do not, however, interfere with comprehension. 4 Errors of grammar or word order fairly frequent; occasional re-reading necessary for full comprehension. 3 Errors of grammar or word order frequent; efforts of interpretation sometimes required on reader's part. 2 Errors of grammar or word order very frequent; reader often has to rely on own interpretation. 1 Errors of grammar or word order so severe as to make comprehension virtually impossible.
Mechanics 6 Few (if any) noticeable lapses in punctuation or spelling. 5 Occasional lapses in punctuation or spelling which do not, however, interfere with comprehension. 4 Errors in punctuation or spelling; occasional re-reading necessary for full comprehension. 3 Errors in punctuation or spelling; lead sometimes to obscurity. 2 Errors in punctuation or spelling so frequent that reader must often rely on personal interpretation. 1 Errors in punctuation or spelling so severe as to make comprehension virtually impossible.
Vocabulary 6 Use of vocabulary and idiom rarely (if at all) distinguishable from that of educated native writer. 5 Occasionally uses inappropriate terms or relies on circumlocutions; expression of ideas hardly impaired. 4 Uses wrong or inappropriate words fairly frequently; expression of ideas may be limited because of inadequate vocabulary. 3 Limited vocabulary and frequent errors clearly hinder expression of ideas. 2 Vocabulary so limited and so frequently misused that reader must often rely on own interpretation. 1 Vocabulary limitations so extreme as to make comprehension virtually impossible.
Score Gramm: ___+ Voc: ___+ Mech: ___+ Fluency: ___+ Form: ___= ___ (TOTAL)

Another example of analytic scale was provided by Brown J. D. & Bailey (1984 in Brown H. D. 2004). In this case, the scale served as a model to rate composition tasks (i.e. writing). The scale included five categories, namely: organization, logical development of ideas, grammar, punctuation, spelling, and mechanics, style and quality of expression. Each category was divided into five different levels ranging from "Excellent" (max. points: 20) to "Unacceptable" (max. points: 1). The total score is, once again, given by the sum of the points in each category for a total of 100 points.

There are several advantages to analytic scoring, but there are also a number of disadvantages. As Hughes (2003: 102) maintains, analytic scoring "disposes of the problem of uneven development of subskills in individuals". Besides, analytic scoring forces scorers to consider some aspects of a performance, which they would otherwise ignore. Last, analytic scoring is rather reliable since scorers are given "five [or more] 'shots' at assessing the student's performance" (Hughes, 2003: 103).

However, the main disadvantage of analytic scoring is that it is fairly time-consuming (definitely more than holistic scoring). Hughes (2003: 103) mentions a second disadvantage, namely that "concentration on the different aspects may divert attention from the overall effect" of the performance.

Having said this, when it comes to choosing which scoring procedure to adopt, the purpose of the test/assessment should always be taken into consideration. For instance, "if diagnostic information is required directly from the ratings given, then analytic scoring is necessary" (Hughes, 2003: 105).

Another aspect that needs to be taken into consideration are the circumstances, in which the scoring will take place:

> "if it is being carried out by a small, well-knit group at a single site, then holistic scoring, which is likely to be more economical of time, may be the most appropriate. But if scoring is being conducted by a heterogeneous, possibly less well-trained group, or in a number of different places, analytic scoring is probably called for" (Hughes, 2003: 105).

1.4 Non-traditional/Alternative Assessment: old and new paradigms

With these important clarifications in mind, it now seems important to introduce the concept of non-traditional/alternative assessment, which will be the focus of this dissertation. As discussed in *§1.2*, there are two approaches to assessment: one is summative, and the other is formative. Although both are essential, they serve different

purposes. While summative assessment judges the 'final product' of learning, formative assessment judges the process of learning.

With regard to alternative assessment, many scholars agree that it should only serve formative purposes, rather than summative ones, in that alternative assessment "supports and guides learning as it takes place" (Little 2002: 186). Having said this, other scholars argue that alternative assessment can be valuable for summative purposes too. In particular, portfolios seem to be the most suitable for summative purposes, since they contain samples of what should be the 'best production' of a student: "a selection of work that in the owner's judgement best represents his/her second/ foreign language proficiency" (Little 2002: 182). Therefore, portfolios could be used to assess the 'final product' of a course or a course unit. Nevertheless, researchers also agree that portfolios are part of formative experience. This particular aspect will be discussed in *§1.4.5*. On the contrary, observation, peer assessment and self-assessment seem to be best suited to serve formative purposes. Indeed, these can be implemented during a course with the pedagogical function of making "the language learning process more transparent to the owner", and fostering "the development of learner autonomy" (Little 2002: 182).

It is undeniable that assessing in heterogeneous classes is becoming increasingly difficult, and that heterogeneity in the modern EFL classroom calls for new, non-traditional forms of assessment. On the one hand, teachers want to favour the conditions to ensure significant learning experiences to every student, having the learning differences between students in mind. On the other hand, however, teachers also acknowledge that the individualization of learning paths, which appears to be more feasible thanks to alternative assessment, requires great effort and, especially, time.

In spite of this, many scholars and experts agree that alternatives in assessment seem to be exceptional tools that can enrich the learning experience and lead to greater learner autonomy. Assessment, being it teacher assessment or non-traditional assessment, allows students to monitor their own learning (or, having it monitored by teachers). Indeed, comments, observations, suggestions and, in some cases, criticism, make students more aware of where they are in their learning compared to where they should be (Hattie 2012). This monitoring of students' learning progress is an essential part of the whole learning process, in that it encourages students to adopt self-regulatory behaviours to progress in their learning.

Among the vast array of alternative assessments that enhance learner autonomy, peer assessment and self-assessment seem to be the most obvious ones. Indeed, self-assessment emphasises the connection between an individual and his/her work and, therefore, the actions that an individual has to take to improve his/her work and, consequently, in his/her learning process overall. Instead, although peer assessment consists in students assessing the performance of their peers, even this practice can have positive repercussions on the self: by judging the work of others, students will become more aware of their own skills and of the steps that they need to take to reach a higher level. Having said this, portfolios and learner diaries are also exceptional methods to foster learner autonomy and, therefore, can be used for self-assessment. For example, Little (1991: 52) notes that learner diaries serve "the necessary purpose of giving retrospective shape to the learning process, making it tangible, something that can be recalled and talked about". This huge benefit of learner diaries can be applied to portfolios and e-portfolios as well, and will be further discussed in *§1.4.5*.

The next section will be devoted to an overview of learner autonomy.

1.4.1 Learner Autonomy: some definitions

In this regard, it now seems appropriate to introduce another concept, which will be addressed further in Chapter Two as one of the main benefits of adopting peer assessment in the EFL classroom, namely: learner autonomy (or, learner independence).

Holec (1981: 3 in Novello 2014: 20) provided a definition of learner autonomy as the "ability to take charge of one's own learning". Ten years later, Little (1991: 4) expanded this definition and defined learning autonomy as

> "a capacity for detachment, critical reflection, decision-making and independent action. It presupposes, but also entails that the learner will develop a particular kind of psychological relation to the process and content of his learning."

Besides, Little (1991: 1) further suggests that the main features of autonomous learners lie in the fact that they

> "understand the purpose of their learning program, share in the setting of learning goals, take initiatives in planning and executing learning activities, and regularly review their learning and evaluate its effectiveness."

In other words, learner autonomy entails a variety of self-regulatory behaviours that "develop – through practice – as a fully integrated part of the knowledge and skills that are the goal of learning" (Little et al's, 2003: 4). As we can see from this last definition, another term, which is often associated with learner autonomy, is 'self-regulation' (in this case, self-regulatory behaviours). It is defined as "the degree to which individuals are active participants in their own learning" (Dörnyei and Skehan 2003 in Griffiths 2008: 86).

Among the self-regulatory behaviours one can, for instance, include language learning strategies, which can be described as

> "specific actions taken by the learner to make learning easier, faster, more enjoyable, more self-directed, more effective, and more transferable to new situations" (Oxford 1990: 8 in Griffiths 2008: 84).

Brown (2000: 122-123) distinguishes between learning strategies, which "relate to input — to processing, storage, and retrieval" and communication strategies, which "pertain to output, how we productively express meaning, how we deliver messages to others". Among learning strategies, Brown mentions cognitive and socio-affective strategies. Instead, to communication strategies belong avoidance strategies and compensatory strategies.

Oxford (1990) divides learning strategies into two groups: direct strategies and indirect strategies. Among the first group, one can find strategies related to memory, cognitive strategies and compensation strategies. Instead, to the second group belong metacognitive strategies, affective strategies and social strategies. Learners consciously adopt one (or more) strategies, in accordance with their preferred learning style (see *§1.1.2*).

1.4.2 Common Features in Alternative Assessment

According to García and Pearson (1991: 357), the main goal of alternative assessment is to "gather evidence about how students are approaching, processing, and completing real-life tasks in a particular domain", 'real-life' being the keyword here. Indeed, much dissatisfaction with traditional assessment procedures and testing originates from the belief that these are useless, in that they test artificial and non-authentic language, which will never occur in natural communicative events (Meyer 2002; Römer 2004). Instead, as Huerta-Macías (2002: 339) maintains, alternative assessment differs

from traditional testing in that "students are evaluated on what they integrate and produce, rather than on what they are able to recall and reproduce".

The growing discontent and dissatisfaction with traditional assessment methods, which is currently felt all over the education system, has resulted in an evident need to reassess the role of assessment. Therefore, new[7] non-traditional/alternative forms of assessment have been recently praised. Indeed, as Richards and Renandya (2002: 335) point out,

> "interest in the use of non-traditional forms of assessment in the classroom reflects the changing paradigm in education in general, and in second language teaching in particular. The old paradigm is slowly giving way to a new one."

Table 3 below, taken from Richards and Renandya (2002) exemplifies the above-mentioned changes in education:

Table 3: Old and New Paradigm in Education (Richards, Renandya 2002).

Old Paradigm	New Paradigm
1. Focus on language	1. Focus on communication
2. Teacher-centred	2. Learner-centred
3. Isolated skills	3. Integrated skills
4. Emphasis on product	4. Emphasis on process
5. One answer, one-way correctness	5. Open-ended, multiple solutions
6. Tests that test	6. Tests that also teach

Similarly, Brown (2004: 13) also compares traditional and alternative assessment features, as shown in *Table 4* below:

Table 4: Traditional and Alternative Assessment: A Comparison (Brown 2004).

Traditional Assessment	Alternative Assessment
One-shot, standardized exams	Continuous long-term assessment
Timed, multiple-choice format	Untimed, free-response format
Decontextualized test items	Contextualized communicative tasks
Scores suffice for feedback	Individualized feedback and washback

[7] Discussions on alternative assessment are fairly recent. One of the major contributors to these discussions were Elbow and Belanoff, who published *Portfolios as a substitute for proficiency exams* in 1986. Their work was a seminal research on the use of portfolios as alternatives to traditional tests at University level. "Their research triggered a wave of interest in portfolio use in language teaching settings" (Fox 2017: 139) and other researchers and scholars started discussing other alternatives in assessment. Among these researchers and scholars, Fox (2017) mentions, for example, Valdés and Figueroa (1994), Huerta-Macías (1995), Delandshere and Petrosky (1998), Gipps (1999), Hamp-Lyons and Condon (2000), Rea-Dickins (2001), Maslovaty and Kuzi (2002), Lynch and Shaw (2005), Black and William (2006), and Cheng and Fox (2013).

Norm-referenced scores	Criterion-referenced scores
Focus on the "right" answer	Open-ended, creative answers
Summative	Formative
Oriented to product	Oriented to process
Non-interactive performance	Interactive performance
Fosters extrinsic motivation	Fosters intrinsic motivation

As can be clearly noted from a comparison of the two tables, both emphasize that the focus of alternatives assessment procedures is on the so-called communicative competence[8]. Since the acquisition of the communicative competence should be the ultimate goal of a language course (see, for instance, Little 1991[9], Brown 2004[10] and Novello 2014[11]), these alternative forms of assessment seem perfectly fitting: language courses, indeed, aim at providing useful and authentic communicative tools to students. In fact, alternative assessment is also known under the name of 'authentic assessment' (and 'informal assessment').

In this respect, the notion of authenticity should now be explained: Bachman and Palmer (1996: 23 in Brown 2004: 28) define authenticity as "the degree of correspondence of the characteristics of a given language test task to the features of a target language task". Brown then specifies that

> "when you make a claim for authenticity in a test task, you are saying that this task is likely to be enacted in the 'real world'. Many test items types fail to simulate real-world tasks" (2004: 28).

Besides, both tables show that alternative assessment is oriented to process, rather than product. Similarly, Hattie points out that alternatives to traditional assessment are "assessment *for* learning, rather than assessment *of* learning" (2012: 141), in that they should be exploited by teachers to gain important information both about the learners and

[8] This term was coined by Dell Hymes in 1967.
[9] "The principal goal of second and foreign language teaching has always been to enable learners to use the language in question as a medium of communication […]. Communicative efficiency in the target language community depends on learners having the independence, self-reliance and self-confidence to fulfil the variety of social, psychological and discourse roles in which they are cast" (Little, 1991: 27).
[10] "Lately there seems to be a trend to supplement traditional test designs with alternatives that are more authentic in their elicitation of meaningful communication" (Brown, 2004: 13).
[11] "L'obiettivo dell'insegnamento linguistico è quello di far padroneggiare (vale a dire essere competente a tal punto da utilizzare le conoscenze in contesti diversi ad un livello efficace per lo scopo stabilito) al discente tale competenza, la quale comprende tutte quelle competenze specifiche (o sotto-competenze) che entrano in gioco nei comportamenti linguistici" (Novello, 2014: 57).

their teaching and make adjustments if necessary, while the learning process is still going on.

Lastly, the data provided in both tables also suggest that while traditional assessment procedures focus on the correctness of the answer and only allow one possible correct answer, alternative assessment allows for more than just one correct answer, thus also encouraging creativity in responding to a task.

In spite of what has been said so far about alternative assessment, Brown (2004: 13) points out that one should not think that "everything on the left-hand side [of table 4] is tainted while the list on the right-hand side offers salvation to the field of language assessment". Thus, he simply suggests that teachers incorporate more "updated" assessment procedures into the more traditional ones.

Among the vast array of non-traditional forms of assessment, Brown (2004) mentions learner journals, portfolios, teacher-learner conference, self-assessment and, of course, peer-assessment (which will be the focus of Chapter Two). Besides, Huerta-Macías (2002) adds observation, checklists, reading logs, debates, and role-plays to the list. The following paragraphs (*1.4.3-1.4.6*) will be devoted to a more thorough elucidation of some of these procedures.

First, however, it seems important to emphasize that choosing the right form of assessment is a complex undertaking, which involves decisions based on several aspects, such as "the format of the course and the educational setting, learners' needs, preferences of the teachers, institutional or curricular requirements, available resources" (Çelik, Türkan 2014: 424).

1.4.3 Observation

Teachers observe their students constantly and systematically. Indeed, through observation, they can obtain a wealth of useful information and form intuitions about their students' performances. The main advantage of systematic observation lies in the fact that students are not aware that they are being observed. Therefore, "the naturalness of their linguistic performance is maximized" (Brown 2004: 267).

Just like every other assessment procedure, observation requires careful planning to be effective: it is usually associated with checklists and observation grids, in which teachers are supposed to record what they observe about a student and the specific

objectives of the observation should be decided beforehand. As Novello (2014: 53) maintains, "observation grids are very beneficial, in that teachers will not only have a final grade, but also specific information to support the student in his learning process". For instance, teachers can use these grids to jot down some notes on the most critical aspects of the performance, and then share their notes with the student to make him/her aware of what to focus on more next.

Many aspects of a performance can be controlled through observation, such as discourse-level skills, oral production skills, grammatical correctness, pronunciation and intonation, and interaction with classmates. However, as Brown (2004: 268) points out, it is important "not to overestimate the number of different elements you can observe at one time — keep them very limited". Moreover, it is advisable to select only a few students at a time to be able to devote the right amount of time and attention to every student.

1.4.4 Learner Diaries/Journals

Learner diaries consist of entries, which are regularly written by students, where they can 'store' their "thoughts, feelings, reactions, assessments, ideas, or progress toward goals" (Brown 2004: 260). Diary writing is a pedagogic tool, in that it can be a tool for self-reflection: students can, for example, set their goals, record their progress by keeping track of how far they have come towards the above-mentioned goals, practice writing, "using writing as a 'thinking' process" (Brown 2004: 261), and self-assess. Diary entries should be written in the target language, where it "comes to be used not only as a medium of classroom communication but also as a channel of learning and a tool for reflection" (Thomsen 2003: 29 in Han 2011: 200).

As mentioned above, learner diaries can also be used by students to self-assess. Therefore, they seem to be excellent tools in fostering learner autonomy: "it is impossible to achieve learner autonomy without reflection. […] Besides being an important activity in itself, writing supports reflection" (Han 2011: 200).

Having said this, some claim that journals are "too free a form to be assessed accurately" (Brown 2004: 262), since it could be difficult to set up criteria for evaluation. However, teachers should make clear to their students that journals are not meant to be perfectly written accounts, since grammatical correctness is not their purpose. Instead,

the main assessment criterion will depend on the purpose of the journal: language-learning logs, responses to readings, or self-assessment reflections (Brown 2004). For instance, in the case of 'responses to readings' teachers could take the ability of the students to write a critical commentary on the reading as the main criterion for assessment.

Diary writing, just like other alternatives to traditional assessment, may cause some confusion at first; this is why there need to be some guidelines to instruct the students and to allow them to fully exploit diary writing as a pedagogic tool. For instance, models or suggestions on what to incorporate in the journal could be provided to the students.

1.4.5 Portfolios and the European Language Portfolio (ELP)

Portfolios are "collections of examples of work that, as a collection, reveal both the capability and the progress of a learner" (Cameron 2001 in Novello 2012: 107). Typical portfolios can either contain the students' "total writing output to represent his or her overall performance" (Peñaflorida 2002: 347), or just part of their personal contributions, thus providing a fair estimate about a student's progress. Portfolios usually contain written samples. However, electronic or e-portfolios, which result from "the widespread use of increasingly sophisticated technologies" (Fox 2017: 135) can also contain spoken production.

Portfolios are useful especially in that they can give a 'snapshot' of how much a student's writing has progressed throughout a course; indeed, they have the advantage of usually containing several types of written tasks. However, as Peñaflorida (2002: 348) rightly points out, "conferencing between students and the teacher is an inherent activity in portfolio assessment": through one-to-one conversations, teachers can "uncover potential misunderstandings" (Kroll 1991: 259 in Peñaflorida 2002: 352), request clarification, discuss critical aspects as well as praise the student.

Of course, portfolios do not constitute the assessment. As Herman et al. (1992: 88) point out,

> "the 'assessment' in portfolio exists only when 1) an assessment purpose is defined; 2) criteria or methods for determining what is put into the portfolio, by whom, and when, are explicated; and 3) criteria for assessing either the collection or individual pieces of work are identified".

Portfolios as collections of written samples may also be very useful for students, in that they can go back to their past work and see for themselves how much they have progressed. Indeed, as Brown (2004: 37) points out, "once the pressure of assessment has come and gone, students may be able to look back on their written work with a fresh eye".

In the same way, e-portfolios which contain spoken production are equally useful for students, who can easily access their past production and see how much they have progressed.

In the context of portfolios, it seems important to also mention the Council of Europe's European Language Portfolio (ELP). The ELP has three obligatory components: "a *language passport* […], a *language biography* […], and a *dossier*" (Little 2002: 182). The ELP can be used by students to "directly rate their own language proficiency" (Bailey 2017: 330) and, therefore, has an enormous value for self-assessment. Indeed, according to the official website of the Council of Europe, the ELP was developed, among other things, "to support the development of learner autonomy" by allowing "users to record their language learning achievements and their experience of learning and using languages"[12]. Similarly, Little (2002: 183) notes that

> "using the ELP necessarily engages the owner in self-assessment; and the basis for self-assessment is provided by the common reference levels of the CEF [Common European Framework]. These describe second/ foreign language proficiency in relation to five communicative activities at six levels.

In particular, Little (2002: 186) states that "the basis for self-assessment in the ELP is provided by 'can do' statements that describe communicative behaviour". These statements communicative[13] behaviours are *listening*, *reading*, *spoken interaction*, *spoken production* and *writing*, and the six levels are A1 (*Breakthrough*), A2 (*Waystage*), B1 (*Threshold*), B2 (*Vantage*), C1 (*Effective Operational Proficiency*), and C2 (*Mastery*) (Little 2002: 183).

The next section will be devoted to a more in-depth discussion of self-assessment, which was mentioned several times in the current section.

---

[12] DOI: https://www.coe.int/en/web/portfolio
[13] See 'communicative competence' in *§1.4.2*.

1.4.6 Self-Assessment

As the name suggests, self-assessment consists in assessment by learners on their own written or oral performance. As Gardner (1999: 50) notes, the term simply denotes "the mode of administration, i.e., assessments which are self-administered".

Self-assessment, together with peer assessment, is one of the most criticized assessment procedures. The main issues that are often raised against these types of alternative assessment are in terms of their validity, reliability and objectivity (see *§ 1.5*). However, self-assessment also has numerous benefits, first and foremost the enhancement of learner autonomy (see *§ 1.4.1*) and of self-regulation skills:

> "autonomous learners decide what to learn, when to learn and how to learn. […] Autonomous learners take responsibility for their learning and this includes taking responsibility for monitoring their progress" (Gardner 1999: 51).

More specifically, students can check the effectiveness of their learning strategies toward the achievement of their objectives, reflect on their goals, identify critical areas of their learning and, thus, make adjustments while the learning process is still going on.

Additionally, self-assessment is also a useful tool to raise students' language awareness, by making them reflect on those areas of the language that might need improvement.

However, students may not welcome self-assessment at first. Indeed, past evidence has shown that they may perceive themselves as inadequate and unprepared to carry out self-evaluation. Nonetheless, this issue can be overcome. For instance, teachers can facilitate self-assessment by providing specific guidelines on where students should focus their attention. Indeed, as Gardner (1999: 55) clarifies, self-assessment "is not about leaving students to fend for themselves. It is about teachers creating opportunities for students to make responsible choices".

1.5 Validity, Reliability and Objectivity

Valuable though these new forms of assessment may be, many teachers are still sceptical about employing them regularly in the classroom. Huerta-Macías (2002: 338) noted, in fact, that some issues "related to validity, reliability, and objectivity are often raised as objections to alternative assessment".

To begin with, 'validity' is a criterion which refers to the usefulness of a test. The term is described by Gronlund (1998: 226) as "the extent to which inferences made from assessment results are appropriate, meaningful and useful in terms of the purpose of the assessment". More specifically, validity is meant to describe the extent to which a test measures what it purports to measure. For instance, it would be wrong to test reading by asking students to answer open-ended questions using their own words; indeed, in this case we would not actually be testing their reading skills, but rather their writing skills. Instead, to test reading it would be more appropriate to have multiple-choice exercises. Brown (2004: 22) suggests that validity could be established by examining "the extent to which a test calls for performance that matches that of the course or unit of study being tested". Moreover, he also mentions 'criterion-related validity', which is "the extent to which the 'criterion' of the test has actually been reached" (2004: 24). Lastly, 'face validity' is also an important facet of validity, and it refers to "the extent to which students view the assessment as fair, relevant, and useful for learning" (Gronlund 1998: 210 in Brown 2004: 26). If a test is well-constructed, feasible within time limit, has clear and transparent instructions, face validity is likely to be high. On the contrary, if a test is too challenging, complicated and not well-designed, face validity is going to be very low.

Moving on to reliability, it is a criterion which refers to the consistency of a test. Indeed, the term denotes the consistency of the results through subsequent administrations of a test. According to Novello (2014: 46), reliability "requires that a test obtains the same result with different evaluators". In addition, as Huerta-Macías (2002: 340) argues, it follows that "if a procedure is valid, then it is reliable".

Several elements can affect reliability, such as the conditions in which a test takes place and errors in assigning points: in this regard, Brown (2004) mentions student-related reliability, rater (or, scorer) reliability, test administration reliability, and test reliability. Student-related reliability refers to "physical or psychological factors, which may make an 'observed' score deviate from one's 'true' score" (2004: 21). Instead, rater reliability refers to "human error, subjectivity and bias" and "lack of attention to scoring criteria, inexperience, inattention" (2004: 21). As has already been stated in *§1.3*, scorer reliability is one of the biggest issues in assessment. According to the *APA Dictionary of Psychology*, scorer reliability indicates

> "the extent to which independent evaluators produce similar ratings in judging the same abilities or characteristics in the same target person or

> object. It is often expressed as *correlation coefficient*. If consistency is high, a researcher can be confident that similarly trained individuals would likely produce similar scores on targets of the same kind. If consistency is low, there is little confidence that the obtained scores could be reproduced with a different set of raters."

The conditions in which a test takes place can also play an important role in affecting reliability: street noise, desk and chairs conditions, or ambient lightning conditions can disturb the test-takers. Lastly, some problems may relate directly to the nature of the test; for instance, "if a test is too long, test-takers may become fatigued by the time they reach the later items and hastily respond incorrectly" (2004: 22).

Having said this, reliability can also be ensured, for instance by having more than just one rater or through the appropriate use of assessment criteria, which should be transparent and shared with the learners (see *§ 1.1.1*). Besides, Huerta-Macías (2002: 341) claims that another means to ensure reliability lies in the so-called 'triangulation', which refers to "the collection of data or information from three different sources or perspectives". As *Figure 2* below shows, Huerta-Macías suggests among suggests the students themselves, the parents and the teacher as sources of data. Of course, this is just an example of how triangulation works; another possibility would be, for instance, to have a peer (or more than one) instead of 'family'.

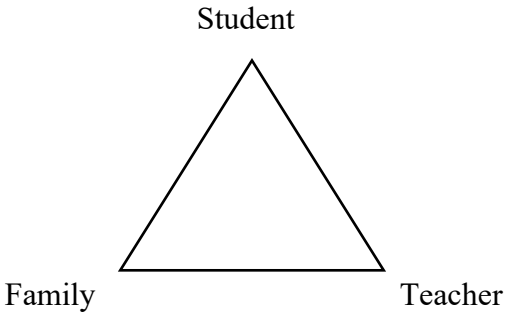


Figure 2: An example of triangulation of data in an EFL environment (Huerta-Macías 2002).

Last but not least, 'objectivity' indicates the degree to which a test (or an assessment) is unbiased, neutral, impartial and independent of external influence. Many argue that alternative assessment procedures (especially peer- and self-assessment) are not objective. However, as White (2009: 24) points out, every type of assessment can be subject to erroneous judgements, irrespective of their nature (teacher assessment, peer

assessment, self-assessment). Indeed, "since assessment involves making judgments, it will inevitably be subject to some error and bias".

In this context, Huerta-Macías (2002: 342) also maintains that traditional forms of assessment are no more objective nor reliable than their non-traditional counterparts:

> "a standardized test merely represents agreement among a number of people on scoring procedures, format, and/or content for that specific test. In other words, these individuals are not really objective; they just collectively share the same biases. […] There is no reason, then, to consider alternative assessment as being any less objective than traditional testing.

Like reliability, objectivity can also be encouraged through pre-determined criteria: "clear assessment criteria can go a long way toward encouraging objectivity" (Brown, 2004: 277). Moreover, guiding questions or checklists can also compensate for the initial disorientation of both teachers and learners towards these new assessment practices, thus encouraging objectivity.

1.6 Theories and Concepts in Foreign and Second Language Teaching

A good teacher needs to be aware of the biological mechanisms of language acquisition. The function of the brain in the process of second language acquisition has long been the focus of much research. Neurolinguistic concepts such as hemispheric lateralization, the hypothesis of the critical period, left- and right-brain dominance, bimodality and directionality are the basics for second language teaching.

The importance of these theories also has repercussions in language assessment. For instance, it seems fundamental that teachers know the difference between 'acquisition' and 'learning' and thus be able to observe (through assessment), whether their students are internalizing new knowledge (acquisition) or simply memorizing it (rote learning[14]). Moreover, the sections devoted to the explanation of the *LASS*, of scaffolding and the *ZPD* are related to the concept of cooperative learning, peer assessment and peer feedback, which will be dealt with more in-depth in Chapter Two.

The next sections will be concerned with some other very important theories and concepts in (second) language acquisition, which provide theoretical support for peer assessment.

[14] See footnote 3.

1.6.1 Krashen's Second Language Acquisition Theory (SLAT)

In 1977[15] Stephen Krashen postulated five different hypotheses to form his Second Language Acquisition Theory (SLAT).

According to the first hypothesis, the acquisition-learning hypothesis, language learners have two different modes for internalizing a target language: acquisition and learning. Acquisition occurs at the subconscious level, it is an "intuitive process of constructing the system of a language, not unlike the process used by a child to 'pick up' a language" (Brown 2000: 278). Acquisition allows the learners, for example, to form new sentences by using the language creatively. On the contrary, learning is a conscious, systematic and structured process. According to the second hypothesis, the monitor hypothesis, learning has a monitoring function: "it is a device for 'watchdogging' one's output" (Brown 2000: 278). The third hypothesis is the natural order hypothesis, and it refers to the 'sequentiality' of languages, whereby some forms/structures have to be acquired/learnt before others[16]. The fourth hypothesis is the well-known input hypothesis "*i* + 1*"*, "i" standing for "input":

> "an important condition for language acquisition to occur is that the acquirer *understand* (via hearing or reading) input language that contains structure 'a bit beyond' his or her current level of competence… If an acquirer is at stage level *i*, the input he or she understands should contain *i* + 1" (Krashen 1981: 100 in Brown 2000: 278).

Put more simply, "the language that learners are exposed to should be just far enough beyond their current competence that they can understand most of it but still be challenged[17] to make progress" (Brown 2000: 278). Furthermore, the input should be comprehensible to become intake, which can be defined as the subset of all input that a learner has managed to save into his long-term memory store. Of course, it goes without saying that intake will always be a bit less than the initial input.

Lastly, the fifth hypothesis is the affective filter hypothesis: according to this hypothesis, acquisition will only occur in a low-anxiety environment, where the 'affective filter' is low. Otherwise, "too much anxiety may hinder the process of successful second language learning" (Brown 2000: 152).

---

[15] The input hypothesis was then further discussed by Krashen in 1981, 1982, and 1985.
[16] For example oral skills such as listening and speaking precede written skills such as reading and writing.
[17] See *§1.1.2* "il piacere della sfida" (Balboni 2015).

The next section will be centred on two cornerstones of linguistics: Noam Chomsky's nativist approach, which resulted in the language acquisition device, and Jerome Bruner's language acquisition support system.

1.6.2 Language Acquisition Device (LAD) and Language Acquisition Support System (LASS)

According to Chomsky, human beings are born with a genetic capacity that predisposes them to acquire languages in a (seemingly) short span of time (especially the first language); this is the so-called 'innateness hypothesis'. Chomsky labelled this innate capacity of the human being "Language Acquisition Device" (LAD). Therefore, it can be said that language learners are predisposed to language acquisition.

Bruner rejected Chomsky's view, claiming that the *LAD* alone is not enough. Instead, he argued for a need for a support to the *LAD*: the well-known "Language Acquisition Support System" (LASS). In the case of first language acquisition, the support comes from the parents. In the case of second/foreign language learning, it comes from the teacher and the peers, who serve as 'catalysts' for the learner, in that they can provide support to facilitate the learning. In particular the fact that also peers play an important role in language acquisition is very important for the purpose of this dissertation.

1.6.3 The concept of 'scaffolding' and Vygotsky's Zone of Proximal Development (ZPD)

Vygotsky holds a similar view to Bruner in that he also believes that social interaction plays a fundamental role in cognitive development. The concept of scaffolding relates to Vygotsky's Zone of Proximal Development (1978). Therefore, there is an evident need to first define what the ZPD is, that is the distance between learners' existing developmental state and their potential development: "the distance between the child's actual cognitive capacity and the level of potential development" (Vygotsky 1978: 86).

Put more simply, Slavin (2003: 44 in Brown 2000: 24) explains that "the ZPD describes tasks that a learner has not yet learned but is capable of learning with appropriate stimuli", i.e. with the assistance of more competent peers or adults (such as teachers). Indeed, as Vygotsky (1978: 90) claims:

> "an essential feature of learning is that it creates the zone of proximal development; that is, learning awakens a variety of internal developmental

> processes that are able to operate only when the child is interacting with people in his environment and in cooperation with his peers."

Let us now take into consideration the related notion of 'scaffolding'. The term originates from construction industry, where it is employed to indicate the temporary frames/structures that are needed during the construction of a building. However, this term is often employed in language teaching as a metaphor to indicate the above-mentioned support that learners receive during their process of language acquisition. As argued by Hattie (2012: 144), the purpose of scaffolding is to

> "provide support, knowledge, strategies, modelling, questioning, instructing, restructuring, and other forms of feedback, with the intention that the student comes to 'own' the knowledge, understanding and concept."

According to Wood, Bruner and Ross (1976: 90), the scaffolding process entails more than simply "an assisted completion of the task". Instead, they claim that "it may result, eventually, in development of task competence by the learner at a pace that would far outstrip his unassisted efforts".

Of course, scaffolding does not mean to immediately hand the solution to the learner. Instead, thanks to their expert assistance, tutors could and should assist the "novice" (Wood, Bruner, Ross 1976), the tutee, in many ways, such as simplifying the task, keeping the novice focused toward an objective, marking critical features of the task, and only at the end should they model solutions to the task. The concept of scaffolding will be addressed further in Chapter Two.

To conclude, this chapter was intended to shed some light on the newly available non-traditional forms of assessment, as well as introducing important practices in classroom assessment. This chapter has provided some core definitions, such as testing, assessment and feedback, formative and summative assessment, and holistic and analytic scoring. Furthermore, it was also intended to give an overview of some of the key issues related to validity, reliability and objectivity, which are often raised as objections to alternative assessment. Lastly, the final section of this chapter dealt with important theories related to language acquisition, and language teaching and assessment.

Chapter Two will be dedicated to an in-depth analysis of peer assessment, with a prime focus on its advantages and drawbacks. Furthermore, possible solutions to the

issues of PA by different experts will be provided. Some important concepts related to assessment and alternative assessment that were discussed in this chapter will be revisited and expanded, and especially that of 'learner autonomy' (or, learner independency): indeed, learner autonomy seems to be one of the biggest advantages of such assessment procedure. Then, eight studies on peer assessment by different academics will be reviewed.

Chapter 2: Peer Assessment and Peer Feedback

"*Tell me and I forget, teach me and I may remember, involve me and I learn.*" — Confucius/B. Franklin[18]

Chapter Two will be devoted to an in-depth analysis of peer assessment and peer feedback. First, I will make a clarification of the difference between *teacher talking time* (TTT) and *student talking time* (STT), to introduce the idea of involving students in their learning (in this case, through peer assessment), as well as a short introduction to student-centred learning. Second, I will provide some definitions of peer assessment and make a distinction between peer assessment and peer feedback. Then, I will address the key issues that are often raised against peer assessment that were also addressed in Chapter One, namely: validity, reliability and objectivity, and I will point out possible solutions to these issues that have been offered by different educational experts. As mentioned in Chapter One, peer assessment belongs to the so-called alternative assessment procedures. Therefore, those issues related to alternative assessment, are also present in peer assessment. Next, the numerous advantages that may come from adopting peer assessment along with traditional teacher assessment will also be taken into consideration. Last, eight studies on peer assessment, which will partly serve as models for my own study on peer assessment in Chapter Three, will be reviewed.

2.1 Student-centred learning, Teacher Talking Time (TTT) and Student Talking Time (STT)

Student-centred approaches, also known as learner-oriented approaches, as the name itself suggests, comprise those teaching methods that shift their focus from teachers to learners, thus making students the real protagonists of their own learning process. Student-centred approaches imply a much wider involvement of students in their learning than in traditional teacher-learning, also when it comes to assessment, since students are expected to take initiatives in learning and assessment activities and adopt self-regulatory behaviours to guide their own learning process. Therefore, a more learner-centred approach can foster autonomy.

[18] Some people attribute this quotation to Benjamin Franklin, some to Confucius. Irrespective of who actually is the author, this quotation is a true statement for anyone to live by, but especially for educators. This statement seemed perfectly fitting given the focus and aim of this chapter.

To start the discourse on peer assessment, I would like to shortly introduce the distinction between *teacher talking time* (TTT) and *student talking time* (STT). This clarification aims at pointing out how little students 'act' in the language classroom, being this speaking, participating, or assessing, and how much teachers should make efforts to find more opportunities for students to participate more actively in class. Among these opportunities that allow for a more active participation of students in the language class we can mention peer assessment, which will be discussed thoroughly in the following sections.

A fairly recent claim within the education field is that language classrooms are dominated by the so-called teacher talking time. This expression denotes the amount of time that a teacher spends talking in class. Indeed, it seems that teachers rather than students are the real protagonists in the language classroom environment: they talk most of the time, whereas the time dedicated to student's talk is usually very little. Of course, this is a major issue, since the acquisition of the communicative competence, which is (or, should be) the ultimate goal of a language course, can only be acquired by students through practice. In this context, of course, by 'practicing' I mean 'communicating', 'talking', "allowing learners opportunities to speak and learn from speaking" (TeachingEnglish | British Council)[19]. However, when teachers consciously or unconsciously monopolise classroom discourse, the role of students is merely that of respondents. Therefore, in this case, "opportunities for developing the speaking skill are severely limited" (TeachingEnglish | British Council)[20]. Moreover, during teacher talking time, students' autonomy is also very limited, in that it is teachers who decide when students will be given space to talk:

> "if the teacher is constantly dominant and controlling, the learners take no responsibility for their own learning but learn what the teacher decides and when" (TeachingEnglish | British Council)[21].

As already stated in Chapter One, teachers should aim at encouraging their students to be more autonomous and responsible learners, with the goal of generating the right mindset for lifelong (or, at least, long-term) learning. Lifelong learning has been defined as a process to help learners

[19] For more information see: https://www.teachingenglish.org.uk/article/teacher-talking-time-0
[20] For more information see: https://www.teachingenglish.org.uk/article/pros-cons-teacher-talking-time
[21] For more information see: https://www.teachingenglish.org.uk/article/pros-cons-teacher-talking-time

> "to build up the attitudes, knowledge and skills they need to become more independent in thought and action, and also more responsible and co-operative in relation to other people" (Council of Europe 2001: xii in Huttunen 2003: 122).

Put more simply, lifelong learning can be defined as the continuous acquisition of new knowledge, even outside a "formal" education environment. Whenever teachers manage to increase student engagement, for instance by establishing the relevance of the topics that they are teaching and by sparking the interest of their students, then long-term learning is more likely to happen. One very straightforward way of increasing student engagement would be, therefore, to have him or her take part in classroom activities not just as 'spectator', but rather as 'host', 'presenter'. On the contrary, if teachers keep having the dominant role in the language classroom, learners will never be able to make that step toward autonomy, which would then also lead to sharpening and refining various skills. It seems, therefore, essential that teachers take into consideration this issue, and start creating more opportunities for student-centred activities that aim to reach learner autonomy. One way of doing this could be, for instance, through peer assessment.

Nevertheless, it seems important to point out that the percentage of time that a teacher spends talking in class depends on the stage of the learning (i.e. proficiency levels): if the students have just begun learning a foreign language, then it goes without saying that the language they will be able to produce is very little. At this stage, teachers might want to provide numerous explanations as well as authentic snippets of the target language to their students. Instead, at a higher stage, learners will be expected to take the floor and organize and manage their speaking production. In this respect, as Liu and Carless (2006: 281) point out, learning involves some form of socialization:

> "one important way we learn is through expressing and articulating to others what we know or understand. In this process of self-expression, we construct an evolving understanding of increasing complexity".

Topping (2005: 637) also shares the same view as he claims that learners "might never have truly grasped a concept until having to explain it to another, embodying and crystallizing thought into language". As already stated in Chapter One, Vygotsky (1978) also strongly believed that social interaction plays a fundamental role in cognitive development. Having said this, even at a higher stage, teachers do not completely disappear from the scene: in fact, they still need to be present to ensure the quality of their

learners' speaking and direct it toward a higher standard of quality. Besides, it seems important to emphasize that very often students do not have the opportunity to speak the target language outside the classroom[22], or, should they have this opportunity, it is limited to social media (video call). Therefore, it is extremely important that student talking time in the broader sense becomes a primary concern of teachers. All this discourse should be kept in mind even when talking about peer assessment.

2.2 Peer Assessment and Peer Feedback

As asserted in the section above, the top priority of teachers should be student involvement in the classroom activities and student autonomy[23]. Autonomous students can make choices about various aspects of their learning. These can include, for example, planning learning goals and adopting strategies to achieve these goals, monitoring one's learning, and even objectively assessing one's and others' skills (self-assessment and peer assessment). Indeed, peer assessment is one effective way of ensuring student involvement, in that students are placed in a teacher-like role as they are asked to assess their peers' performances. Peer assessment is commonly described as a procedure, through which students assess their classmates' performance and provide constructive feedback on the performance. Topping (1998: 250) also provides a clear definition of peer assessment, namely as "an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status". By 'acting as teachers', students will perceive that they are an essential part of their own learning process as well as that of their peers, thus becoming more motivated and actively engaged. Yet, it seems important to point out that students are not "surrogate teachers" (Topping 2005: 631) and, more importantly, it is not just the 'best students' who should act as peer helpers. On the contrary, the best solution appears to be "deploying helpers whose capabilities are nearer to those of the helped, so that both

[22] Of course, I am referring to a hypothetical EFL context, where the target language is not spoken in the country. Therefore, the only source of input (i.e. what is taught to the students, language-wise) comes from the teachers, who have the power to decide what, how and when to teach certain topics and aspects of a language, whence the absolute necessity to provide authentic examples of the target language. On the contrary, in an ESL context, the target language is spoken in the country. Therefore, students will have numerous opportunities to practice their speaking skills also (and, probably, especially) outside of the school walls.

[23] See *§1.4.1* for more information on learner autonomy.

members of the pair find some cognitive challenge in their joint activities" (Topping 2005: 632).

Put more simply, it seems that students with similar status (in terms of learning) should be matched for peer assessment. The term 'similar status' also suggests that peer assessment must be carefully planned in order for peers with similar learning paths to evaluate each other. Otherwise, students may feel judged and will manifest negative feelings toward this procedure and might not take the whole activity seriously. Besides, as suggested earlier, equal status is more likely to indicate equal-opportunity involvement of learners. Jung (2016: 10) also shares this view and maintains that this procedure should only occur between learners with "equivalent language ability and progression" to ensure balance and effectiveness.

Peer assessment belongs to the so-called formative assessment procedures discussed in Chapter One. To sum up, formative assessment "aims to improve learning while it is happening" (Topping 1998: 249). On the contrary, summative assessment occurs at the end of a course/unit, thus not providing useful information to direct, support and facilitate students toward improvement, nor can it be used by teachers to form future plans. Peer assessment seems more fitting for formative purposes since, for instance, it can enable "earlier error and misconception identification and analysis" (Topping 1998: 255) and push students to develop strategies to close their knowledge gaps.

Peer assessment is an extension of 'peer learning', which consists in the "acquisition of knowledge and skill through active helping and supporting among status equals or matched companions" (Topping 2005: 631). As Topping (2005: 632) further explains, peer learning and, more generally, cooperative learning, "is more than 'working together'"; it has been described as "'structuring positive interdependence' in pursuit of a specific shared goal or output" (Slavin 1990 in Topping 2005: 632). Consequently, in cooperative learning, "individual growth depends upon but also feeds into the collaborative construction of skills and knowledge" (Little 2003: 223).

Given the definitions of peer assessment mentioned at the beginning of this section, it seems important to point out that peer assessment activities are extremely variable, and different peer assessment modes exist. For instance, peer assessment can be done individually or in pairs/small groups, anonymously or in public, it can be mutual or a one-way direction (although the latter is not recommended), and the feedback can be

global or specific. Moreover, peer assessment can have different modes, namely: face-to face (verbal), pen-to-paper (written), asynchronous computer-mediated-communication, or synchronous computer-mediated-communication (Hansen Edwards 2013: 736). These modes have several advantages as well as some disadvantages (listed in *Table 1)*, and teachers must choose them carefully and in accordance with the purpose of feedback. For example, if teachers want their students to provide instant feedback to their peers, then face-to-face peer assessment and synchronous CMC may be more suitable. Last but not least, peer assessment can be conducted on every aspect of language learning, such as speaking production, writing, and listening. All these variables must be carefully taken into consideration beforehand and in accordance with the scope of the activity.

Table 1: Benefits and drawbacks of different modes of PA[24]. Taken from Hansen Edwards (2013).

Format	Face-to-face (verbal)	Pen-to-paper (written)	Asynchronous (CMC)	Synchronous (CMC)
Timing	+ instant feedback	- too little giving criticism face-to-face	+ more time to read and reflect - delayed feedback	+ instant feedback - enough time?
Social interaction	+ active participation			+ active participation + CMC may increase collaboration
Affect	- uncomfortable giving criticism face-to-face		+ computer-based mode more motivating + less anxiety as not face-to-face	+ computer-based mode more + less anxiety as not face-to-face
Feedback	+ opportunities to clarify meaning	+ written record of feedback - no opportunity to clarify meaning	+ written record of feedback - no opportunity to clarify meaning	+ written record of feedback + opportunity to clarify meaning
Language	+ can use L1 to clarify meaning			+ supports interaction

[24] In this table, + stands for the advantages and – stands for the drawbacks.

	+ verbal skills developed + supports interaction			

Having said this, it seems that peer assessment is particularly effective for assessing writing. As Peñaflorida (2002: 345) puts it, "the traditional way of evaluating papers where the teacher is the only reader for whom the students write […] is now being gradually replaced by the so-called extended readership". The term 'extended readership' indicates that the assessment of a written assignment is no longer the sole responsibility of teachers, but rather it undergoes multiple readers and assessors, each of whom provides feedback, suggestions and constructive criticism on the task. This procedure is thought to ensure a more complete and helpful feedback, since different assessors might focus on different aspects of the language and of the task: "students would receive more feedback from peers and more quickly than when academics are providing comments" (Liu and Carless 2006: 281). Besides, as Peñaflorida (2002: 351) adds, "students may not be as skilled as their teachers at responding to each other's work, but they are excellent in providing the one thing that writers need most – an audience".

An important distinction should now be made with regard to peer feedback and peer grading/assessment. As Liu and Carless (2006: 280) point out, peer feedback consists in a "communication process", whereby students exchange their opinions and "enter into dialogues related to performance and standards". Peer assessment, instead, entails much more than that: it entails students also providing grades on their peers' performance, as if they were actually teachers. The distinction that the two scholars remark on, then, is that

> "peer feedback is primarily about rich detailed comments but without formal grades, whilst peer assessment denotes grading (irrespective of whether comments are also included. Whether grades are awarded or not, the emphasis is on standards and how peer interaction can lead to enhanced understandings and improved learning" (Liu and Carless 2006: 280).

Although both are important, a combination of the two is often suggested. Nonetheless, the use of peer assessment or peer feedback depends on the focus of the task and the results that one wishes to obtain. For instance, if teachers want their students to simply receive suggestions or remarks on their performance (e.g. on a draft, before handing in the final version), then peer feedback seems more indicated. Instead, if teachers expect their students to be able to also judge whether their peers' work is up to

the standards of the course, then peer assessment seems to be the best option. More specifically, Falchinov (2007: 132) observes that peer assessment "on a product or a performance, should be based on the criteria of excellence for that product or event". Therefore, students who assess their peers' work, should always judge against the criteria of excellence. These criteria of excellence, that is those criteria, which establish the 'desired' task accomplishment, should be pointed out by the teacher.

Although many scholars claim that asking students to provide grades is challenging for different reasons, according to Topping (2005: 640), giving qualitative feedback is actually "much more cognitively demanding for the assessor" than giving grades. Indeed, students have to carefully 'judge' their peers' performance and provide useful suggestions to them. Nevertheless, qualitative feedback "is likely to be both more socially comfortable and more useful to the assessee" than grades. In spite of this, as Liu and Carless (2006: 287) further claim, "to advocate the abandonment of peer grading altogether is to ignore the centrality of marks to the whole student experience of assessment".

For the purpose of my dissertation, I will refer to 'peer assessment' as an activity which includes both the practice of providing feedback (peer feedback) and that of marking peers' performance (peer grading).

For peer assessment to be helpful and effective, careful planning and implementation, and clear objectives and criteria are required. Boud et al. (1999: 417) observe in this respect that special caution needs to be raised when it comes to the implementation of peer assessment, since "assessment can easily inhibit the processes it is designed to enhance if it is not implemented sensitively". With regard to assessment criteria, as Herman et al. (1992: 44) observe, "in the absence of criteria, assessment tasks remain just that, tasks or instructional activities". Besides, criteria clarify instructional goals and, therefore, provide a sound rationale for the whole activity as well "guide students' judgments and make public to students, parents, and others the basis for these judgments" (Herman et al. 1992: 45). Clear criteria also help teachers (or students) to mark consistently, thus obtaining reliable and valid results.

Because peer assessment needs careful planning and implementation, it can be very time-consuming and demanding, especially for teachers. Nonetheless, the time that one might 'lose' in this phase (planning), will be made up for once both teachers and

students become acquainted with and master the technique of peer assessment. Peer assessment is time-consuming and demanding for students too, in that they have to think, compare, contrast and communicate (Topping 1998). However, this increased time that students have to spend on a task will help to "consolidate, reinforce, and deepen understanding in the assessor" (Topping 1998: 254).

Evidently, another necessary aspect to be taken into consideration when it comes to peer assessment is that teachers need to cultivate a collaborative and non-threatening environment, in which students' contributions are highly valued and cherished: "the creation of a course climate in which the giving and receiving of peer feedback is a normal part of teaching and learning processes" is highly recommended (Liu and Carless 2006: 288). On the contrary, if students feel that they are being judged and that their work is being brought into question, then they will probably feel resentful and they will not be willing to explore the potential benefits of peer assessment. Additionally, students may also lack faith in their peers' ability to assess them and those assessing may feel inadequate and lack confidence in their own ability.

The next section will be devoted to the discussion of the exceptional outcomes that could result from adopting peer assessment as an upturning educational method, as well as the demystification of those popular issues, which are often raised against alternative forms of assessment, namely objectivity and reliability.

2.3 Advantages and Drawbacks of Peer Assessment

Many scholars and educational theorists (see, for instance, Herman et al. 1992, Topping 1998, Boud et al. 1999, Cameron 2001, Little 2003, Brown 2004, White 2009, Peng 2010, Spiller 2012, Novello 2014, Jung 2016) have listed numerous positive outcomes that may result from the adoption of peer assessment as an ordinary assessment procedure. To name a few, increasing motivation, enhancement of ownership of the assessment process and, therefore, of learner autonomy, deeper learning, fairness (being assessed by more people) and development of transferable skills (such as cooperative skills, metacognitive skills, self-regulation skills, communication skills, critical enquiry and reflection). Indeed, as maintained by Boud et al. (1999: 417), "peer learning typically pursues learning outcomes traditionally hard to assess, for example, those related to group work, oral communication, planning and self-assessment".

Furthermore, it seems noteworthy to mention that through peer assessment "students can help each other to make sense of the gaps in their learning and understanding and to get a more sophisticated grasp of the learning process" (Spiller 2012: 11). Therefore, peer assessment also has positive repercussions on self-assessment[25], since students, by judging the work of others, are also likely to become more aware of their own skills and, consequently, improve their own performance. In this respect, Little (2003: 223) argues that

> "the capacity for private reflection grows out of the practice of public, interactive reflection, and the capacity for self-assessment develops partly out of the experience of assessing and being assessed by others".

In this respect, Boud et al. (1999: 421) also claim that, in peer assessment, "peers provide rich information which is then used by individuals to make their own assessments".

As already mentioned in Chapter One, one of the many benefits of using alternative forms of assessment is that these can enhance learner autonomy. More specifically, peer assessment seems to be one of the best suited procedures to encourage and foster learner autonomy, together with self-assessment. It seems, indeed, that students develop very important skills when they are asked to evaluate their peers' performances, such as critical thinking and reflection, which are likely to facilitate autonomy and lifelong learning. With this in mind, as has already been stated, peer assessment is also essential to foster self-assessment. In this respect, Little (2002: 187) observes that

> "if learners do not gradually develop the capacity to engage in reflection and evaluation in their second/foreign languages, we cannot expect them to progress to the more advanced levels of proficiency, which presuppose such a capacity".

For all of the above reasons, Peng (2010: 89) claims that peer assessment has "significant pedagogical value", because it "enables learners to take part in the evaluation process and gives learners opportunities to participate in and evaluate their peers' learning process and products".

[25] See Chapter One, pp. 21-22.

Proof that students are aware of the benefits that result from peer assessment lies, for instance, in a study by Gan (2011 in Hattie 2012). Hattie (2012: 150) mentions that, in this case, students regarded giving and receiving feedback as a

> "potentially enriching experience, because it allowed them to identify their learning gaps, collaborate on error detection and correction, develop their ability to self-regulate, including monitoring their own mistakes, and initiate their own corrective measures or strategies".

Table 2 below, taken from Hansen-Edwards (2013), summarizes some of the main benefits and drawbacks of peer assessment.

Table 2: Benefits and drawbacks of peer assessment. Taken from Hansen Edwards (2013).

Areas	Potential benefits	Potential drawbacks
Metacognitive/ Cognitive	○ More time on the task, on thinking, reviewing and summarizing. ○ Greater understanding of the nature and process of assessment. ○ Development of autonomy. ○ Higher order thinking processes.	
Time	○ Saves teacher's commenting time	○ Requires time for training students.
Affect	○ Increases motivation. ○ Develops student's ownership of the assessment process.	○ Unwillingness to assess peers, especially if friends. ○ Students may not have enough confidence in their own language skills to give feedback. ○ The teacher may be perceived to be the one responsible for giving feedback.
Feedback	○ Greater quantity of feedback. ○ Faster feedback. ○ Triangulation of ratings/feedback if self- and teacher assessment is also used.	○ Peer assessment may not be accepted as accurate, reliable, and professional. ○ Students may question the accuracy of grading as well as linguistic, rhetorical, and content feedback coming from their peers.

Social interaction	○ Increased negotiation skills. ○ Collaboration skills. ○ Active learner roles.	
Linguistic development	○ Development of verbal communication skills. ○ More opportunities for L2 use.	○ Students may not have linguistic knowledge to comment on grammar and may not know how to express feedback.

In spite of all this, peer assessment might also entail some drawbacks, most of them being the same disadvantages of all the other alternative assessment procedures that were discussed in Chapter One, namely: validity, reliability, and objectivity. In addition to these, however, peer assessment poses several other issues due to its 'mutual' and 'collaborative' nature[26]: while other alternatives mainly focus on the student only (self-assessment, diary and portfolio writing, and observation), peer assessment focuses especially on the interaction between students, besides focusing on the student's own performance. Indeed, in peer assessment, students are required to assess their peers work and share their observations with each other. Often, this fact raises some issues. For instance, as maintained by White (2009: 8), some students might express a fear of "possible social embarrassment (especially concerning identifying weaknesses in the work of peers) and the fact that PA[27] may be cognitively challenging and straining". Moreover, concerns are also often raised about the potential bias of peer assessment, in that "reliability of scoring may be affected by student bias caused by the relationship between the assessor and the person being assessed" (White 2009: 22).

Another issue, which is often raised against peer assessment, relates to the inexperience and inadequacy of students to assess their peers. More specifically, how could learners, who are still in the process of acquisition, be capable of making accurate and thorough assessments of their peers' performance? Indeed, past evidence (see, for instance, Cheng and Warren 2005; Peng 2010; Jung 2016) has shown that in some cases students expressed dislike in criticizing friends and getting arbitrary markings. Moreover, some were afraid to give wrong scores to their peers, due to their own perceptions of inability and lack of competence and qualification to assess their peers correctly.

[26] "Peer assessment is fundamentally a collaborative assessment practice" (Double et al.'s 2019).
[27] Peer Assessment.

Liu and Carless (2006) also discuss the issues regarding peer assessment. According to the two experts, resistance to peer assessment on the part of students is often voiced in terms of reliability, perceived expertise, power relations, and time. The issue of reliability has already been discussed in Chapter One. As regards perceived expertise, students often feel inadequate in their role as assessors and think that only teachers "possess the necessary knowledge and expertise to conduct reliable assessment" (Liu and Carless 2006: 284). Power relations, instead, concern the idea that "to assess is to have power over a person" (Brew 1999 in Liu and Carless 2006: 285). With this in mind, students might dislike assessing their peers' performances (as well as having their performance assessed by their peers) because they do not want to have power over their peers, or their peers having power over them. Last but not least, as already discussed, time is also an issue in peer assessment, which "may act as a discouragement" toward implementing this activity.

The next section will be devoted to possible solutions that have been offered by educational experts to the above-mentioned drawbacks of peer assessment.

2.4 Experts' suggestions and proposals to overcome peer assessment's issues

To overcome the above-listed drawbacks of peer assessment, numerous suggestions and proposals have been provided by different educational experts. For example, Brown (2004) suggests that teachers provide their students with some guidelines to ensure successful peer learning experiences. Besides, Brown (2004) also argues that telling students the purpose of the assessment, defining the tasks and what is expected from students clearly, and encouraging impartial evaluation of performance would also help students to overcome their fears.

As mentioned above, criticisms involving reliability, validity and objectivity are often raised when it comes to peer assessment. According to Peng (2010: 89), however, this is mainly due to the fact that many teachers are still not aware that "reaching high validity and reliability is not the main goal of peer assessment". On the contrary, as observed by Peng (2010), the functions and roles of peer assessment and teacher assessment are very different, and the former should function, therefore, simply as a "support tool" (Hansen Edwards 2013: 732) to the latter: indeed, peer assessment is a "supplementary assessment method for involving and empowering students rather than a

substitution for teacher assessment" (Peng 2010: 90). Topping (1998: 262) similarly argued that peer assessment should be regarded merely as an added value to the learning process:

> "the incorporation of peer assessment as a learning tool besides the usual teacher assessment not only can change learners' perspective toward various types of assessment but may also lead to outcomes at least as good as teacher assessment and sometimes better".

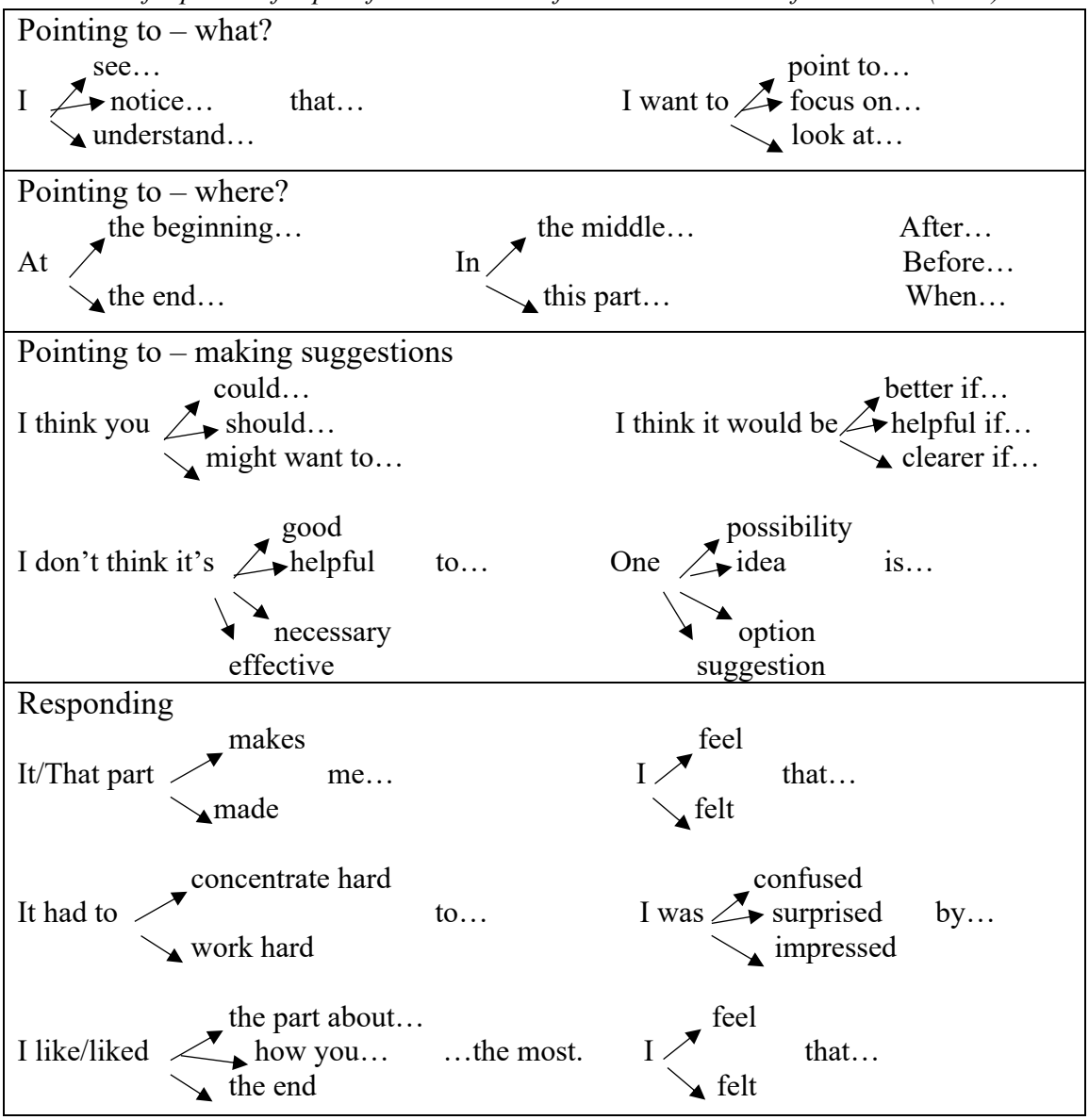
As mentioned above, many scholars have put forward the suggestion to adopt peer assessment as a tool that can support more traditional forms of assessment (i.e. teacher assessment), rather than as a radical substitution of these. Moreover, at least in the initial stage, peer assessment may be very strictly controlled by teachers: although these new forms of assessment have shifted the focus from teachers to learners, teachers still play a very important role, in that it falls upon them to establish "a learning environment where teachers and learners are jointly responsible for the outcome" (Dam 2003: 135).

To overcome the above-mentioned issues of inexperience with assessing peers, Keller and Westfall-Greiter (2014: 9) argue that it is essential to explain to the students what feedback is (or should be), namely: "concrete and specific, focused on quality, helpful, user-friendly (clear and easily understandable), relevant, and respectful". Moreover, they point out that useful phrases (see *Table 3* below) for peer feedback could be handed out to students, so that they know on which aspects of their peers' performance they should focus. Indeed, it seems that "especially at the beginning, learners of a foreign language need these readily accessible language 'chunks'[28] as phrase builders to communicate quickly and effectively". Therefore, providing handouts with useful phrases for peer feedback along with the task will help students carry out more thorough and detailed assessments on their peer's work.

As stated in Chapter One, it is important that feedback is specific rather than general so that students believe that it is relevant for them.

[28] Michael Lewis in the early 1990s develops the *Lexical Approach*, according to which the most important part of foreign language acquisition lies in the ability to master and to produce the so-called "lexical chunks": "instead of words, we consciously try to think of collocations, and to present these in expressions. Rather than trying to break things into ever smaller pieces, there is a conscious effort to see things in larger, more holistic ways" (Lewis, 1997: 204 in Keller and Westfall-Greiter, 2014: 10).

Table 3: Useful phrases for peer feedback. Taken from Keller and Westfall-Greiter (2014).



Similarly, Hattie (2012: 144) also suggests that teachers provide students with prompts for feedback. Hattie (2012) refers to many forms of prompts: organizational prompts, elaboration prompts and monitoring progress prompts, which can be found at all levels of the performance: task, process and self-regulation[29] (see *Table 4* below). Prompts by teachers may take different forms, such as "guiding questions, sentence openers, or question stems that provide cues, hints, suggestions, and reminders to help

[29] See Chapter One, p. 13.

students to complete a task" (Hattie 2012: 147). More specifically, prompts can serve two functions, namely scaffolding and activation:

> "prompts act as scaffolding tools to help learners by supporting and informing their learning processes. Prompts can be designed to target procedural, cognitive, and meta-cognitive skills of the learner; they can provide new or corrective information, invoke alternative strategies already known by the student, and provide directions for trying new learning strategies. In this sense, prompts can be conceived as 'strategy activators'" (Hattie 2012: 148).

Table 4: Prompts to help students give appropriate feedback to peers. Taken from Hattie (2012).

Level of prompt	Examples
Task	o Does his/her answer meet success criteria? o Is his/her answer correct/incorrect? o How can he/she elaborate on the answer? o What did he/she do well? o Where did he/she go wrong? o What is the correct answer? o What other information is needed to meet the criteria?
Process	o What is wrong and why? o What strategies did he/she use? o What is the explanation for the correct answer? o What other questions can he/she ask about the task? o What are the relationships with other parts of the task? o What other information is provided in the handout? o What is his/her understanding of the concepts/knowledge related to the task?
Self-regulation	o How can he/she monitor his/her own work? o How can he/she carry out self-checking? o How can he/she evaluate the information provided? o How can he/she reflect on his/her own learning? o What did you do to…? o What happened when you…? o How can you account for…? o What justification can be given for…? o What further doubts do you have regarding this task? o How does this compare to…? o What does all of this information have in common? o What learning goals have you achieved? o How have your ideas changed? o What can you now teach? o Can you now teach another student how to…?

Having said this, Hattie (2012) also points out that such prompts can also be used by students to monitor and reflect on their own learning, and not only on that of their

peers. In this case, prompts can therefore be very useful tools also to foster learner autonomy and lifelong learning.

Another helpful way of controlling peer response is through the use of checklists. Herman et al. (1992: 64) define checklists as lists of "dimensions, characteristics, or behaviors that are essentially scored as 'yes-no' ratings". These dimensions can be ticked to indicate whether a characteristic or behavior was present or not. Furthermore, checklists differ from rating scales, in that they usually contain more dimensions, which are nonetheless "quite narrow and concrete" (Herman et al. 1992: 64). On the contrary, as can be noted in Chapter One, rating scales usually contain more elaborate and detailed descriptors.

Although there appears to still be some hesitation among teachers towards non-traditional forms of assessment, more and more educational experts seem to acknowledge that both the teaching and the learning process can significantly benefit from a consistent employment of peer assessment in the class. Indeed, alternative assessment in general (and, therefore, also peer assessment) provides useful diagnostic information about learners' learning ability, which traditional forms of assessment may fail to provide, such as students' pace of learning or the learning strategies that work best for them in accordance with their learning styles.

As briefly mentioned above, one of the main weaknesses of alternative forms of assessment is generally assumed to be objectivity. Not only teachers, but also students are often sceptical about peer assessment in that they believe that evaluations made by teachers are more accurate, which, however, may not always be true. Besides, as already pointed out in Chapter One, all assessments can be subject to erroneous judgements, irrespective of their nature (teacher assessment, peer assessment, self-assessment…).

Peng (2010: 91) claims that in peer assessment two possible situations may occur: "one is that students may either be too critical or too flattering. The other is that they simply do not know how to make an adequate assessment". Nevertheless, many researchers and academics (to name a few, Freeman 1995; Topping 1998; Patri 2002; Cheng and Warren 2005) suggest that learners simply need appropriate training and practice in peer assessment to achieve objectivity. Indeed, through practice, learners will presumably become more accurate in their assessments. Spiller (2012: 10) likewise claims that "students need practice to gain confidence and become more competent at

peer assessment". This view is also supported by Keller and Westfall-Greiter (2014: 4), according to whom "the more students implement assessment procedures, the more they will also recognize the benefits of using them"[30]. Through familiarity with the evaluating criteria and assessment practice, students are, indeed, expected to have greater evaluating skills and, thus meet the goals of peer assessment (i.e. assessing their peers' work effectively and proficiently).

As stated in the previous section, concerns are also often raised about the potential bias of peer assessment. However, as evidence shows, the agreement between teacher and student grading is often very close (Donato 1994; Topping 1998; White 2009; Peng 2010; Gan 2011; Spiller 2012**;** Jung 2016). In this regard, Donato (1994: 51) observes that "learners are capable of providing guided support — feedback — to their peers during collaborative L2 interactions in ways analogous to expert scaffolding", thus demonstrating that students are often more suited to this type of activity than they themselves think. Nevertheless Kroll (1991: 259 in Peñaflorida 2002: 351) observes that

> "because ESL[31] students lack the language competence of native speakers of English who can react instructively to their classmates' papers, peer responding in the ESL classroom must be modeled, taught, and controlled in order for it to be a valuable activity."

As mentioned earlier, there are numerous ways of guiding and controlling peer feedback, which can facilitate the process and increase the trustworthiness of peer assessment. Among those that have already been mentioned, Topping (1998) adds 'providing quality training' as suggestion to implement peer assessment correctly and to encourage objectivity. According to Topping (1998: 266), quality training should cover several aspects of peer assessment, such as:

> "objectives, general organization, developing and using criteria and any associated materials, sustaining an effective group process, giving and receiving positive and negative feedback in different forms, action in response to feedback, and arrangements for evaluation".

The next section will be devoted to a review of some of the most important studies on peer assessment, which will also partially serve as a basis for my own study on peer assessment in the next chapter.

---

[30] Translated from German.
[31] English as a Second Language. This claim fits EFL (English as Foreign Language) students too.

2.5 Studies on Peer Assessment

An extensive body of research related to the study of peer assessment is available (see, for instance, Freeman 1995; Topping 1998; Boud et al. 1999; Falchinov 2000; White 2009; Peng 2010; Spiller 2012; Jung 2016). Most peer assessment literature, though, is concerned with two issues in particular: evaluating students' perceptions of peer assessment procedures and the extent to which this alternative form of assessment can be credible and sustainable in terms of objectivity, reliability and validity. As Liu and Carless (2006: 282) add, "the existing literature on peer assessment has also been dominated by studies of peer-tutor grade correlations".

This section aims at summarizing the outcomes of some important studies on peer assessment, starting with a study by DeNisi et al. (1983) on the effects of peer evaluations on a group performance. The studies will be reviewed in chronological order (oldest to newest), and the rationale for mentioning each one of them in my dissertation will be pointed out, as well as the extent to which they have/will influence(d) my own study on peer assessment.

In the studies that have been reviewed in the next sections, different skills are being assessed. More specifically, the studies by Mendonça and Johnson (1994), Azarnoosh (2013), Jung (2016) assess writing skills (i.e. written compositions), whereas in the studies by White (2009) and Peng (2010) speaking skills were taken into consideration (i.e. oral presentations). Furthermore, two meta-analysis have been reviewed, namely by Falchinov and Goldfinch (2000) and by Double et al. (2019), which did not focus entirely on either written or oral skills.

Two studies, namely DeNisi et al.'s (1983) and Mendonça and Johnson's (1994) only took peer feedback into consideration, whereas others, such as Falchinov and Goldfinch's (2000), White's (2009), Peng's (2010), Azarnoosh' (2013) and Jung's (2016) focused on peer grading.

2.5.1 A study of Peer Assessment by DeNisi et al. (1983)

DeNisi et al. (1983) carried out a study predicting that there would be "positive effects for individuals who learn that their peers have rated them positively […] and negative effects for individuals who learn that their peers have evaluated them negatively" (DeNisi et al. 1983: 458).

The participants were 143 undergraduate students, who were divided into small groups of 3 to 5 people (for a total of 34 small groups) and were told that they would have worked on two tasks as a group. Moreover, after each task, they were asked to fill in a questionnaire, in which they were asked to evaluate the other members of the group by name. More specifically, they had to evaluate the following aspects: satisfaction with the contribution of the other group members, group cohesiveness, perceived performance and peer ratings using a 7-point scale. After each subject completed the first questionnaire (after task 1), average peer rating feedback was given, but all peer feedback was false. Indeed, it had been manipulated by the researchers, who then randomly assigned the groups to either positive (17 groups) or negative feedback (the remaining 17 groups). As has been mentioned earlier in this section, the hypothesis of the researchers was that "knowledge of peer ratings would affect group member interactions, perceptions, performance, and subsequent peer ratings and that the nature of these effects would depend on the sign of the peer ratings" (DeNisi et al. 1983: 461). Indeed, evidence showed that knowledge of peer ratings did affect the aspects that were mentioned before (group cohesiveness, peer ratings…). Students who received (false) positive feedback, raised scores on all of these aspects for task 2, whereas students who received (false) negative feedback lowered their scores.

Although this study seems to suggest that peer feedback can have serious consequences on the reliability of subsequent feedback, several limitations to this study, as the researchers themselves suggest, should be taken into consideration. First and foremost, obvious limitations derive from the manipulation of peer feedback by the researchers, who gave really low average scores to the negative feedback groups: however, "rarely would one peer rate another so poorly" (DeNisi et al. 1983: 463). Moreover, it was argued that negative feedback might actually motivate students to 'do more' to improve future evaluation, rather than be a cause of discouragement. Lastly, the time factor has also probably influenced the results of this study, and further research is needed over longer periods of time to study the eventual deterioration of group members relations or, on the contrary, to find out that group members have "become accustomed to poor peer ratings and ignore them completely" (DeNisi et al. 1983: 463).

This study was one of the first to be conducted on peer assessment; from there, several other studies of peer assessment have been carried out. Nevertheless, DeNisi et

al.'s (1983) research have inspired my own investigation of students' perceptions of peer feedback and peer assessment.

2.5.2 A study of Peer Assessment by Mendonça and Johnson (1994)

Mendonça and Johnson (1994) conducted a study on peer interactions occurring during peer review. According to Kroll (1991 in Mendonça and Johnson 1994: 746), "peer interactions that occur during peer reviews represent an important component of effective L2 writing instruction". As stated earlier in this chapter, peer assessment is a valuable form of assessment for assessing all aspects of the language, but especially for assessing writing.

The subjects of this study were a cohort of 12 advanced non-native speakers of English, who were enrolled in a writing class and whose fields of study were very diverse, including education, mechanical engineering, and agriculture. The students were matched by the two researches and six pairs were created. Four out of six pairs consisted in students from the same field of study, whereas the remaining two pairs were formed by students of different fields of study. Peer review (i.e. peer feedback) on the writing assignment was communicated exclusively orally. The aim of the review was to provide feedback to the drafts, so that students could then adjust them and write a second revised draft of the assignment. The students were given guided questions by the teacher (see *Table 5*) to provide students with guidance during peer feedback. At the end of the activity, students were encouraged to make their own decisions on whether or not to incorporate their peer's comments in the revised draft.

Table 5: Guided questions for peer feedback. Taken from Mendonça and Johnson (1994).

1. Before starting the peer review, explain to your partner what your paper is about.
2. What is the main idea of your partner's paper?
3. Is there any idea in his/her paper that is not clear?
4. What suggestions could you give to your partner?

A second aspect of this study consisted in a post-interview: students were individually interviewed by the two researchers so as to understand, whether they found peer reviews helpful or not and whether they used their peer's comments in their final draft.

The first and second draft were compared by Mendonça and Johnson for evidence that students had incorporated peer's comments and labels, such as R/PR (Revised/Peer Review), NR/PR (Not Revised/Peer Review) and R/NPR (Revised/Not in Peer Review) were placed in the second draft.

Results from the comparison showed five different types of interactions[32] between students: question, explanation, restatement, suggestion and grammar correction. Moreover, in 'question', the two researchers found requests for explanation and comprehension checks, and the 'explanations' were concerned with unclear points, opinions and content. In total, 578 interactions were found in the peer reviews. Of these 578 interactions, grammar correction was the least used one (only twice).

In *Table 6* below, some representative examples taken from Mendonça and Johnson (1994) of several of these interactions are presented, such as 'request for explanation', 'comprehension check', 'explanation of content' and 'suggestion'.

Table 6: Examples of interactions during peer review (Mendonça and Johnson 1994).

○ Lee: I don't understand pesticides. [*request for explanation*] ○ Sandra: Oh, you don't know what is a pesticide? [*comprehension check*] ○ Sandra: It's something that you put eh... that farmers use to put in the crop... to kill... [*explanation of content*]
○ Gang: Language minority parents. Eh, I can't understand what is mean ... what is means... language minority parents. [*request for explanation*] ○ Rosa: Language minority parents? [comprehension check] I am talking about parents that have their first language is other than English. [*explanation of unclear point*]
○ Pablo: Yeah, so yeah... maybe you can add some more things in some sections or just delete the sections and put all together. [*suggestion*] ○ Jean: I think this is also a kind of transition between each other... giving the topic... this is another... [*explanation of opinion*]

In the next phase, Mendonça and Johnson compared the two drafts, and found three scenarios:

> "(a) students used what had been discussed during the peer reviews, (b) students did not change a given part of their written texts even though part had been discussed in the peer reviews, and (c) students a given part of their texts without input from peer review."

[32] "Negotiations" (Mendonça and Johnson 1994: 751).

Evidence of non-revisions possibly suggest that "sometimes students may not trust their peers' comments and thus not use them as they revise; however, they may do so when the teacher makes similar comments" (Mendonça and Johnson 1994: 762).

Last but not least, in post-interviews Mendonça and Johnson found that a significant majority of students perceived peer feedback as useful and beneficial. Only two students stated that they did not find peer review useful, but only because they had been matched with students from different fields of study. Instead, the remaining 10 students reported that "even though peers in the same field of study can offer more ideas, peers who are in different fields are better able to pinpoint parts that are not clear in their drafts" Mendonça and Johnson (1994: 765).

Some limitations to this study concern the small sample size (only 12 students) that may hinder generalizations. Furthermore, only advanced students participated in peer reviews; perhaps, beginners or intermediate students would give different answers and less diverse interactions would occur. In spite of these limitations, this study will hopefully be useful for my own study, and especially the identification of different types of interactions. Moreover, post-interviews will also be taken into consideration in my study with the same purpose of Mendonça and Johnson's, in the form of a written questionnaire.

2.5.3 A study of Peer Assessment by Falchinov and Goldfinch (2000)

Falchikov and Goldfinch (2000) conducted a meta-analysis of 48 quantitative peer assessment studies that compared peer and teacher marks, and demonstrated that students are able to make reasonably reliable judgements, for the most part. The two researchers collected over 100 studies from the following databases: *Bath Information Data Service* (BIDS), *Educational Resources Information Center* (ERIC), *MEDLINE*, *Psychinfo*, *Socinfo*, and *FirstSearch*. All of the studies had been performed in higher education settings and included quantitative evaluation (through percentages or numerical grades or letters). Moreover, the studies were classified according to several "independent variables" (Falchikov and Goldfinch 2000: 291), such as study identifiers, population characteristics, what is assessed, the level of the module/course, how the assessment is carried out and the criteria used, the design quality and the number of peers involved.

On average, Falchinov and Goldfinch (2000) found clear evidence of agreement between peer and teacher marks. More thorough analyses, however, suggested that some variables affected the agreement between peer and teacher assessment and the validity of the results. To recap, validity is defined by Gronlund (1998: 226) as "the extent to which inferences made from assessment results are appropriate, meaningful and useful in terms of the purpose of the assessment". For example, analytical evaluation proved to be less valid than holistic evaluation. Furthermore, the nature of the task also seemed to influence the validity of peer assessment: "peer assessment of academic products and processes seems to correspond more closely to faculty ratings than peer assessment in the context of professional practice" (Falchinov and Goldfinch 2000: 315). Well-designed studies also resulted in higher peer-teacher agreement, whereas it seemed that the level of the course (advanced/introductory) did not affect the validity of peer assessments, in general. Last but not least, it was noted that "student familiarity with, and ownership of, criteria tends to enhance peer assessment validity" (Falchinov and Goldfinch 2000: 315), thus confirming what has been stated numerous times in this chapter about providing students with clear assessment criteria.

Among the limitations to this study, Falcinov and Goldfinch (2000) include repeated experience of peer assessment which, in this case and for obvious reasons, was absent. Indeed, as Falchinov (2003) later points out, replication with same cohort is an essential part of peer assessment, because it helps to monitor the development of student peer assessment. Then, the two researchers suggest investigating gender differences in peer feedback, as "gender effects are present in a wide variety of social and academic situations, and it is possible that they may also play a part in peer assessment" (Falcinov and Goldfinch 2000: 318). Last but not least, bias investigations ("investigations into friendship or enmity effects" Falcinov and Goldfinch 2000: 318) are also suggested.

This study has been of great inspiration for my own study; more specifically, bias investigation will be an aspect I will focus part of my study on, and will be carried out via a questionnaire.

2.5.4 A study of Peer Assessment by White (2009)

As already stated in this chapter, peer assessment can be carried out on most aspects of the language, although researchers agree that it is particularly effective for

assessing writing. Nonetheless, this study by White (2009) was conducted among a group of 55 female students between the age of 20-21 of a public speaking class in Tokyo[33].

As a course requirement, each student was required to do two short presentations of about 8-10 minutes. Before making the final presentations, though, students were divided into small groups of about 4 people, where they had to do mini presentations to prepare for their final ones. In these groups, students also had to report on their own progress and get feedback from the other members of the group. To give feedback, students were handed out a 14-points scale, designed by Yamashiro and Johnson (1997)[34], with the following criteria: voice control, divided into projection, pace, intonation and diction, body language, divided into posture, eye contact and gesture, content of oral presentation, divided into introduction, body and conclusion and, lastly, effectiveness, which was divided into topic, language use, vocabulary and purpose. According to White (2009: 15), the mini-presentations and subsequent group feedback were meant to help

> "students prepare for their performances, but also served as training sessions in the use of the assessment criteria. It was hoped that such use of the 14 points in mini-presentations (and class activities) would serve to help students internalize the key assessment criteria to be used".

The final grade would then consist in teacher assessment (60%), peer assessment (30%) and self-assessment (10%). Therefore, "almost half (40% in total) of the final course grade was based on student-generated assessment input" (White 2009: 15).

Lastly, a survey was conducted to gather students' perceptions on peer assessment. Overall, survey data showed that a vast majority of students (96%) had a positive perception of peer assessment: assessment items were found to be clear and easy to follow, peer ratings were found to be fair and reasonable, and assessing peers proved to be helpful for own presentation planning. Moreover, especially comments were found to be useful and effective to make a better second presentation. In spite of this, a significant minority (36%) expressed a dislike for peer assessment, in that they found that their scores had been influenced by relationships with presenters. In addition, some were uncomfortable with having peers assess their performance. All in all, however, the positive aspects of peer assessment were noted by most students and results showed that students did indeed benefit from their peers' comments.

---

[33] Tokyo Woman's Christian University (TWCU).
[34] *Public Speaking in EFL: Elements of Course Design*.

Even though this study was conducted on a speaking task, rather than a writing task, I have taken inspiration for my own questionnaires from the survey structure and the particular questions that it sought to answer.

2.5.5 A study of Peer Assessment by Peng (2010)

A study by Peng (2010) sought to explore college EFL students' perceptions of peer assessment in relation to their proficiency level as well as score agreement between teacher- and peer assessment. Like in the study by White (2009), this study was centred on group oral presentations. Participants were 43 high-intermediate (HI) students of English and 45 low-intermediate (LI) students of English, enrolled in different faculties at a University in northern Taiwan. All participants had been trained to evaluate their peers beforehand: they had had the opportunity to discuss assessment criteria and had designed peer feedback forms. A written survey to explore students' perceptions of peer assessment was conducted before and after the group presentations using a 5-point Likert scale[35]. The survey was conducted in Chinese, so as to avoid language barriers. Moreover, semi-structured interviews were also made to explore more in-depth participants' opinions in regard to peer assessment: "the data obtained from the interviews provided in-depth information that the five-point Likert scale survey could not offer" (Peng 2010: 92).

Participants were asked to listen to the group presentations and write down any comments on it. Then, as a group, they were asked to discuss the performance and assign a grade to the performers. Furthermore, after each group had presented, they were asked to evaluate the other group members using a within-group peer assessment form.

Results from the surveys showed that perceptions on peer assessment from both group of students (LI and HI) became significantly more positive after the presentation: students recognized peer assessment as a valuable activity to improve oral skills, since it created an opportunity to converse in English, group participation and interaction skills. Moreover, on average, students were considered eligible to assess their peers. Nevertheless, some students felt that their judgement was biased: "it's hard to evaluate

[35] Likert scales (1932) consist in several statements, to which points are assigned, which should cover most aspects of what is being surveyed. Respondents need to state their degree of agreement or disagreement with each statement. For example: 'my family should ban meat from their diet' (strongly disagree 1 2 3 4 5 totally agree).

my friends. I feel obligated to reward them good grades", or "sometimes peer assessment is subjective", or "people may have different standards; for instance, some peers may give me 80 points thinking that it's a good grade while I think otherwise" ((Peng 2010: 99).

As to language proficiency differences in terms of attitudes toward peer assessment, data from the pre- and post-surveys showed that in the pre-surveys, the HI students reacted less favorably to peer assessment than LI students. Nevertheless, the situation was reversed in the post-surveys. All in all, there were no significant differences in attitudes between the two groups of students. Furthermore, in terms of grades, there were also no significant differences between the two groups. With all this in mind, the results of this study were at odds with others, since teacher grading was consistently higher than peer grading. It is generally thought, instead, that usually students tend to over-mark their peers.

This study was particularly interesting in its purpose to also investigate proficiency differences in attitudes toward peer assessment and grading.

2.5.6 A study of Peer Assessment by Azarnoosh (2013)

A study by Azarnoosh (2013) was conducted to answer to the following research questions: what is the correlation between peer and teacher ratings? And, does friendship affect peer assessment?

Participants to this study were 26 students from a University in Iran, who had to write an essay for their course. To score the compositions, a revisited scale by Jacobs et al. (1981 in Hughes 2003) was used, which considered five criteria and 4 points each (for a final grade of 20), namely: content, organization, vocabulary, language use and mechanics. It was, therefore, an analytic grading scale, which students were introduced to and with which students practiced before the start of the activity. Pre- and post-questionnaires were also distributed to compare the answers to several questions regarding the peer assessment process adopted in this study. To identify friendships in the classroom, students were asked to write down three names of close friends within the classroom; these, were then used to match students during peer assessment. Indeed, students were evaluated by more than one peer, one being a friend, and the other not being one. In addition to peer raters, the teacher (researchers) and two other raters (EFL instructors) assessed the papers.

Students had the opportunity to choose between three topics, which were offered in class, and submit their written assignment by the following week.

Rating findings (see *Table 7* below) showed that there were no relevant differences between peer and teacher ratings, and "the mean scores for corrections were quite close to each other" (Azarnoosh 2013: 6).

Table 7: Peer and Teacher ratings (mean). Taken from Azarnoosh (2013).

Comparisons	Mean
Friend and non-friend corrections[36]	17.18
Teacher corrections	17.36
Friend correction	17.37
Non-friend correction	17.00

As to the questionnaire results, data showed significant changes in attitudes toward peer assessment before and after the activity: while in the pre-questionnaire some students seemed reluctant and doubtful toward the activity, in the post-questionnaire the same students seemed to have changed their mind, and actually stated that they found peer assessment helpful and motivating.

Last but not least, friendship did not seem to affect peer ratings, which are quite similar between friends and non-friends. Nevertheless,

> "although students named their intimate friends, they did not deny their overall friendship with others who had been their classmates for at least 2 years, so this might have affected their ratings unconsciously" (Azarnoosh 2013: 7).

In spite of this obvious limitation to this study, the idea of trying to identify possible friendship biases in assessment results seemed very interesting and well-conceived, and I would like to also address this issue in my own study.

2.5.7 A study of Peer Assessment by Jung (2016)

A study by Jung (2016) aimed at investigating differences between peer- and teacher-assessment in short essays of about 600 words. The study was carried out at a University in South Korea, where 104 written essays from 26 students were collected. Students proficiency ranged from high (6) to intermediate (14) to low (6). Questionnaires

[36] Raters were considered as a whole (friend and non-friend corrections) as well as individually (friend corrections and non-friend corrections).

were also handed in to students, to detect those criteria that students found most difficult to assess. Assessment criteria for assessing the essays were five, namely: content, coherence and logicality, structure and organization, grammatical accuracy, and range of vocabulary.

Four assessment approaches were proposed in the study "in order to find out how each type of assessment operates to help students develop their L2 writing" (Jung 2016: 3): a 'free assessment approach' (self-assessment), a 'controlled assessment approach' (peer- or group-assessment), a 'guided assessment approach' (teacher-assessment), and an 'integrated assessment approach' (peer/teacher-assessment). More specifically, students were offered the opportunity to carry out self-assessment on their written assignments, before sending it to their peer and teacher for evaluation. Before writing their short essays, students were instructed on how to write a good essay. Participants were also provided with topics and guidelines for writing by their teacher, and were matched for peer assessment according to their level of written English.

Students wrote two essays for peer-review and teacher-review (essay one to be submitted before the mid-term exam, and essay two to be submitted after the exam):

> "after completing their essay, students exchanged their work by providing feedback following the given criteria and then they resubmitted their work to the teacher, and finally the teacher gave each student comments and marks on their piece of writing and students revised it accordingly" (Jung 2016: 10).

Table 8 below summarizes the stages of peer- and teacher review in the study by Jung (2016).

Table 8: Peer and Teacher reviews. Taken from Jung (2016).

Stages	Stage description
Pre-writing	When the teacher gave a guideline and assessment criteria about the essays to students who made a plan on how and what to write related to the articles.
Drafting	After composing a rough draft, students exchanged their work with their partner for peer-assessment from the given criteria.
Revising	After receiving peer-assessment with marks, the teacher asked students to do self-assessment, so students examined which parts they made mistakes or errors.
Editing	Students again resubmitted their essays to the teacher who corrected grammatical errors and other factors in their work as teacher assessment and finally the teacher returned the revised essays with marks to each student.

Results from the first essay show that students gave slightly higher scores (124.2/130) than the teacher did (122.2/130). On the contrary, results from the second essay show that teacher scores were slightly higher (126.5/130) than students' scores (125.1/130). More specifically, in the first essay, students gave their partner the highest score for 'range of vocabulary' and the lowest score for 'grammatical accuracy'. Instead, the teacher gave the highest score for 'content' and the lowest score for 'grammatical accuracy' as well. In the second essay, students gave their partner the highest score for 'content' and the lowest score for 'grammatical accuracy'. Likewise, the teacher gave the highest score for 'content' and the lowest score also for 'grammatical accuracy'. Lastly, questionnaire results showed that over half of the students (69%) thought that comments from peers were helpful to develop their writing skills and that assessment criteria were clear and well-designed. Nevertheless, some students thought they had received a much lower grade than they deserved, thus doubting their peers' adequacy to assess.

One possible limitation to this study is that assessment differences with regard to different proficiency levels have not been taken into consideration. In spite of this, this study is another example of how much students' writing can benefit from receiving comments by more than just one person (a teacher).

2.5.8 A study of Peer Assessment by Double et al. (2019)

Double et al. (2019) carried out a meta-analysis to evaluate "the effect of peer assessment on academic performance when compared to no assessment as well as teacher assessment". To be included in this meta-analysis, several criteria had to be met: all studies had to examine the effect of peer assessment, the assessment could be delivered in any form (written or spoken), all studies needed to compare the effect of peer assessment with a control group (experimental or semi-experimental studies) and all studies "needed to examine the effect of peer assessment on a non-self-reported measure of academic performance". Several electronic databases were searched for inclusion, such as *PsycInfo*, *Google Scholar*, and *ERIC*. In the end, 55 articles (studies) were included for eligibility in the meta-analysis. Different variables were extracted from the articles and listed in a table, such as publication type (conference, dissertation, journal, report), educational level, subject, grading, and anonymity. Data was analysed both in complete

form as well as for each educational level (primary, secondary, tertiary) and results were standardized for the sake of consistency.

In general, results showed a remarkable positive effect of peer assessment on academic performance[37]. Moreover, "these findings also suggest that that the benefits of peer assessment are robust across many contextual factors, including different feedback and educational characteristics". Having said this, it seems that feedback characteristics, such as anonymity and practice with peer assessment, that have always been deemed extremely important to achieve effective assessment, did not significantly "moderate the effect of peer assessment when analysed across studies". Last but not least, findings also indicated that "peer grading was beneficial for tertiary students but not beneficial for primary or secondary school students". Therefore, it seems that peer comments, rather than peer grading, holds a much bigger role in peer assessment.

With all this in mind, Double et al. observe that peer assessment is a valuable 'tool' for formative assessment, in that it can provide "new information about the learning process to the teacher or student, which in turn facilitates later performance (Pellegrino et al. 2001 in Double et al. 2019).

This study made me wonder whether higher education students that will be investigated in my study will be able to give sensible scores and comments to their peers' performance.

To conclude, this chapter was devoted entirely to peer assessment (in this dissertation peer feedback + peer grading). Advantages and disadvantages of this recent assessment method have been discussed in detail, as well as possible solutions to the disadvantages. Moreover, several studies on peer assessment have been reviewed, which will be relevant for my own study of peer assessment in the next chapter.

Chapter Three will be an in-depth analysis of questionnaire results by a cohort of second-year master students of foreign languages at the University of Padova, with which I sought to investigate perceptions of peer assessment.

[37] Double et al.'s considered academic performance as both "traditional outcomes (e.g., test performance or essay writing) and practical skills (e.g., constructing a circuit in science class)".

Chapter 3: Two Studies of Peer Assessment with EFL second-year Master Students and first-year Bachelor Students

"Alone we can do so little; together we can do so much." — H. Keller

Chapter Three will be divided as follows: the first part (3.1-3.5) will be an overview of the findings of a questionnaire on peer assessment that was distributed among a cohort of second-year master students of languages at the University of Padova. The aim of this questionnaire aimed to investigate students' perceptions of a peer assessment task they were asked to carry out during a *Spoken English* module. The module belonged to the course *Lingua, linguistica e traduzione inglese 2* of a language and literature degree course at the University of Padova, which is especially suited for students who are aspiring schoolteachers. The peer assessment was meant to raise students' awareness of what assessment entails and to make them more in control of the assessment process and, consequently, to encourage learner autonomy. A secondary aim of my study was to understand whether students who are aspiring schoolteachers would ever propose a peer assessment activity to their future students. These sections will be structured as follows: first, I will review the course aim and activities. Then, I will shortly describe the participants and the questionnaire. Lastly, questionnaire results will be discussed and possible limitations to the study will also be taken into consideration.

The second part of the chapter (3.6-3.11) will be an in-depth analysis of my own study on peer assessment with first-year bachelor students of foreign languages. The aim of this study was to investigate students' perceptions of peer assessment, as well as teacher/student rating agreement. This chapter will describe whether the students felt that peer assessment activity was helpful to prepare for the final exam of *General English* and if their peers' feedback was perceived as objective and adequate. In this case, as will be explained later in this chapter, peer assessment was conducted on writing skills. In addition, a secondary aim was to understand whether there are any differences in the perceptions of peer assessment between first-year bachelor students and more experienced second-year master students, and especially between those students who are aspiring schoolteachers.

As far as the methodology adopted in this study is concerned, the models were Mendonça and Johnson (1994), Peng (2010), and Azarnoosh (2013). In the case of Mendonça and Johnson (1994), students were given very precise guidelines for feedback, which I also gave to the students. In this study by Peng (2010), two questionnaires were distributed among the students to explore students' perceptions of peer assessment before and after the activity took place, and I also designed two questionnaires for the same purpose. Last but not least, as regards the study by Azarnoosh (2013), friendship biases were investigated, which inspired me to also investigate students' perceptions of biases in peer assessment. The starting point of this study was the presupposition that students, by judging the work of others and comparing it to their own performance, can improve their [writing] skills by identifying their "learning gaps" (Hattie 2012: 150), as well as become more involved in the assessment process.

The sections devoted to my project will be structured as follows: first, the study with the first-year bachelor students will be thoroughly explained and discussed, as well as the tools that were used to conduct the study, namely: a pre- and a post-questionnaire, a written task, and an evaluation grid, which were designed specifically for the sake of this study. Then, findings will be discussed and possible limitations to the study will be taken into consideration.

3.1 A study of peer assessment with second-year master students

3.1.1 The Course

The course was held during the winter semester 2019-2020 by Professor Katherine Ackerley. Students who wished to do the continuous assessment[38] (C.A.) had to attend the classes of the Spoken English module, in which peer assessment of students' presentations was adopted. The course focused on several aspects of spoken English, such as varieties of spoken English, teaching of spoken English, assessment of spoken English, phonetics and phonology, and pragmatics. Students were expected to read course materials before coming to class and be ready to discuss the topics. Moreover, they were also expected to complete tasks before, in, or after class, such as watching short videos

[38] An oral presentation and a written essay to be completed during the course, so as to receive a final grade by the end of February instead of doing the whole exam later in the academic year.

and answering to some related questions, explaining readings in class (flipped classroom reading), and, as briefly mentioned earlier, developing evaluating grids.

As regards assessment, students had to prepare and present a 10-minute presentation on one of the topics mentioned above, as well as write a 2500-word essay on a topic related to one of the aspects of the course. The presentation could be prepared and presented individually or in pairs, according to students' personal preference, but students were then assessed individually by both their teacher (Professor Ackerley) and two groups of peers. Before carrying out their assessments, students were asked to design both an analytic and a holistic grading scale, based on the 'can do' statements of the *Common European Framework of Reference for Languages* (CEFR). Students could choose to do so individually or in small groups and they could choose the parameters, evaluation criteria and the aspects of their peer/s performance they wanted to focus on. In the end, all students decided to work in groups, so the work was collaborative.

All rubrics were then posted in *Moodle* before the presentation session started. This way, students who had to present on a certain day knew exactly who were going to be the groups of peers assessing them in advance and, therefore, they also knew the evaluation criteria they were going to be assessed against in advance. These criteria, as mentioned above, could be decided freely by the student/s preparing the evaluation grid and, therefore, slightly differed between groups and individuals. For instance, some students decided to focus on pronunciation, public speaking skills and lexical repertoire, while others chose to base their grids on signposting, non-verbal communication, intonation and pauses. As stated in the other two chapters, sharing evaluation criteria with students is an essential step toward ensuring objectivity and transparency, as well as encouraging students' faith in this assessment process.

The roles during the presentations were those of presenter (or presenters), rater, and active audience member, meaning that some students were asked to come up with questions for the presenter/s. Furthermore, these roles were all decided upon beforehand and scheduled in advance, so that students would come prepared to the lesson.

As regards the weight of peer assessment for the final score, two groups were responsible for assigning half of the score for the presentation. More specifically, the final score consisted in 50% on the presentation and 50% on the essay. Given that not all of the groups gave a global rating and not all of them used the same rating system in the

analytic grid (some gave 5 points for each criterion, some gave 4, some used adjectives rather than points), the scores had to be interpreted for the sake of consistency by the teacher, who assigned a final score for each student based on the average resulting from the scores provided by the groups. The teacher checked for major discrepancies not only between the two groups of raters, but also between the students' evaluations and her own observations. However, a certain uniformity was noted. Where numerical scores were not available on a group's assessment grid, the teacher converted evaluations such as 'excellent' and 'good' into numerical scores. Bonus points were also available for the group reflective writing (self-reflection within a group) and one for actively intervening during the peers' presentations.

3.1.2 Participants

The questionnaire participants were a cohort of second-year master students of languages in an English course at University of Padova. Students who attended this course shared the study of the English language, but also studied one or more languages, as well as the related literature. Other frequently studied languages by these students were, for example, French, Spanish, and German. Three Erasmus students (two female students from Romania and one male student from Poland) also attended this course. Nevertheless, given the anonymous nature of the questionnaire, it was not possible to establish whether these three students also responded to the questionnaire.

Although there were twenty-four course participants[39], only twelve students (exactly half of the students) responded to the questionnaire. The age of the respondents ranges between 18 and over 30 years old. More specifically, half of the students are between 18 and 23 years old (50%), 41.7% of students are between 24 and 29 years old, whereas only one student is over 30. The findings also reveal that a vast majority of students who responded to the questionnaire were females (83.3%), while only two male students responded.

Furthermore, a large majority of the respondents stated in the questionnaire that they are aspiring schoolteachers (83.3%), whereas one student answered 'not sure' and one answered 'no' to this question. Given this information, the module part which focused

[39] This is the number of participants who were scheduled for the presentations in *Moodle*, a platform used by the University of Padova where course materials and forums for each course can be found.

on designing specific evaluation grids and on peer assessment seems particularly fitting to prepare students for their future career as schoolteachers.

The data from the questionnaire also revealed that all students except two were familiar with peer assessment and had had other peer assessment experiences at University, and some of them even in school. Some, however, specified that they had previously reviewed their peers' writings (peer review/peer feedback) but had never assigned points to their performances (peer grading/peer assessment).

3.1.3 The Questionnaire

The questionnaire (see *Appendix A*) aimed to find out this particular group of second-year master students' perceptions of peer assessment. At the beginning of the questionnaire, I briefly introduced both myself and my project, so that students knew the general purpose of the questionnaire and how their answers were going to be used in my dissertation. The questionnaire was entirely anonymous and consisted of 20 questions: the first three aimed at investigating the respondents' age, sex, and future job aspiration. More specifically, question three sought to find out whether or not the respondents are aspiring schoolteachers. This latter element was especially important, because I wanted to investigate whether or not aspiring schoolteachers found peer assessment so helpful that they would propose some kind of peer assessment to their future students. The remaining questions were all related to the peer assessment activity done during the course, such as whether the respondents found peer assessment objective or whether they thought their judgement was biased, as well as investigating their opinions on designing evaluation grids 'from scratch'.

The majority of the questions were open-ended questions (12), some were multiple choice questions (6), and two were scales from one to ten. Usually, in open-ended questions respondents were asked to motivate their answer to one of the multiple-choice questions asked previously or to one of the two scales. Moreover, all questions except one were compulsory and labelled with an asterisk (*). The only question that was not compulsory was question four, which was addressed only to those students who had answered 'yes' or 'not sure' to the previous question (about being aspiring schoolteachers). In fact, question four sought to find out whether or not those students

who are aspiring school teachers would ever propose a peer assessment activity to their future students.

3.1.4 Results

As stated above, one of the purposes of this questionnaire was to find out whether the peer assessment activity was helpful for students, in particular for those who are aspiring schoolteachers and, consequently, if these students would ever propose a similar activity to their future students. Interestingly, 83.3% of the students said that they would like to pursue such a career, and all of them agreed that peer assessment was (and is) useful and they would, therefore, probably propose such activity in the future. *Table 1* below lists some of the answers to this question. All answers from this group of students will be recorded in this chapter as they were written by them in the questionnaire, including typos and errors of grammar and/or word order. Moreover, each student has been assigned an anonymous label.

Table 1: Perceived usefulness of PA and possible replicability of this activity in the future.

In my opinion, peer assessment is extremely useful because it permits students to improve their skills thanks to their peers' observations on their performance. I would absolutely propose a similar activity to my future students since this could help them to understand what their level is and how to study to obtain better results. (S7)
Absolutely yes. Peer assessment is a good opportunity for students to build educational partnership with both peers and teachers, to develop critical thinking and to learn to examine knowledge, attitudes and learning skills. (S1)
Yes, I think that peer reviewing is a useful tool for both teaching and learning. As learners, getting to know your peers' mistakes enables you to ponder upon some of the most common mistakes that learners make. As teacher, it enhances practice for you future job (of course), but it's also a kind of activity that I would propose to my students, because it allows to compare themselves with someone with a similar level as theirs and learn from them. (S9)
Yes, I would argue it definitely improves the students' perception of what to consider when correcting somebody's work. (S8)
Yes, I found this activity very useful, because I think that it raises awareness on the difficulties involved in assessing the students' abilities. (S10)
I found the assess-your-peer activity quite useful because it turned out to be a challenging one. It made me aware of the difficulty of assessing someone as objectively

as possible, on the one hand, and of the importance of a constructive feedback on the other. (S2) Yes, I believe it has been useful, because I think students may successfully help each other and may correct themselves without the teacher's help. I think I will apply it in the future since students may become more autonomous in that way. (S4) Yes. It was very useful in understanding the pros and cons of peer assessment in teaching a language. (S12)

As the data provided in Table 1 reveals, students agreed on the usefulness of peer assessment and listed different benefits that may come from adopting peer assessment even with younger students. Some of the points that were made by the respondents were that assessing one's peer may increase learner autonomy, in that students will become more aware of their level by comparing themselves to students with similar levels, as well as help to obtain more objective feedback and make students aware of what assessment entails and the difficulties related to assessing students' skills. Moreover, some students are also aware that PA helped them to understand assessment and its inherent difficulties.

Only a couple of students seemed hesitant about proposing a similar activity to their future students, the reasons being listed in *Table 2*.

Table 2: Reasons for distrust of PA.

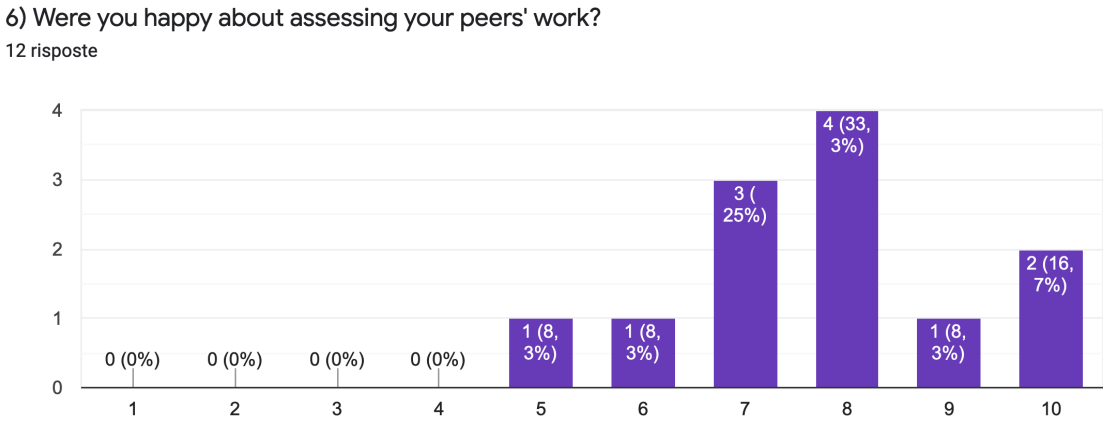
Your peers are students, just like you, so it's not always motivating because sometimes all you need is your teacher's feedback. (S5) It depends on their [future students] level of proficiency in English. I don't think that students' opinion is quite as authoritative as the teacher's one. (S6) I think that peer review [rather than peer grading] is more suited for younger students because providing a grade is a difficult task whereas providing feedback is easier. (S3)

The data in Table 2 shows that the reasons that were mentioned by students for their partial distrust of peer assessment revolved around motivation — possible lack of motivation — and the perceived inadequacy of students to assess their peers, especially if their "level of proficiency in English" (S6) is low. One student added that peer review is more appropriate than peer assessment for younger students [her future students], because asking students to provide feedback is easier than asking them to also provide grades (S3). In spite of this, these students did not state they would never propose such an activity to their future students, but they were aware of the potential difficulties that

peer assessment might entail and that have been thoroughly discussed in the previous chapter.

Next, *Figure 1* below illustrates students' responses to question six, which sought to investigate students' degree of happiness with the PA activity. The data reveals that a vast majority of the students (83.3%) were happy or very happy about assessing their peers' presentation, thus giving a score from 7 to 10. Nevertheless, two students seemed less enthusiastic about it (scores 5 and 6).

Figure 1: Student's degree of happiness with assessing peers' presentations.



Students were then asked to motivate their answers to this question, and those who claimed that they were not so happy about assessing their peers' work stated that the main reason for that was that they felt under pressure about judging their friends. Moreover, one student also pointed out that he had sometimes questioned the usefulness of the activity:

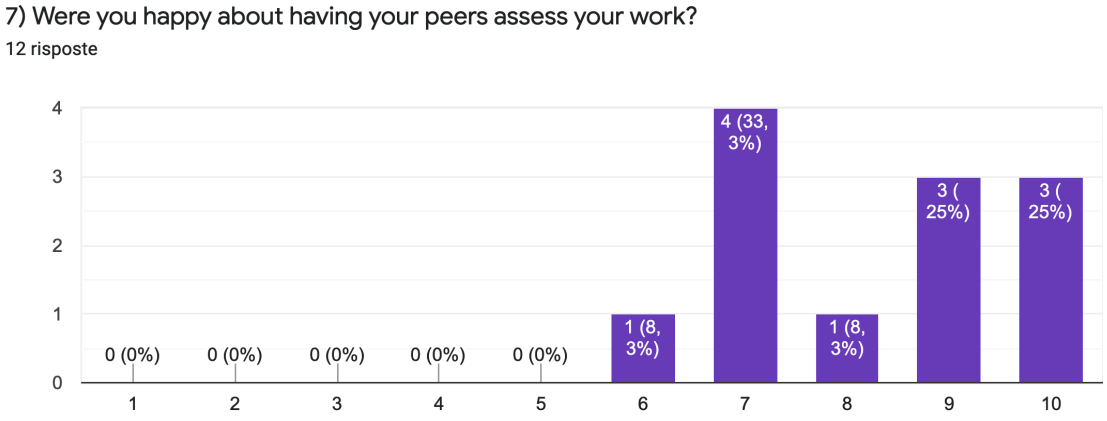
> "I sometimes perceived it as not necessary at all, given the little time I had to perform plenty of tasks. Additionally, some of my peers' work was abnormally bad, and I felt somehow offended in 'wasting my time' in such activity." (S8)

On the contrary, the two students who were very happy (10/10) about assessing their peers' presentation explained that this was so because "receiving feedback from someone whose level of proficiency is like yours is constructive" (S6) and "it was a great chance to check whether the criteria we set in order to assess our peers were correct or needed improvement" (S10). It is interesting to note that one student (S6) interpreted this question with regard to how useful they thought their feedback was going to be for the

presenter, whereas the other student (S10) answered this question thinking about the validity of the rating scale that her group designed. In particular, the first student noted the importance of carrying out peer assessment between students with similar status, which is often discussed in the literature on peer assessment (Topping 2005; Jung 2016).

On the contrary, *Figure 2* below illustrates students' degree of happiness with having their peers assess their presentations. In this case, the data shows that a big majority of the students were happy about it (91.6%), but just over half of the students were very happy about it (58.3%), giving scores from 8 to 10, whereas one student seemed a bit reluctant and scored six on the scale.

Figure 2: Degree of happiness with being assessed by peers.



Consequently, the following question sought to find out why it was so. This particular student explained that the reason why she was not so happy about the activity was that she thought that only teachers should be responsible for assigning grades: "it sure is interesting and very useful to have a feedback from a peer but I'm not sure I love the idea that their assessment may play such an important part in the final grade. I think it should be the teacher's responsibility" (S2). Moreover, that same student also added that "inexperienced judges tend to be too harsh sometimes", thus suggesting the possible inadequacy and unreliability of peers' judgements, especially if they have not been trained properly and thoroughly for this kind of assessment ("inexperienced judges").

Instead, the three students who were enthusiastic (10/10) about having their peers assess their work observed that sometimes one's perception of his/her own performance is influenced by self-esteem or confidence, and, therefore, it might be hard to understand

what skills really need improvement. In this case, then, peer feedback seems extremely helpful. Moreover, one of these three students added that knowing that a peer was going to assess her work, as well as the teacher, encouraged her to do her best. This same student also claimed that assessing a peer helped her to reflect on her own performance: "it was a pleasure to assess my peers' work since this gave me the chance of reflecting on some important aspects of oral skills and public speaking in my own speech as well" (S7).

Lastly, one student who was quite pleased with the activity (and gave a 7 to the activity) said that she liked it but she would have preferred to receive more detailed feedback[40] from her peers, instead of just getting a grade: "maybe, if possible I would have wanted to have a more detailed feedback from my peers to know how to do better in the future" (S2).

All students who claimed that they were very happy about assessing their peers' work, also claimed that they were happy about having their peers assess their work. Interestingly, though, a couple of students who were not so happy about assessing their peers' work, had a different opinion when they were asked to indicate their degree of happiness with having their work assessed by their peers (student 5 and student 8).

Table 3 below illustrates a comparison of results between question six and seven, which revolved respectively around the degree of happiness with assessing peers' presentations and having one's presentation assessed by peers.

Table 3: Comparing results from question six and seven of the questionnaire.

Student	1	2	3	4	**5**	6	7	**8**	9	10	11	12
Q. 6	8	7	8	7	**5**	8	10	**6**	9	10	8	7
Q. 7	9	6	7	7	**7**	10	10	**9**	9	10	7	8

The reasons that these two students gave about their answer to question seven were that the feedback received was "constructive, reasonable, appropriate, professional and helpful"[41], thus actually helping them to improve their skills.

With regard to question nine, I asked students whether they felt adequate and prepared in their teacher-like role (i.e. when they were asked to assess their peers using a

[40] All groups of students wrote notes on their grids that could be shared with peers upon request. However, this lack of detailed feedback was mainly due to the current situation with Coronavirus, since students could have received the detailed feedback during Professor Ackerley's office hour at the end of February, which were, however, moved online and it was not possible to give students access to the notes.
[41] S3 and S8.

rating scale, which they had to design from scratch). Answers to this question, some of which I have reported in *Table 4* below, show that just half of the students did feel up to the task, but the other half of the students did not feel well enough prepared and thought that preparing an evaluation grid was no easy task. The first three statements belong to the students who did not have any problems with assessing their peers, whereas the other three belong to those who had some difficulties during the process.

Table 4: Students perceptions of own adequacy in assessing peers' work.

Personally, I was comfortable preparing an evaluation grid and assessing my peers' work. (S7)
I had some experience and theorical information about it so I knew I could do it. (S8)
It wasn't easy. It was the first time for me, so I didn't know exactly what I was supposed to do, but with some help from the teacher it turned out I could do a good job. (S1)
No, I felt puzzled and lost sometimes. The help of Professor Ackerley and the fact of working in groups were essential. (S3)
I think I need to reflect more on some aspects. (S10)
No, I did not feel fit for this activity at all. (S5)

One student in particular argued that having "theoretical information" (S8) about peer assessment and grading scales helped him during the process of designing the rating scale, thus suggesting that students need to be instructed on how to prepare evaluation grids. This aspect has also been pointed out in Chapter One and Chapter Two.

In spite of the difficulties encountered during the PA activity, most students claimed that they would not have preferred to have already had an evaluation grid and simply filled it out, thus acknowledging the importance of designing a grid from scratch. For instance, one student said "no, this was a considerable opportunity for me and I enjoyed preparing the evaluation grid and assessing my peers' work" (S7), while another student claimed that designing her own grid made her "more aware of what it takes to assess well and fairly" (S12). Indeed, literature on peer assessment shows that having clear and precise grids helps teachers to deliver fair and objective assessments.

Additionally, two students admitted that at the beginning they would have preferred the second option, but looking back at the activity now, they are happy they could design their own grid, even for their future career as schoolteachers: "it would have

been easier at first, but I think that sooner or later we'll have to create or adapt evaluation grids and now I can say to have a model to rely on" (S3), and "at the beginning of the activity I would have preferred it, but now I think that the preparation our own grids has been an enriching element for the experience" (S1).

Having said this, three students said they would have preferred to have already had a grid, claiming that "it's been really hard and challenging to create one, even though we were a group" (S5) and that "it would have been easier to do so" (S4). Furthermore, one student in particular said that she was lacking preparation on how to design a valid rating scale, because she did not know "how to estimate correctly the weight of each imperfection in the performance"[42] (S10). Interestingly, this student is the only one who is not interested in becoming a schoolteacher.

In addition, students were also asked to identify the main benefits of designing a tailored evaluation grid. The most frequent answer was the freedom of choice about which criteria to use and on which aspects of the language to focus. For example, one student observed that

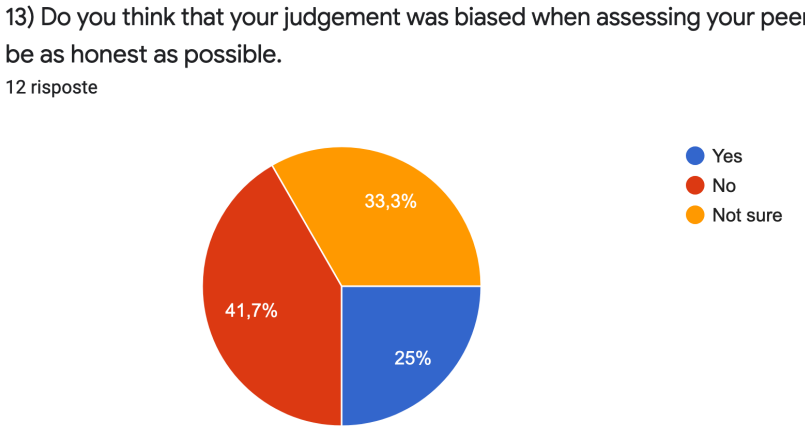
> "in order to create your own grid, you have to think about what you want to evaluate and according to what parameters. So, it is a valid instrument to make you focus from the very beginning on what you will have to consider and evaluate" (S1).

The next question sought to investigate the perceived objectivity of peer assessment. In this case, only three respondents thought of peer assessment as objective, whereas six of them were not sure. Instead, three students claimed peer assessment cannot be objective, and their main reasons for this answer were the fear for repercussions when assessing a friend and the fact that different "assessors might have different opinions on the performance" (S7). Having said this, one student pointed out that having precise grids and rubrics for assessment helped her toward achieving more objective judgements. As stated above, the literature on peer assessment also suggests using very clear and precise assessment criteria to ensure objectivity in the process (see, for instance, Kroll 1991; Topping 1998; Brown 2004; Hattie 2012).

[42] As stated in *§3.1.1*, the grids were based on the *CEFR*, which was developed by the Council of Europe in 2001 to describe in broad terms language learners' abilities. The descriptors for each of the four skills (writing, speaking, reading, and listening) comprise several 'can do' statements, thus pointing at what a student can do with a language, rather than what he/she cannot do.

Some studies on peer assessment (Cheng and Warren 2005; White 2009; Peng 2010; Jung 2016) have emphasized that students often dislike peer assessment, in that they find it subject to biases, especially when it comes to assessing a friend. Therefore, my next two questions revolved around possible biases in peer assessment. First, I asked students whether they thought that their judgement was biased when assessing their peers' presentations (see *Figure 3*).

Figure 3: Students' perceptions of own biases.



The results of this question reveal that almost half of the students (41.7%) thought that their judgement was not biased when assessing their peers' performances, and that 33.3% of students were not sure, while three students admitted that their judgements were biased.

When asked to express their opinion on why evaluations in peer assessment might be biased, a large majority of the respondents (9 students) claimed that friendship was the main reason for biased judgement/assessment, thus confirming what the literature on peer assessment also says about this issue (see, for instance, White 2009; Peng 2010; Azarnoosh 2013).

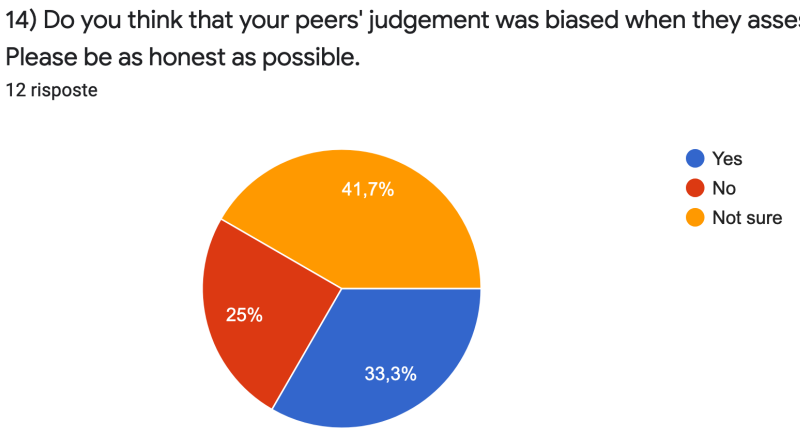
Table 5 below lists some of the answers to this question. Interestingly, one student also pointed out that he might have been biased because he knew the academic performance of his peers and, consequently, might have been influenced by that in his evaluation, and another student claimed that "dislike for the presentation's topic" (S11) might also be one of the many causes for biases. Moreover, another student added that "social rules and politeness" (S12) might influence the way in which students express their evaluations about peers' performances.

Table 5: Possible reasons for biases.

As I said before, I think liking or disliking one of your peers definitely affects your judgement. Not only because of sympathy of antipathy, but also given the fact we generally knew our peers' 'scholastic performance'. (S4) Friendship, social rules and politeness come in to play, hence the biased assessments. (S12)
I think peers feel compelled to be kind to their friends because they might fear repercussions on their friends' marks, for example. (S6)
I honestly think many students were influenced by the perception they previously had of their peers, e.g. being friends. (S8)
In my experience, friendship, dislike and competitiveness are the main causes for biases. (S7)
I think that dislike for the topic can be a cause for bias - it could be something I'm not interested in or understand and my evaluation can be affected by that. (S10)

Then, I also asked students whether they thought that their peers' judgement was biased when assessing other students' presentations. Figure 4 below illustrates the answers given by the students.

Figure 4: Students' perceptions of peers' biases.



In this case, the findings reveal some differences with the previous question, the most striking difference being that only 25% of students though that their peers' assessments were not biased. Instead, 33.3% of students claimed that their peers' evaluations were definitely biased, while a significant number of students (41.7%) were not able to evaluate whether or not their peers' assessments were influenced. Two students in particular claimed that the peers assessing them were "really kind" (S5) in

their scores and that "the feedback was more positive than I expected" (S10), thus suggesting that, possibly, some students tended to be over-generous in their assessments. Moreover, this latter student also pointed out that, in her opinion, "students' judgements are biased by their desire to be kind with others, maybe expecting the same kindness in return".

Lastly, students were asked to provide any suggestions on how to avoid biases. *Table 6* lists some of the suggestions that some of the respondents gave on this important issue. The most frequent answers were to form groups of students who do not know each other well (one student called these groups "mixed groups"[43]), and to have more than just one assessor for each performance/presenter (S1 and S9). One student added that she could not think of any way to avoid biases in oral tasks but pointed out that written tasks could be submitted anonymously (S11), thus significantly lowering the possibility of biased judgements. Furthermore, one respondent also suggested that training students in assessing their peers might also help to avoid biases (S5). This latter aspect in particular has also been emphasized in the literature on peer assessment (see, for instance, Freeman 1995; Topping 1998; Patri 2002; Cheng and Warren 2005).

Table 6: Student' suggestions on how to avoid biases during PA.

I would say peer assessment would be better performed by groups of students who do not know each other. (S8)
I found it useful, but there's a lot of work to be done in order to become good at that. (S5)
It is quite impossible to avoid bias completely. Again, if there are more reviewers there is a lesser chance of bias. (S1)
Working in mixed groups could be a solution to avoid biases. (S3)
I think it's impossible to be 100% unbiased; for this reason, I think that working in group could be a solution. (S10)
In my opinion, the higher the number of reviewers, the higher the level of objectivity and unbiased behaviour. (S9)
The important thing is to recognize and separate personal subjective judgment from critical thinking and be sure that when subjectivity comes into action, you are ready to stop it. (S4)

[43] S3.

You've got grids and rubrics that are quite precise, so I think that if one sticks to them, they'll give you an objective evaluation and prevent biased judgements. (S2) I honestly don't know how to avoid biases on oral evaluations, but they can be avoided in written evaluations through turning in works anonymously. (S11)

3.1.5 Discussion and Limitations

This study attempted to observe the perceptions of a cohort of second-year master students of languages on the peer assessment activity which was carried out during a Spoken English module at the University of Padova.

As appears from the questionnaire findings, all the interviewed students expressed positive thoughts on peer assessment and found it helpful overall, especially to improve their speaking skills, but also to have a better understanding of what classroom assessment entails (i.e. issues and responsibilities). Some students in particular said that peer assessment "raises awareness", which was actually the reason why the teacher decided to carry out this activity in the first place: "I really liked the activity and I felt the responsibility upon my shoulders. I think it's usefull to try it in class as a student before doing it as a teacher. It raises awareness" (S3), and **"**I found this activity very useful, because I think that it raises awareness on the difficulties involved in assessing the students' abilities" (S10). Other students expressed the same view, though not using these specific terms: "it made me more aware of what it takes to assess well and fairly" (S12), "I would argue it definitely improves the students' perception of what to consider when correcting somebody's work" (S8), and "it was a pleasure to assess my peers' work since this gave me the chance of reflecting on some important aspects of oral skills and public speaking in my own speech as well" (S7), thus confirming that peer assessment can have positive repercussions on self-reflection (and on self-assessment)[44].

Nevertheless, a couple of students seemed hesitant to carry out assessments on their peers' performances, for a number of reasons. The reasons named by the students for not completely trusting or enjoying peer assessment were, for instance, the fact that they were expected to assess friends as well as having to assess notoriously 'good' students, that is students who usually obtain good marks. This last aspect in particular was also named as a possible cause for biased evaluation by a couple of students. Indeed,

[44] See Chapter One, p. 22.

assessing students who usually get high grades might be difficult for their peers, because one might perceive himself/herself as inadequate in this teacher-like role. Interestingly, in the studies of peer assessment that have been reviewed in Chapter Two, no evidence of this particular issue (i.e. assessing students who notoriously obtain higher grades) has been found. However, as stated in Chapter Two, students often do feel inadequate in their teacher-like roles, thus usually not having high expectations about peer assessment.

Overall, this particular group of students seemed to be aware of the difficulties that assessment entails, and especially peer assessment, and were able to provide solid suggestions on how to achieve objectivity and avoid biases. Some students, for example, noticed that the opinions or prejudices that they previously had on their peers influenced their assessments at first; nevertheless, these students also stated that they tried to do their best to prevent that subjectivity hindered providing objective evaluations, once they acknowledged that they were being biased. Interestingly, one student claimed that social rules were also one of the possible causes for biases. However, the term used by this student is vague and very broad, thus making her comment difficult to analyse.

Unfortunately, this study has several limitations. First and foremost, the limited number of respondents compared to the total number of course attenders (exactly half of the course attenders filled in the questionnaire) cannot be considered representative of the whole class and, consequently, made drawing conclusions and providing generalisations difficult. For example, the questions related to age, sex and whether students were aspiring schoolteachers did not show significant differences in the answers. Only one student out of twelve respondents was over thirty, and only two male students responded to the questionnaire. Moreover, almost all students are aspiring schoolteachers, except one who answered ‘not sure’ and one who said she is not. Last but not least, this study only collected students’ responses through a questionnaire and several months after the course with Professor Ackerley had ended. Perhaps, if I had asked students to fill in this questionnaire right at the end of the course or if I had designed the peer assessment task myself, the results might have been different.

In spite of this, the questionnaire findings show an overall positive attitude of students toward peer assessment and the results can be thought of as showing a tendency of this particular group of students toward peer assessment.

The next section will be on my own study of peer assessment, with which I have sought to find out first-year bachelor students' perceptions of peer assessment and to compare their answers with those of the more experienced second-year master students surveyed in these sections.

3.2 A study of peer assessment with first-year bachelor students

3.2.1 Participants

The participants of this study were six students of foreign languages enrolled in an English course at the University of Padova[45] who are preparing to take the written exam of *General English* in June. A big majority of the students who responded were female students (5), whereas only one male student participated. Three students were aged between 18 and 20, two students were aged between 21 and 29, whereas only one student was over 40. Furthermore, two students claimed that their level of written English was a B1, two claimed that it was a B1+, one claimed that it was a B2 and one claimed that it was a C1, which allowed me to form fairly balanced pairs for peer assessment. Indeed, as stated in Chapter Two, peer assessment should occur between students who have "equivalent language ability and progression" (Jung 2016: 10) to ensure a successful mutual learning experience.

3.2.2 The two Questionnaires

Two anonymous questionnaires were distributed to the students to collect their opinions about peer assessment before and after the activity took place. Most questions were obligatory and were marked with an asterisk (*).

The pre-activity questionnaire (*Appendix B*) consisted of 11 questions: the first three questions were all related to the students' personal information, namely their age, sex, and level of English. Instead, the remaining questions regarded their past experience with peer assessment (if any) and their opinions about the objectivity of peer assessment.

[45] I am aware that statistically my results are not relevant and that it is not possible to make generalizations with only six students participating in this study compared to the total number of students enrolled in the course (about 400). However, due to the special circumstances brought about by the spread of Coronavirus all classroom lessons have been suspended. Consequently first-year bachelor students did not have the possibility to meet me in person and have only been contacted via *Moodle*, which has hampered data collection.

The questions asked were similar to those asked the group of second-year master students to allow an easier comparison of the results.

The post-activity questionnaire (*Appendix C*) consisted of 12 questions: the first two were related to their age and sex in order to identify the different answers of the students to the first and second questionnaire to see whether there were any significant changes (e.g. the perceptions of peer assessment of the male student in the first questionnaire and the perceptions of peer assessment of this same student after the activity took place). Instead, the remaining questions regarded their experience with peer assessment, namely their feelings about assessing their peer's performance, and whether they were satisfied with their peer's evaluation. Last but not least, the post-activity questionnaire also sought to find out whether the students would have preferred to just provide feedback and not a grade to their peer, or whether they were happy with providing both.

3.2.3 Writing Task and Evaluation Grid

The writing task was designed specifically as practice for the final exam of *General English*, which consists in writing a listening report. Students were given very precise instructions and guidelines on how to write their report, which were largely based on the instructions that they had been given by their teachers during the course (e.g. time to be spent in the task, number of words to be written, and report structure). *Table 7* below illustrates the instructions and guidelines for the report as they were given to the students.

Table 7: Writing task: instructions and guidelines for listening report.

Task instructions
For the purpose of this study and as practice for your final exam of *General English*, you will be asked to write a listening report following some guidelines which I will provide (see below). Then, you will also be asked to assess one of your peers' listening reports: you will be asked to give a score from poor to excellent to your peer's work by filling in an evaluation grid. You will also be asked to provide constructive and detailed feedback to your peer to justify the score. This type of assessment where students assess other students is called peer assessment. Peer assessment can be very useful, because it allows you to receive feedback from more than just your teacher. Moreover, by assessing the work of others, you may become more aware of your own [writing] skills. I would like to clarify that, at the end of the activity, you will receive a score both from your peer and from me. However, these scores will NOT be taken into account by your professors: they will simply serve

you as indicators of how well you have accomplished the task and as preparation for your final exam of *General English*.

Guidelines for listening report

Listen to this interview with Angelina Jolie on her fairly recent film '*Unbroken'*: *https://www.npr.org/2014/12/17/371445206/angelina-jolie-on-her-films-unbroken-hero-he-was-truly-a-great-man?t=1588157070691*

Listen to the video once carefully and without taking any notes. Then, listen to it again and take notes. The notes will be useful for summarizing the content of the video. Do NOT look at unknown words in the dictionary. Then, write a full text of about 300[46] words touching on all of these aspects:

- write a short summary of the content of the video (about 120-150 words),
- take notes of difficult words and/or difficult passages and write a short paragraph on which strategies you used to bypass the issue (e.g. listening to the word/passage again, looking at the word/passage in its context to grasp its meaning…),
- the amount of time you spent listening to the video and how many times you listened to it (once was enough, twice…),
- was it difficult to listen to? Why? Why not?
- write the total amount of words at the end.

You have 50 minutes[47] to listen to the interview and write your report. Please make sure that you write a cohesive text using linking words.

Before you start writing your report, make sure that you take a close look at the evaluation grid so as to be aware of the evaluation criteria you will be assessed against by your peers and by me.

As shown in Table 7, the writing task consisted in a listening report of about 300 words. The audio recording for the listening report was an interview with film director Angelina Jolie on her film *Unbroken*. I chose this particular interview because, generally, the listening in the exam is an interview with a book author, so I wanted the whole activity to be relevant for the students and as similar as possible to the exam environment.

After the students had completed their report, they were asked to send it to me via e-mail. I then swapped the reports and sent them to the other student for the actual peer assessment part of the project.

To assess the reports, students were asked to fill in an evaluation grid (see *Table 8*) which was designed specifically for this task. The criteria in the evaluation grid were decided upon according to what I assumed were the relevant aspects of a listening report. The criteria were distributed in the grid from what I thought were the easiest to assess (punctuation and spelling) to the hardest to assess (overall task accomplishment).

[46] In the exam students are expected to write between 250 and 350 words.
[47] In the exam students have one hour to complete their listening report.

Moreover, I decided not to put numerical grades but only evaluations (poor, mediocre, good, outstanding) in the grid so as not to intimidate those students who had never carried out peer assessment. The grid was given to the students at the same time as the report instructions, so that they could become familiar with it and know exactly on which aspects they were going to be assessed. Indeed, it is usually maintained that if students know how they are going to be assessed, it is going to be much easier for them to accept the results of the assessment.

In addition, students were also asked to provide feedback to their peer using some useful phrases taken from Keller and Westfall-Greiter (2014)[48]. For the peer assessment part, students were also given very clear and precise instructions to avoid misunderstandings. Moreover, these students had not been trained to give feedback nor grades during the course, so I assumed they needed special guidance to carry out successful evaluations.

I also filled in the evaluation grid myself and provided feedback to all participants, so that they could compare their peer's suggestions with my feedback. Moreover, I also wanted to investigate teacher/student rating agreement with these students, which will be discussed later.

Table 8: Analytic evaluation grid for peer assessment.

Criteria	Poor	Mediocre	Good	Outstanding
1. Spelling and punctuation	Very frequent and repeated lapses in spelling and/or punctuation.	Frequent lapses in spelling and punctuation.	Few lapses in spelling and/or punctuation.	No lapses in spelling or punctuation.
2. Grammatical accuracy	Severe errors of grammar or word order, which seriously interfere with comprehension	Several errors of grammar or word order, which partially interfere with comprehension	Some errors of grammar or word order, which do not interfere with comprehension	Very few/No errors of grammar or word order. Efforts of interpretation not required.
3. Vocabulary	Improper language use. Several misused	Occasional improper language use and some	Rare improper language use and/or very few misused words. Re-	Proper and accurate language use. No re-reading necessary.

[48] See Chapter Two, p. 53.

	words, and necessary re-reading of text.	misused words. Re-reading of text necessary in some passages.	reading of text not needed, and efforts of interpretation not required.	Own words used to summarize video.
4. Text organization and cohesion	No evidence of linking words, which interferes with comprehension and affects text cohesiveness. Very basic sentences.	Minimal evidence of linking words that nonetheless help the overall comprehension and cohesiveness of the text. Basic sentences.	Appropriate use of linking words, although same ones tend to be repeated. Good connection between thoughts. Elaborate sentences but not always correct.	Excellent and various use of linking words. Thoughts logically connected and text flows well. Elaborate sentences, but very clear and concise at the same time.
5. Overall task accomplishment	Purpose of task not understood. Guidelines for listening report not followed. Word limit not respected.	Partial understanding of task purpose. Forgot to discuss one or more aspects in the guidelines. Word limit not respected but within an admissible span.	Purpose of task understood. Task accomplished well. All aspects in the guidelines discussed. Word limit respected.	Task excellently accomplished. Word limit respected and guidelines followed. Very high level of written English.
Final score:				

3.2.4 Questionnaires Results

Before discussing the results of my study, I would like to specify that the activity was completely anonymous and that the students did not know who the student assessing them was and had no contact with each other whatsoever because I acted as an intermediary (e.g. I received the reports and swapped them and contacted them via e-mail individually).

The pre-activity questionnaire results show that almost all students except one knew what peer assessment is and had already been asked to assess, with or without

grades, peers' performances. Having said this, the students had different reactions about peer assessment: the male student (S2) claimed that, on that occasion, he thought of himself as being helpful to the other student. On the contrary, three other students student claimed that, on that occasion, they had felt inadequate and uncomfortable: "I don't think I have the skills to correct someones[49] work, and also I am completely aware about my difficulties of writing in English" (S1), "I felt unconfortable because i don't think that i have the righ skills to assessing anyone" (S6). The one student who had never done peer assessment, however, claimed that she was "interested" (S5).

The following questions were related to the perceived objectivity of peer assessment. More specifically, I asked students whether they thought that peer assessment was equally objective as teacher assessment. Overall, two students thought peer assessment was almost as objective as teacher-assessment and scored an 8 in the scale, two students scored a 7, one student scored a 6 and one scored a 5. The following question asked students to motivate their answers, which are listed in *Table 9*.

Table 9: Students' motivation for their answer to question 8.

The students generally tend to be more comprensives than a teacher. (S4)
In my opinion try to assessing a peer will never be equally objective as a techer could do, because you may feel unprepared, not sure about your abilities or even making a displeasure to him/her. Revisions need attention and confidence to create no doubts and be reliable. (S5)
I think it should be rather objective, but pheraphs not as objective as theacher's since students may try to 'sugarcoat' their revisions or ignore some mistakes. (S3)
I think it depends on the skills of the classmate who is correcting. Another think is that we feel ourselves, and what about me, I would never be too evaluate in corrections because I don't what to make someone feel bad, especially if we are friends. (S1)
Peer assessment is not objective at all because it is always compare to own opinion that it might be incorrect. (S6)
Yes, because the assesment creteria used by a teacher or by a student are standard. (S2)

One student claimed that the objectivity of peer assessment depends on the skills of the assessor and that she personally would not want to make someone "feel bad about

[49] Students' answers will be recorded as they were written in the questionnaires.

his work, especially if we are friends" (S1), thus suggesting that friendship might be a cause for biased evaluations. Furthermore, interestingly, another student claimed that students tend to be more "comprensives" (S4), possibly meaning understanding and, therefore, more generous than teachers in their assessments. On the contrary, one student claimed that peer assessment is as objective as teacher assessment because "the assesment creteria used by a teacher or by a student are standard" (S2).

All students then agreed that anonymity in peer assessment could be a successful solution to avoid biased assessments.

As far as the post-activity questionnaire is concerned, results show that all students were happy with the peer assessment activity overall and were very comfortable both with assessing another peer's report and having another peer assess their report[50]. Nevertheless, most students (5) claimed they would have preferred to just provide feedback (and not also a 'grade'[51]) to their peer's report, because "I can provide an advice but I haven't the skills for assigning a grade" (S1), "basically because I don't know how to translate the indicators into a number[52]" (S3) and "assign a grade, evaluate other people's works, means to judge it without justifications in some way. I was glad there was a feedback too" (S5). Two students then added that the guidelines for peer feedback and the evaluation grid helped them to achieve fair evaluations: "it was essentials for me to have a guideline, because is hard to give an opinion about so many different aspects about an assesment" (S5), "it gave extra confidence about making sure the assessing was going right" (S3). Another student claimed he was very comfortable both with assessing his peer's listening report and with having his work assessed by his peer, because he was given precise instructions and guidelines:

> "guidelines helped me a lot to giving a grade to my peer, so for me it was essential key and the author of this task was very clear in given us her indications and she also was very available to be clear in anything" (S2).

Last but not least, all students agreed that giving a grade was more difficult than providing feedback because it was the first time that they had done so.

---

[50] Possibly, the complete anonymity which was ensured to the students throughout the whole project has contributed towards this feeling of 'comfort'.
[51] Poor, mediocre, good, outstanding.
[52] They were not asked to do this anyway because I knew they might find it difficult, therefore they were only asked to provide an 'adjective' instead of a numerical grade.

3.2.5 Task Results

As regards the listening reports, overall all students accomplished the task well. The first pair was formed by a female student (S1) and a male student (S2), both claiming to have a B1 level in written English. The report written by S2 was a very detailed summary of the content of the audio file, but he forgot to touch on some aspects that were mentioned in the guidelines (notes of difficult words and/or difficult passages, strategies used to bypass this issue, and the amount of time you spent listening to the video). Instead, S1 touched on all the aspects mentioned in the guidelines but forgot to write the total number of words at the end of her report. Moreover, from how the students accomplished the writing task, it seems that the female student had a B2 level rather than the B1 that she claimed she had, whereas the male student seems to have a significant lower level of English than the other student (A2+/B1).

With regard to the evaluation grids, S1 gave the final score 'mediocre' to her peer. More specifically, she gave him a 'good' in spelling and punctuation, and a 'mediocre' in the remaining aspects. Interestingly, she also gave a numerical grade (20/30) to her peer's listening report, although this was not required.

As regards the feedback provided by this student, *Table 9* illustrates the suggestions that this student gave to her peer.

Table 9: Feedback of S1 to S2.

You said you didn't understand very well the audio so you should practice more on Listening Reports. I noticed you sometimes don't organize exactly the sentences and this makes them not fluent. Maybe try to use more linking words. You should also divide properly your text into several paragraphs to make it clearer. There were also some misused terms: I think that you should write 'words' instead of 'vocabularies' in the last paragraph and sometimes you could write synonyms instead of repeating the same words. And also in the last paragraph you wrote 'she talk' instead of 'she talks'. Moreover these sentences are not very clear. Can you rewrite them to explain them better? "When the film is realized Zamperini said how great and exceptional is he, this remind the others the self-greatness… In the protagonist's period experience it was unusual being open mind and telling to the others about private life, because also of his generation which is different from now's generation that can discuss freely, but Zamperini was not stressed to telling his story."

S2 gave the final score 'good' to his peer's report. More specifically, he gave an 'outstanding' to the criteria 'grammatical accuracy', and 'good' to the remaining criteria.

This student also gave a numerical grade to his peer, namely 26/30. He was not influenced by the comments of S1, because they received each other's grids and the feedback as well as my grid and feedback at the same moment. Furthermore, this student did not use the useful phrases for feedback that I had provided them with in the guidelines for feedback. *Table 10* displays the feedback that this student gave to his peer.

Table 10: Feedback of S2 to S1.

I think that you need just some attention on the punctuation and a little bit on the subject (when there are more subjects) of the sentences when periods are long. Some sentences could be shorter. In this sentence, if you want to start a new paragraph, you should not start with "in fact", but you could use Zamperini or The protagonist: "In fact he suggests to Angelina not to make a film about how great he was, but to make one which could remind people that they have greatness inside themselves." Here you should connect the sentences better using 'who': "This audio is an interview to Angelina Jolie, the director of the fair film *Unbroken*. She is asked about the protagonist Louis Zamperini, who was a soldier of World War II, about his personal experiences". She who? "She talks then about how responsible she felt to product this film and how emotional it was to make him watch a cut of it before he died."

As regards my corrections, I sent both students their listening report back with highlighted passages that indicated misused words, errors of grammar or word order and unclear sentences. I gave them possible solutions for these passages (e.g. correct words, corrected mistakes[53]…) and I also wrote a detailed comment explaining what I think they could do to improve their writing skills. *Table 11* lists my comments to both students on their writing task.

Table 11: My suggestions to the students on their listening reports.

S1	S2
Here are my suggestions on how you could improve your listening report as preparation for the written exam: - try using more linking words to help text cohesiveness (however, on the one hand… on the other hand…, furthermore, in addition, in spite of this, moreover, nevertheless, besides…). - One sentence was not so clear: "Moreover, to the question about the	Here are my suggestions on how you could improve your listening report as preparation for the written exam: - your summary of the audio recording is very detailed and precise; however, you forgot to mention several aspects of the guidelines, such as how many times you listened to the audio file and any difficult words that you have encountered. Next

[53] For example: "In the protagonist's period experience it was unusual being open mind and telling to the others about private life" (original sentence). "At the time when the protagonist lived, being open-minded and telling other people about one's private life was unusual" (my corrected version).

soldier's traumas she answers that there weren't places he couldn't go to or things he couldn't do, he was about going on and making use of all the experiences he lived." Make sure that you re-read your report before you hand it in at the exam to check on any last-minute mistakes and make sure that every sentence is clear and, if not, try rephrasing unclear sentences. - I have corrected some mistakes of grammar (especially prepositions) or word order so that you can improve for next time. I have also written other synonyms or options for certain words. P.S. You forgot to mention the total number of words at the end of your report. Overall, I think that you did a good job, especially at summarizing the audio recording, because you really focused on the most important passages in the interview. Next time try expanding a bit more on the rest of your report and use more linking words to connect sentences.	time, I would suggest reading the instructions more carefully. - Try using more linking words to help text cohesiveness (however, on the one hand, furthermore, in addition, in spite of this, moreover…). - Be careful with word spelling: if you don't know how to spell a word, try using another one with a similar meaning of which you know the spelling. - Some sentences were unclear: in the exam, try rephrasing longer sentences to make them clearer to the reader. In English, it is best to keep sentences shorter and simpler. - I have corrected your mistakes of grammar or word order so that you can improve for next time. Don't worry if you're not satisfied with your results: I saw your efforts. Maybe you could watch some videos on YouTube in English or watch some films with English subtitles to improve your understanding of spoken English as well as your spelling.

As far as the evaluation grid is concerned, I also gave the students both a grade from poor to outstanding and a numerical grade. The female student obtained the final score 'good' (25/30). More specifically, I gave her 'outstanding' in spelling and punctuation, and 'good' in all the other aspects. The male student, instead, obtained the final score 'mediocre' (18/30). More precisely, I gave him 'poor' in spelling and punctuation and in vocabulary, and 'mediocre' in grammatical accuracy, text organization and cohesion, and in overall task accomplishment.

The second pair was formed by S3 with a C1 level and S5 with a B2 level. Actually, from what I could tell, I think that this latter student also has a C1 level, so this pair was very balanced. Both reports were very well written and both students accomplished the task excellently. Moreover, these students did not give a numerical grade to each other's reports. With regard to the evaluation grids, S3 gave the final score 'outstanding' to her peer. More specifically, she gave her a 'good' in spelling and

punctuation and in overall task accomplishment, and 'outstanding' in the remaining aspects.

As regards the feedback provided by this student, *Table 12* illustrates the suggestions that this student gave to her peer. Actually, there is only one piece of advice, because S5 wrote a very good report and there were very few inaccuracies.

Table 12: Feedback of S3 to S5.

I am positively surprised by the accuracy of the report. The text is very well written and doesn't lack information. There are almost zero imprecisions or misunderstandings (apart from "living ON his own", or "answering to QUESTIONS" rather than details, or PTSD instead of just PTD). The only criticism I have is the word limit not respected: according to the Instructions I've received, summaries should not be longer than 150 words and the whole listening report should not be more than 300 words total. Therefore, I recommend to keep an eye on that, but overall the report was almost perfect.

As for S5, she gave the final score 'good' to her peer (she gave a 'good' to all the aspects in the evaluation grid). *Table 13* records the feedback that S5 gave to S3.

Table 13: Feedback of S5 to S3.

Overall, I think it's a good listening report, even if there are some minor misunderstandings. The sentences are written in a structure which is simple and easy to read and minimize mistakes. Mind the use of some words like the verb to realize or the noun vision, they're "false friends" for Italian first language speakers. I would suggest varying the use of linking words to keep up the rhythm and preserve the reader's attention.

As regards my corrections, again I sent both students their listening report back with highlighted passages that indicated misused words, errors of grammar or word order and unclear sentences, as well as provided a detailed comment explaining what I think they could do to improve their writing skills. *Table 14* lists my comments to these students on their writing task.

Table 14: My feedback to S3 and S5.

S3	S5
Hello, here are my comments on your report. First of all, I would like to point out that you did a really good job. I especially appreciate your use of linking words to	Here are my comments on your LR. First of all, you did a really good job: you understood the salient passages of the interview and you summarized them using your own words. There was only one

connect sentences: it is varied and sophisticated, so well done! My only slight criticism is that your report is rather short, so maybe you could have added some more details to the summary of the content. I would also suggest avoiding contracted forms (wouldn't → would not, I'm → I am etc). Moreover, a couple of terms could be replaced with more elaborate ones belonging to a more formal register and to academic writing (take a look at how I have rephrased your last sentence for example). Having said this, you did a very good job because you kept the two parts (summary and self-reflection) very balanced.	small inaccuracy in the summary of the content, which, however, you yourself noticed and pointed out. My only suggestion is to try and use more linking words for the sake of cohesion. For example, I think it would be best to divide this sentence and start the new one with 'however', as I suggested: "Furthermore, she recalled he was keen to answer to all kind of details as if he hadn't experienced the post-traumatic disorders. However, we nowadays well know that every victim suffers". So, my only suggestion is to expand your use of linking words. P.S. you forgot to mention the number of words at the end of your report.

As far as the evaluation grids are concerned, again I gave the students both a grade from poor to outstanding and a numerical grade. S3 obtained the final score 'good' (28/30). More specifically, I gave her 'outstanding' in grammatical accuracy and text organization and cohesion, and 'good' to the remaining criteria. S5, instead, obtained the final score 'outstanding' (30/30), as I actually gave her that score in all aspects.

The third pair was formed by two students: S4 with a B1+ level of written English and S6 also with a B1+ level. With regard to the evaluation grids, S4 gave the final score 'good' to her peer: 'good' in spelling and punctuation, vocabulary and overall task accomplishment, 'mediocre' in text organization and cohesion and 'outstanding' in grammatical accuracy. Interestingly, S4 also gave a numerical grade (24/30) to the student's listening report.

As regards the feedback provided by this student, *Table 15* illustrates the suggestions that S4 gave to her peer.

Table 15: Feedback of S4 to S6.

I find that your listening report is quite accurate, maybe I would have omitted the part of the comparison between the two different types of films and their budget and I would have focused on something else (maybe on the time of imprisonment he spent with the Japanese). However, I find that it is a very good job, but from the audio I think I understood that it was not a ship that was shipwrecked, but their plane, so maybe you should review this part (I also had problems understanding this part). As for grammar and punctuation, I didn't find any big mistakes.

S6 gave the final score ‘good’ to her peer: ‘good’ in spelling and punctuation, grammatical accuracy and overall task accomplishment, and ‘mediocre’ in vocabulary and text organization and cohesion. *Table 16* illustrates the feedback provided by this student.

Table 16: Feedback of S6 to S4.

In my opinion the report was really good, she was able to summarize the most important aspects of the interview. There are few errors of distractions, for example the name of the main character, the name “Pacific Ocean” and the movie’s name was wrote without the capital letter. Personally I didn’t found a good cohesion and linking between a few sentences and I think that this caused repetitions. I found a few grammatical errors but the text was comprehensible and I did not found any difficulties to understand the global concept.

As regards my feedback, again I sent both students their listening report back with my corrections, as well as provided feedback. *Table 17* lists my comments to these students on their writing task.

Table 17: My feedback to S4 and S6.

S4	S6
Hello, here is my feedback on your LR. First of all, I would like to point out that you did a good job, so well done! I appreciate that you’ve remembered to mention all aspects in the guidelines, as this shows that you’ve read the instructions carefully. Anyway, here are some suggestions on how you could improve for next time: - do not put the article ‘the’ before the year as I’ve noticed that you’ve made the same mistake more than once (In 1996, not in the 1996). - you concentrated on the most important details in the interview, but you could have expanded your summary a bit more as the total number of words was below the requested amount. - try keeping sentences shorter and use more diverse linking words as this helps text cohesion and fluency (e.g. however, nonetheless, moreover, in addition, furthermore, on the one hand… on the other hand, in spite of this, etc.).	Hello, here is my feedback on your listening report. Overall, you did a good job, so well done! I appreciate that you’ve kept the two parts (the summary of the interview and the ‘self-assessment’ part) quite balanced. As you can see from your LR, in some cases I wrote synonyms and I gave you some options on how you could rephrase certain sentences to make them clearer. I would also suggest using more diverse linking words, because teachers appreciate a text which is well-structured and cohesive. Having said this, you did use a few linking words, which helped text fluency. I just have a few other observations that might help you in the future: - try keeping sentences shorter, which also makes them clearer, - remember that the verb ‘to listen’ always needs preposition ‘to’ after (e.g. I love listening TO music/My mum never listens TO me etc.).

- word order: try not using sentences with 'it' as a subject. Always try using a whole other sentence as the subject (e.g. Instead of saying: "it was an honour for her to watch a great man…", say "watching a great man was an honour for her").	

As far as the evaluation grids are concerned, again I gave the students both a grade from poor to outstanding and a numerical grade. S4 obtained the final score 'good' (25/30), as I gave her 'good' for all criteria. S6, instead, obtained the final score 'good' (26/30): 'good' in all aspects except for 'outstanding' in grammatical accuracy.

Mendonça and Johnson (1994) identified several types of oral interactions that occurred during peer feedback[54]. I was also interested in identifying different types of (written) interactions during peer feedback between these two students.

In this case, students did indeed use some of the interactions noted by the two scholars. For example, S1 used 'suggestion' ("you said you didn't understand very well the audio so you should practice more on Listening Reports"), 'grammar correction' ("she talks" instead of "she talk"), and 'request for explanation' ("these sentences are not very clear and they are wrong. Can you rewrite them to explain them better?"). S2 also used some of the interactions mentioned by Mendonça and Johnson (1994), namely 'suggestion' ("some sentences could be shorter") and 'request for explanation' ("she who?"). Furthermore, this student also gave a suggestion to his peer on how she could improve text cohesiveness in a certain sentence: "here you should connect the sentences better using 'who'". As regards the second pair, S5 also gave suggestions ("I would suggest varying the use of linking words to keep up the rhythm and preserve the reader's attention" and "mind the use of some words like the verb to realize or the noun vision, they're 'false friends for Italian first language speakers"). In the third pair, S4 gave a suggestion to her peer, namely "from the audio I think I understood that it was not a ship that was shipwrecked, but their plane, so maybe you should review this part". Instead, S6 corrected some grammar mistakes, such as deleting the article 'the' before years and a past tense: "watched" instead of "watcht"[55].

[54] See Chapter Two, p. 60.

[55] This student sent me her peer's listening report back with highlighted words and passages (misused words and wrong sentences) and possible solutions for them. This is why these comments have not been listed in Table 16.

3.2.6 Discussion and Limitations

Given the results, it appears that all students enjoyed the activity overall and were happy with assessing their peer's listening report. Moreover, all students claimed that they were comfortable with having their peer assess their work. In addition, most students recognized the usefulness of having precise guidelines for peer feedback (i.e. 'useful phrases'), even though they did not use them extensively, and a rating scale in guiding them during the assessment process.

Overall, it seems that all students were capable of giving helpful and accurate advice to their peer on how they could improve their writing skills as preparation for the final exam in June. Indeed, students were often able to identify incorrect sentences and give partially correct solutions. For example, S4 wrote this sentence: "Needless to say, realizing such intense scenes were also challenging to manage from a financial perspective as an Independent Film". Her partner — S6 — was able to correct the sentence except for one minor slip: "Needless to say, shooting (to realize= comprendere, rendersi conto) such intense scenes was also challenging from a financial perspective as an Independent Film". I would have added, for example, "as this was/*Unbroken* was an independent film". Another example comes from the listening report of S3:

> "Zamperini grabbed attention on him in the 1936 in the Berlin Olimpics games, but the main event of his life was when his plane crushed in the 1943 and he survived for 47 days in the pacific Ocean in the shark infested water".

Her peer, S5, corrected this sentence as follows:

> "Zamperini grabbed the attention on him in 1936 in the Berlin Olympic games, but the main event of his life was when his plane crushed in 1943 and he survived for 47 days in the Pacific Ocean in the waters infested by sharks/in the sharks' infested waters".

As can been noted in the example above, S5 identified a wrong sentence and was able to correct some mistakes, such as the article 'the' before a year, the spelling of 'Olympic Games', and the capital letter in 'Pacific'. However, when she corrected the last part of the sentence, she gave a wrong suggestion, namely "in the sharks' infested waters". The right solution, in this case, would be 'in the shark-infested waters'.

Interestingly, all students claimed that providing a grade was more difficult than providing feedback. One student in particular, S5, claimed that she was happy they also

had the opportunity to provide feedback to "justify" the grade: "assign a grade, evaluate other people's works, means to judge it without justifications in some way I was glad there was a feedback too". This is actually at odds with Topping's claim, namely that giving qualitative feedback is actually "much more cognitively demanding for the assessor" than giving grades (2005: 640).

The results (see *Table 18*) that the students gave to each other as well as the results that they obtained from me also show a clear evidence of agreement between peer and teacher marks (both in poor to outstanding grades and in numerical grades, when these were provided also by students). However, it can be noted that the students who formed the first pair, namely S1 and S2, gave slightly higher numerical grades than I did. S1 in particular claimed in the pre-activity questionnaire that she does not like making other people "feel bad" about their work. This could, perhaps, be the reason why she gave a slightly higher grade to her peer than I did, although I am aware that it is not a big difference. Instead, in the third pair, S4 also gave a numerical grade. In this case, however, the grade that she gave to S6 is slightly lower than the grade I gave to S6: 24/30 and 26/30 respectively. Again, this is not a big difference.

Table 18: Peer/teacher rating agreement.

	Teacher	S1	S2
S1	Good (25/30)	/	Good (26/30)
S2	Mediocre (18/30)	Mediocre (20/30)	/

	Teacher	S3	S5
S3	Good (28/30)	/	Outstanding
S5	Outstanding (30/30)	Good	/

	Teacher	S4	S6
S4	Good (25/30)	/	Good
S6	Good (26/30)	Good (24/30)	/

Lastly, the different level of written English between S1 and S2 and between S3 and S5 did not hinder the peer assessment activity, even though, in the first pair, the difference was big. Indeed, although both students claimed they had a B1, the listening report revealed significant differences in the actual level of the students, but both students

were capable of providing helpful feedback to their peer, nonetheless. In the second pair, instead, the two students had a very similar level of written English.

This project has several limitations. First, as stated already, the very small number of students compared to the total number of course participants (about 400 students) limits the possibilities of making generalizations about peer assessment and makes my results statistically not relevant. Second, the students had not been trained during the course to carry out assessments; if they had been trained, they might have provided more detailed feedback. Having said this, the activity was still successful overall, because all students — from the lower-level ones to the higher-level ones — were able to provide helpful feedback to their peers. Last but not least, it is not possible to compare the results of these students with the results of the master students, because this last group of students designed the grid from scratch and only answered to a questionnaire, whereas the bachelor students had a grid ready to be filled in.

The next and final chapter will be devoted to the discussion of the results of a questionnaire on peer assessment, which was distributed among a cohort of university English teachers in the department of foreign languages of Padova University.

Chapter 4: A study of Peer Assessment with University English Teachers

Chapter Four will be devoted to the discussion of the results of a questionnaire on peer assessment and peer feedback, which was distributed among a cohort of teachers of English in the language department[56] of the University of Padova.

This chapter will be structured as follows: first, I will describe the participants of the questionnaire. Then, I will illustrate the questionnaire that was administered to the teachers. Last but not least, I will discuss the results. In the conclusion of my dissertation, I will discuss more thoroughly similarities and differences in the perceptions of peer assessment between these three different groups of people, namely first-year B.A. students and second-year M.A. students.

4.1 Participants

The participants to the questionnaire were a cohort of teachers of English[57] in the language department at the University of Padova. The questionnaire went to 14 teachers, but 9 responded to the questionnaire. The respondents have taught different classes, namely 1st-year bachelor, 2nd-year bachelor, 3rd-year bachelor, 1st-year master and 2nd-year master. Most teachers have taught both bachelor and master students, two teachers have taught classes to students of all years (1st-year bachelor, 2nd-year bachelor, 3rd-year bachelor, 1st-year master and 2nd-year master), whereas only one teacher has only taught bachelor students.

All the questionnaire participants stated that they have carried out peer assessment with their students and, thus, have had previous experience with this particular form of alternative assessment. More specifically, 7 teachers have also asked their students to give a grade to their peers' performance.

4.2 The Questionnaire

The questionnaire was anonymous and consisted of ten questions. A big majority of the questions were multiple-choice questions, while only two were open-ended.

[56] *Dipartimento di Studi Linguistici e Letterari* (DISLL).
[57] Both professors and CELs (collaboratori ed esperti linguistici).

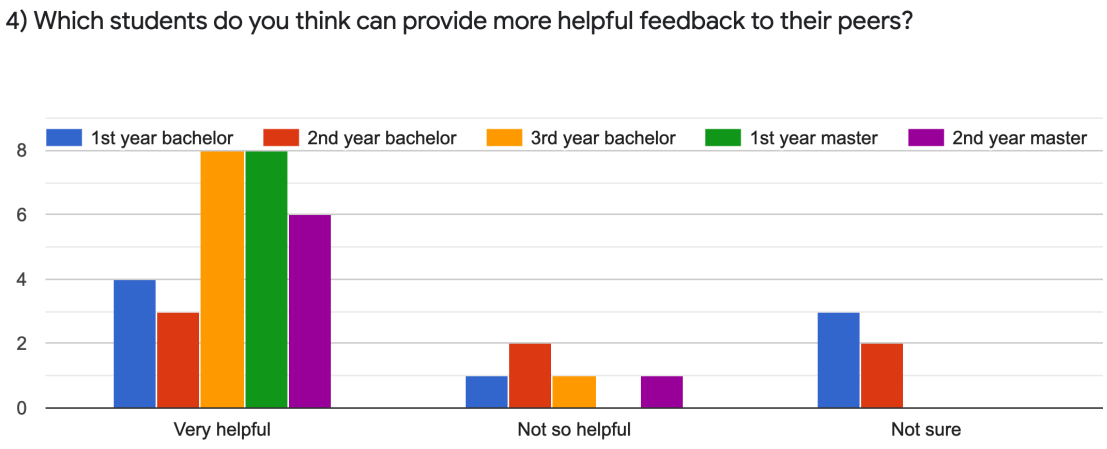
Moreover, all of the questions except one were obligatory and marked with an asterisk (*).

The first three questions aimed at collecting some background knowledge about the teachers, such as which classes they have taught, and if they have ever asked their students to carry out peer assessment or peer feedback. Instead, the remaining questions were all related to gaining some knowledge about the teachers' perceptions of peer assessment, such as whether language proficiency affects students' ability to provide helpful feedback to their peers, which level of students they consider as more suited for peer assessment, and how successful the peer assessment activities have been in their experience.

4.3 Results

As stated above, one of the questions sought to find out the teachers' opinions on which students can provide more helpful feedback to their peers. *Figure 1* illustrates the answers to this question.

Figure 1: Students who are more suited for effective peer feedback.



As can be seen in Figure 1, the majority of the participants thought that the higher the classes the students are in, the more helpful the feedback that they can provide.

The following question is related to question four and sought to find out whether the teachers thought that language proficiency may affect students' ability to provide helpful feedback. Six teachers thought that this was the case, whereas only one teacher thought that language proficiency does not affect students' ability to provide helpful

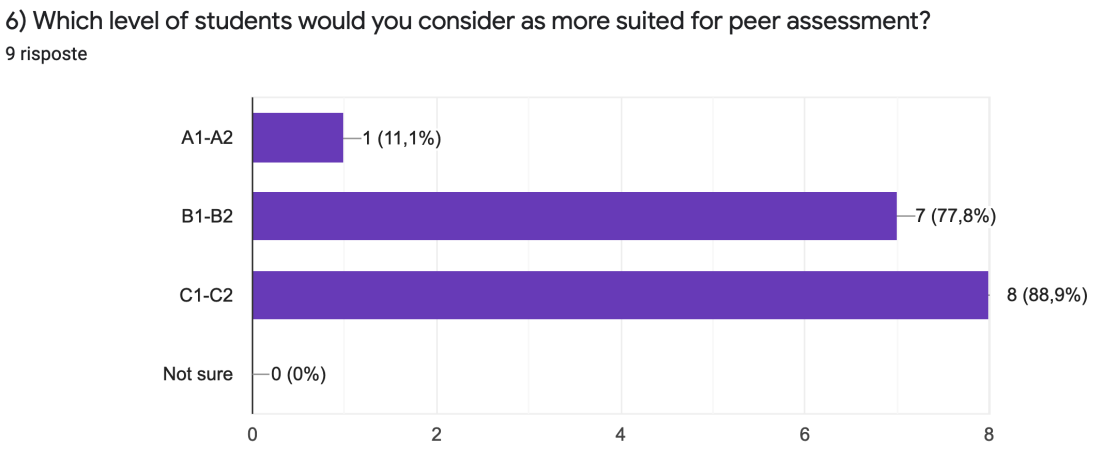
feedback. The remaining wo teachers chose the option 'not sure'. One of the two teachers explained his/her reaction to this question later in the questionnaire:

> "I found it difficult to answer q5 as obviously language proficiency does play its part but is not the main focus. I see peer review activities as promoting autonomy, promoting critical thinking and the importance of 'noticing' how a language works and of building language use and knowledge through mutual exchange, this can be done whatever the proficiency level from simple tasks - e.g. did you understand the same thing I did? to the more complex structuring of say, academic texts" (T2).

Moving on to question six, which was also related to the previous two questions, I asked teachers which level of students they would consider as more suited for peer assessment (*Figure 3*). They could choose between A1-A2, B1-B2, and C1-C2.

As shown in Figure 3, almost all teachers agreed that students with levels ranging between C1-C2 (8) and B1-B2 (7) are more suited for this kind of activity. Only one teacher said that students of all levels can carry out successful peer assessment activities. Unexpectedly, the teacher who thought that students' language proficiency does not affect their ability to provide feedback, said that students with levels B1-B2 are more suited for peer assessment. In this case, I would have expected that he/she had selected all levels.

Figure 3: Levels of students for more successful peer assessment.



Next, questions seven and eight sought to find out the degree of success of the peer assessment activities performed with students as well as which kind of activities they asked their students to do. Question seven was the only one which was not obligatory, because I did not know whether or not all teachers had ever carried out any kind of peer

assessment with their students. However, as already mentioned, all teachers claimed they had asked their students to perform peer assessment at some point.

The results of question seven reveal that a big majority of teachers thought that the activities were successful. More specifically, five teachers thought they were very successful and scored 8 and 9 on the scale, three teachers thought they were successful and scored 7, whereas only one teacher thought they had not been very successful and assigned a 5. As regards the peer assessment activities conducted in class, *Table 1* lists the answers that the teachers gave to this question. I have added the label 'T + number' at the end of each quotation to indicate which teacher said what.

Table 1: Different kinds of peer assessment activities.

More than peer assessment I've got classes to do peer feedback/revision activities where they read a peer's work and write comments, following guidelines. More recently I've asked students assess their peers' oral skills. (T1)
Having them check the formal, content-related and structural adequacy of peers' speech/writing against a list of criteria; having them look for the presence of given components/aspects of language use in peers' speech/writing. (T2)
Activities aimed at promoting awareness: 1) of language use; vocab/grammar 2) structure of written text genres and 3) to work together to understand spoken text. (T3)
I have taken numerous approaches but I find they are most successful when they are anonymous, e.g. having students fill out a Google Form such as this one or edit a Google Document anonymously. (T4)
Feedback on presentations given by other students. (T5)
They've had to assess another student's writing (fluency, accuracy, content, composition, goal achievement etc.) and speaking (fluency, accuracy, content, pronunciation, cohesion etc.). (T6)
I asked them to peer assess their reading reports. (T7)
Give feedback on listening reports, paragraphs, essays, audio recordings, etc in the form of checklists, short answers and comments. (T8)
Peer review of various kinds of texts: newspaper articles, summaries, academic essays. (T9)

As the answers listed in Table 1 reveal, teachers have asked their students to carry out all kinds of peer assessment activities, from assessing peers' presentations to assessing their writing tasks, and to peer revisions.

Last but not least, I gave them room for further observations or comments about peer assessment. *Table 2* lists the comments that teachers left in this question.

Table 2: Teachers' further observations and comments about PA.

I've found that students need very clear guidelines about how to carry out peer assessment/peer feedback activities. [...] Students need advice on what constructive feedback is as, from my experience, they tend to pay compliments, but are afraid to point out changes that need to be made or to suggest corrections. It needs to be made clear that they are not replacing the teacher or doing the teacher's job and that they do not need a high level of English to give feedback. (T1)
A) Proficiency is crucial -- if you don't 'know', you can't spot 'problems' in others' speech and writing. [...] B) Hearsay (students): they don't trust the validity/usefulness of peer feedback, since they think their peers are not knowledgeable enough. C) Hearsay (CELs): students are loath to give accurate feedback to their peers for fear of hurting their feelings [...]. (T2)
I found it difficult to answer q5 as obviously language proficiency does play its part but is not the main focus. I see peer review activities as promoting autonomy, promoting critical thinking and the importance of 'noticing' how a language works and of building language use and knowledge through mutual exchange, this can be done whatever the proficiency level from simple tasks [...]. (T3)
The primary reason I do peer assessment is to raise awareness and help develop students' critical skills. I feel often they cannot be critical enough of their own work or others' because they've been educated in a system where the only type of assessment that matters comes from the teacher [...]. It's also not true, i.e. peers often have insights that teachers don't have so it can actually be enriching. (T4)
Students tend to over-focus on grammar. (T5)
I have found that peer assessment is very useful. By analyzing others' work, it definitely gets students in the habit of critically analyzing their own and thinking about language points or issues they didn't consider with regard to their own work. [...] In many cases it can be more effective than a teacher correcting students' language work because it's a more interactive experience [...]. (T6)
[...] I think that students should be given very clear instructions as to how to go about the assessment of their peers' performance, as they found it difficult to decide what to focus on. Also, the peers should be at about the same level of language proficiency, which in my view should be at least the B2/C1 level. Finally, some students did not want to be too negative about their peers' performance and tended to be 'nice' to them. (T7)
I think peer assessments help not only the receivers of the assessments but also the students doing them especially if the questions \ guides the teacher sets are oriented to the task at hand, that is, lead the student to reflect on language, the skill, etc. (T8)

I feel it is a good way to encourage active collaborative learning and critical thinking. (T9)

The data presented in Table 2 show that teachers had very diverse comments to make about peer assessment, but common threads can be found. Most comments touch on several aspects which are particularly relevant to my dissertation. For example, several teachers noticed that students need very clear guidelines of what they are supposed to do, being this providing feedback, peer reviewing, or providing a grade to a peer's performance: "I've found that students need very clear guidelines about how to carry out peer assessment/peer feedback activities" (T1), "I think that students should be given very clear instructions as to how to go about the assessment of their peers' performance, as they found it difficult to decide what to focus on" (T7), "I think peer assessments help not only the receivers of the assessments but also the students doing them especially if the questions\guides the teacher sets are oriented to the task at hand, that is, lead the student to reflect on language, the skill, etc." (T8). These comments also link in with what the literature on peer assessment says (see, for instance, Kroll 1991; Herman et al. 1992; Hattie 2012; Keller and Westfall-Greiter 2014), namely that students often need clear instructions and models to rely on.

Another interesting comment, which also comes up very often in the literature of peer assessment, was again made by one of these teachers: "the peers should be at about the same level of language proficiency, which in my view should be at least the B2/C1 level" (T7). As discussed in Chapter Two, peer assessment is best performed between students with similar status (Topping 2005; Jung 2016).

One teacher raised an issue in peer assessment, namely that students tend to be overgenerous and "pay compliments" (T1) rather than provide constructive feedback. Again, this teacher stresses the importance of having clear instructions and of explaining students what constructive criticism consists of. Other teachers shared the same view as they claimed that "students are loath to give accurate feedback to their peers for fear of hurting their feelings" (T2), and "some students did not want to be too negative about their peers' performance and tended to be 'nice' to them" (T7).

Two teachers also commented on the fact that students might be even more suited to provide feedback than teachers: "in many cases it [peer assessment] can be more effective than a teacher correcting students' language work because it's a more interactive

experience" (T6) and "peers often have insights that teachers don't have" (T4). These comments have also been pointed out in the literature on peer assessment: students assessing each other may lead "to outcomes at least as good as teacher assessment and sometimes better" (Topping 1998: 262). As pointed out in Chapter Two, however, peer assessment is not meant as a replacement for teacher assessment. This view was also shared by one teacher, who stated that "it needs to be made clear that they are not replacing the teacher or doing the teacher's job" (T1). On the contrary, as observed by Peng (2010: 90), peer assessment is a "supplementary assessment method for involving and empowering students rather than a substitution for teacher assessment".

Lastly, some teachers also explained why they carry out peer assessment activities with their students. The most frequent answers were to encourage critical thinking and develop students' critical skills (T3, T4, T6, and T9). Furthermore, two teachers claimed that the reason why they do peer assessment is that it raises students' awareness (T4) and promotes learner autonomy (T3). Especially these last two reasons were those mentioned by Professor Ackerley as to why she asked her second-year master students, investigated in Chapter Three, to carry out peer assessment during the *Spoken English* module. As mentioned in Chapter One, one of the many benefits of using alternative forms of assessment is that these can enhance learner autonomy. More specifically, peer assessment seems to be an exceptional method to foster learner autonomy, since students, by judging the work of others, will become more aware of their own skills and will presumably be more apt to also improve their own performances. Most of these benefits of peer assessment that were noted by teachers were also noted by Hansen Edwards[58].

I also scheduled a *Zoom* meeting[59] with one of the teachers, Professor Dalziel, to further discuss the topic of peer assessment. Indeed, after I had carried out my study with first-year students, I had some other questions to ask. Firstly, I wanted to understand this teacher's position with regards to which aspects of a written performance students of different levels should be expected to focus on. More specifically, I asked whether students with an A2 level compared to students with a C1 level should be expected to make a 'global' analysis of a peer's work[60] or rather to only focus on a few aspects.

[58] See Chapter Two, pp. 49-50.
[59] I have recorded this meeting with Professor Dalziel's permission so as to be able to quote some of her assertions.
[60] In this case, I made reference to a written performance, such as a listening report.

According to Professor Dalziel, peer assessment consists in "having another reader to give global feedback" on a piece of writing, for instance. Therefore, "it is up to the reader to focus on grammar or vocabulary", depending on what the assessor "spots" in his/her peer's work. In this regard, Professor Dalziel also suggested having students reflect on what constructive feedback is and on how to politely correct mistakes. More specifically, asking students to identify criteria for "good academic writing", for example, could be a good way to engage them more in the process of giving relevant and helpful feedback. According to this teacher, what is important is "having another perspective that is not just the teacher's" and, consequently, what usually happens is that students themselves say that giving feedback is helpful "for the person giving feedback rather than for the person receiving it, because they can go back and reflect on their own writing". The positive repercussions that peer assessment can have on self-reflection and self-assessment have also been discussed in *§1.4*. Furthermore, Professor Dalziel also pointed out that feedback could be done in the first language in case of lower level students[61] (see Peng 2010) to make them more confident in their evaluations.

Another question that came to my mind after doing my project with bachelor students was whether teachers should intervene in case of incorrect feedback. In this regard, what Professor Dalziel suggests is avoiding "adulterating" feedback given by students, but rather teachers could give "feedback on the feedback", pointing out, for examples, the difficulties that students had assessing certain aspects of a peer's performance. Additionally, "it should be made clear to students that they might not always agree with their peers' comments" and that they should take what they think is relevant to them from those comments.

My last question regarded asking students to give grades, being these numerical grades or just adjectives (like I asked bachelor students to do). Although Professor Dalziel has never asked her students to give grades, because she only wanted them to "have comments so that they could improve their work", she believes that "in the comments you have some kind of judgement coming out anyway" and that usually grades are also a "reflection of how comfortable students feel in giving them". Indeed, sometimes students

[61] See Chapter Two, p. 64 for more information.

might give very high grades just because they are afraid of criticizing their friends and feel uncomfortable about doing that.

4.4 Limitations

I am aware that these results are not entirely conclusive, because of the low number of respondents. What is more, the teachers who answered my questionnaire were only those who have carried out peer assessment activities with their students and who find peer assessment beneficial for a number of reasons. Nevertheless, the results show that teachers noted that peer assessment needs careful implementation and very precise guidelines and instructions to be successful.

In the next pages I will draw my conclusions about peer assessment based on the data that I have collected from students (Chapter Three) and teachers (Chapter Four).

Conclusion

This section will be devoted to the conclusion of my dissertation, in which I will discuss the results of my studies with first-year bachelor students of languages and second-year master students of languages of the University of Padova. Furthermore, I will also discuss the findings of a questionnaire, which was distributed among a cohort of teachers of English in the language department of Padova University (*Dipartimento di Studi Linguistici e Letterari*).

With these studies, I aimed to investigate the different perceptions of peer assessment by different target groups, namely students and teachers. More specifically, I wanted to investigate the differences between second-year master students, especially those who are aspiring schoolteachers, and first-year bachelor students. As for teachers, I wanted to see whether the comments they had about peer assessment were in line with what the literature on PA says and with what students also had to say about this topic.

As regards the first group of students, namely the M.A. students, they completed a short questionnaire which sought to examine their opinions about a peer assessment activity on presentations which they carried out during a *Spoken English* module in the winter semester 2019-2020. More specifically, I wanted to see whether there were any significant differences between those students who are aspiring schoolteachers and those who are not. Moreover, I also wanted to find out whether there were any significant differences between male and female students and between younger and older students (the age of the participants ranged between 18 and over 30 years old). However, of the total 12 students who participated, only one was not an aspiring schoolteacher and only one student was over 30 years old. Furthermore, most students were female students, whereas only two male students answered the questions.

All in all, the questionnaire findings reveal no significant differences between students: all of the students agreed on the usefulness of peer assessment, especially to increase learner autonomy and raise students' aware of the assessment process. Only a couple of students seemed a bit reluctant toward the activity, claiming that they felt under pressure about assessing friends and that assigning scores should be the teacher's responsibility. Interestingly, one student claimed that unpractised assessors tend to be "too harsh" when it comes to assessing a peer. However, this is actually at odds with what

some of the teachers interviewed thought, namely that students tend to be overgenerous and pay compliments, rather than giving helpful and constructive feedback.

I also asked students whether they felt up to the task when they were asked to assess their peers using a rating scale, which they also designed from scratch. Only half of the students did feel up to the task, but the other half of the students did not feel well enough prepared, thus suggesting that students need very precise guidelines and instructions about what they are expected to do, as well as a lot of practice. This aspect was also discussed in detail by teachers in their questionnaire.

Lastly, I sought to investigate these students' opinions about biases in peer assessment, given that most of these students knew each other quite well and that there were numerous friendships between them. Indeed, two students claimed that their peers were "really kind" and gave "more positive feedback than expected". Others, however, claimed that negative biases are also often present in peer assessment activities, possibly due to "competitiveness, dislike and fear for repercussions".

As for the second group of students, namely the six B.A. students who participated in my project, they also completed two questionnaires, namely a pre-activity questionnaire and a post-activity questionnaire, which aimed to keep track of any change of opinion about peer assessment. Furthermore, these students were asked to write a short listening report of about 300 words and mutually assess one of their peers' report, with whom they had been matched. Contrary to the master students, the peer assessment activity was conducted on writing skills. The students were given very precise guidelines on how to give feedback and were encouraged to use some 'useful phrases' for peer feedback (taken from Keller and Westfall-Greiter 2014). Indeed, numerous experts (see, for instance, Kroll 1991; Herman et al. 1992; Hattie 2012; Keller and Westfall-Greiter 2014) agree that it is important to give students clear instructions on what they have to do (e.g. what constructive criticism is, how to politely correct mistakes etc.) for peer assessment to be successful. Furthermore, besides writing a detailed comment, students were also expected to fill in a rating scale, which I had designed specifically for this group of students for writing skills, and to give a score from poor to outstanding.

My project with these students was completely anonymous: all students had to send me their reports via e-mail and I then swapped the reports and sent them to the matching peer for correction.

I also wanted to investigate the perceived objectivity of peer assessment with this group of students and compare their answers with those of master students. Overall, the results show that bachelor students are more doubtful of the objectivity of PA than the master students, although they gave relatively high scores to the scale. However, when they were asked to motivate their choice, most of them said that teachers' evaluations are more objective than the students'. Only one student claimed that teacher evaluations and student evaluations are equally objective, especially if clear and precise criteria are involved.

As mentioned above, students were asked to provide detailed feedback on the listening report of their peer as well as to assign a grade from poor to outstanding. Overall, all students claimed that giving a grade was more difficult than providing feedback, which is at odds with Topping's view that giving qualitative feedback is "much more cognitively demanding for the assessor" than giving grades (2005: 640). In spite of this, the rating agreement between the grades given by students and my grades is evident.

As for teachers, they were sent a questionnaire about peer assessment and peer feedback with which I sought to investigate their position toward this alternative method of assessment, especially given their experience with peer assessment at University. In total, nine teachers responded and gave very detailed and answers, which were very relevant and timely for my dissertation, as they touched on several aspects that have been thoroughly discussed in the previous chapters. For example, several teachers noticed that students need very clear guidelines of what they are supposed to do in terms of providing feedback, providing grades, or peer reviewing. Furthermore, one teacher commented on the fact that peers should be at about the same level of language proficiency, which was also discussed in Chapter Two and Chapter Three.

One very interesting comment revolved around students being overgenerous in their feedback, because they might be afraid of criticizing friends. Students, especially master students, did also point out that sometimes students tend to be "too kind" in their evaluations. One of the reasons for this was that they might "expect the same kindness in return". On the contrary, the teachers who expressed this view claimed that the reason is that students might be afraid of offending friends.

Lastly, some teachers explained why they include peer assessment in their courses. The most frequent answers were to encourage critical thinking and develop

students' critical skills, to raise students' awareness and to promote learner autonomy. These benefits of peer assessment have also been mentioned by Hansen Edwards (see Chapter Two, pp. 49-50).

I also had the opportunity to further discuss this topic in a *Zoom* meeting with one of the teachers, Professor Dalziel, allowing me to ask other questions that had come to mind after I had carried out my project with bachelor students (after the questionnaire had been sent to teachers). According to this teacher, peer assessment also has very positive repercussions on self-evaluation, as students will reflect on their own work by looking at the work of others, thus promoting learner autonomy. This particular benefit of peer assessment has also been pointed out in Chapter One and Chapter Two.

Comparing the results of my studies with master students and with bachelor students is difficult for a number of reasons. First of all, the peer assessment activity carried out by master students was very different from the activity which was carried out by bachelor students. Indeed, the first group of students was asked to design a rating scale for oral skills, whereas the second group was given a grid for writing skills. Furthermore, my study with master students was based entirely on a questionnaire, whereas my project with bachelor students was more complete as it consisted in two questionnaires, a listening report and peer assessment of the reports. Moreover, I am aware that the number of participants in the two studies is very different: twelve students out of twenty-four course participants in the case of master students, and only six students out of several hundreds of students in the case of bachelor students.

Having said this, I believe that a similar tendency toward peer assessment can be observed in these two group of students. First, both groups thought that peer assessment helped them in different ways. The main thing that I noticed was that a few master students who are aspiring schoolteachers claimed that doing peer assessment helped them to understand the major issues involved in assessment, and said that playing a 'teacher-like role' was a "considerable opportunity" for them in view of their future job aspiration. Instead, first-year bachelor students, who might not yet have clear ideas about their future career, commented on the fact that doing peer assessment helped them to identify mistakes that would have otherwise gone overlooked and gave them the opportunity to compare their work with that of a peer of similar level, thus not really seeing it as an opportunity for a future career (or, at least, they did not say). Second, bachelor students

agreed that anonymity in peer assessment could ensure fair and objective evaluations, which was also one of the possible solutions for biased assessments that was proposed by one of the master students. Last, in both groups of students some said that they did not feel up to the task, thus suggesting that peer assessment needs training, practice and clear instructions regardless of the age of the students carrying out peer assessment. As some of the teachers interviewed suggested, students also need to be instructed on what constructive criticism is, in order to give helpful feedback.

If I were to carry out a similar study in the future, I would do some things differently. First, I would train students to carry out assessments, so that they could feel more confident and comfortable in their evaluations. Indeed, some students belonging to both groups felt uncomfortable in having to judge their colleagues. Then, I would try having a similar number of students to make a more statistically relevant comparison of the results. Last but not least, I would possibly ask students to carry out peer assessment in smaller groups rather than in pairs, to see if there are any significant differences between them in the feedback that they give to each other. In this scenario, I think that it would also be very interesting to have students assess a friend and someone they do not know as well, to see if there are any significant differences in the two evaluations.

Given the overall positive results of my studies with students and the positive feedback of teachers about peer assessment and peer feedback, it seems that peer assessment is an exceptional tool for a number of reasons. First, it raises awareness about the assessment process and about one's performance. Then, peer assessment encourages students to take charge of their own learning. Consequently, peer assessment also has positive repercussions on self-assessment, thus promoting learner autonomy and lifelong learning. Last but not least, peer assessment can also ensure swifter and richer feedback than teacher assessment.

Given these premises, peer assessment should be implemented even in schools, so that students can become more engaged in their own learning (including assessment) at a younger age.

Bibliography

APA Dictionary of Psychology (online). Retrievable from: https://dictionary.apa.org/

Azarnoosh, M., (2013) "Peer Assessment in an EFL Context: Attitudes and Friendship Bias", *Language Testing in Asia*, Vol. 3, No. 11, pp. 1-10. DOI: https://languagetestingasia.springeropen.com/track/pdf/10.1186/2229-0443-3-11 (last visited: 20/05/2020).

Bachman, L.E., Palmer, A.S., (1996) *Language testing in practice.* NewYork: Oxford University Press. Cited in Brown (2004).

Bailey, A.L., (2017) "Assessing the Language of Young Learners". In Shohamy, E., Or, I.G., May, S. (eds.)*, Encyclopedia of Language and Education* (3rd edition), Vol. 7 (*Language Testing and Assessment*), pp. 323-342. New York: Springer International Publishing.

Balboni, P.E., (1994) *Didattica dell'italiano a stranieri*. Roma: Bonacci. Cited in Novello (2012).

Balboni, P.E., (2006) "Il piacere di imparare, il piacere di insegnare". In Serragiotto, G. (ed.), *Il piacere di imparare, il piacere di insegnare*, pp. 19-28, Vicenza: La Serenissima. Cited in Novello (2012).

Balboni, P. E., (2015) *Le sfide di Babele: insegnare le lingue nelle società complesse*, UTET Università.

Brew, A., (1999) "Towards autonomous assessment: using self-assessment and peer assessment". In Brown, S., Glasner, A. (eds.) *Assessment matters in higher education*. Buckingham: Open University Press. Cited in Liu and carless (2006).

Brown, H.D., (2000) *Principles of language learning and teaching*, 4th Edition. White Plains, NY: Pearson Education.

Brown, H.D., (2004) *Language Assessment: Principles and Classroom Practices*. NY: Pearson Education.

Brown, J.D., Bailey, K.M. (1984) “A categorical instrument for scoring second language writing skills”, *Language Learning*, Vol. 34, No. 4, pp. 21-42. Cited in Brown, H.D. (2004).

Cameron, L., (2001) *Teaching languages to young learners*. Cambridge: Cambridge University Press. Cited in Novello (2012).

Caon, F., Tonioli, V., (2016) “La sfida delle classi ad abilità linguistiche differenziate (CAD) in Italia e in Europa”. In Melero Rodríguez, C.A. (ed.) *Le lingue in Italia, le lingue in Europa: dove siamo, dove andiamo*, pp. 137-154, Edizioni Ca’ Foscari.

Çelik, S., Türkan, S., (2014) “Assessment and evaluation in EFL: Classroom methods, assessment tools, and the instructional implications of high-stakes exams”. In Çelik, S. (ed.), *Approaches and principles in English as a foreign language (EFL) education*, pp. 419-428. Ankara, Turkey: Eğiten. DOI: https://www.researchgate.net/publication/304535480_Celik_S_Turkan_S_2014_Assessment_and_evaluation_in_EFL_Classroom_methods_assessment_tools_and_the_instructional_implications_of_high-stakes_exams_In_S_Celik_Ed_Approaches_and_principles_in_English_as_ (last visited: 09/04/2020).

Cheng, W., Warren, M., (2005) “Peer assessment of language proficiency”, *Language Testing*, Vol. 22, No. 1, pp. 93-121. Cited in Peng (2010).

Council of Europe, (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press. Cited in Huttunen (2003).

Dam, L., (2003) “Developing learner autonomy: the teacher’s responsibility”. In Little et al. (eds.), *Learner autonomy in the foreign language classroom: teacher, learner, curriculum and assessment*, pp. 135-146. Dublin: Authentik Language Learning Resources Ltd.

Dam, L., Legenhausen, L., (2010) “Learners reflecting on learning: Evaluation vs testing in autonomous language learning”. In Param, A., Siercu, L. (eds.), *Testing the untestable in language education*, pp. 120-139. Bristol: Multilingual Matters. DOI: https://www.researchgate.net/publication/292244369_Learners_reflecting_on_learning_Evaluation_versus_testing_in_autonomous_language_learning (last visited: 12/05/2020).

DeNisi, A.S., Alan Randolph, W., Blencoe, A.G., (1983) "Potential Problems with Peer Ratings", *The Academy of Management Journal*, Vol. 26, No. 3, pp. 457-464. DOI: https://www.jstor.org/stable/256256?seq=1 (last visited: 15/05/2020).

Donato, R., (1994) "Collective scaffolding in second language learning". In Lantolf, J.P., Appel, G. (eds.), *Vygotskian Approaches to Second Language Research*, pp. 33-56. Norwood, NJ: Ablex Publishing Corporation. Cited in Thomsen (2003).

Dörnyei, Z., Skehan, P., (2003) "Individual differences in second language learning". In Doughty, C.J., Long, M.H. (eds.), *The Handbook of Second Language Acquisition*, pp. 549-630. Hoboken, New Jersey: Blackwell Publishing Ltd. Cited in Griffiths (2008).

Double, K., McGrane, J., Hopfenbeck, T.N., (2019) "The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies", *Educational Psychology Review*, pp. 1-29. DOI: https://link.springer.com/article/10.1007/s10648-019-09510-3#citeas (last visited: 18/05/2020).

Elbow, P., Belanoff, P., (1986), "Portfolios as a Substitute for Proficiency Examinations", *College Composition and Communication*, Vol. 37, No. 3, pp. 336-339. Cited in Fox (2017).

Falchinov, N., Goldfinch, J., (2000) "Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks", *Review of Educational Research*, Vol. 70, No. 3, pp. 287-322. DOI: https://www.researchgate.net/publication/244504362_Student_Peer_Assessment_in_Higher_Education_A_Meta-Analysis_Comparing_Peer_and_Teacher_Marks (last visited: 15/05/2020).

Falchinov, N., (2003) "Involving students in assessment", *Psychology Learning and Teaching*, Vol. 3, No. 2, pp. 102-108. DOI: https://www.researchgate.net/publication/228609586_Involving_Students_in_Assessment (last visited: 15/05/2020).

Falchinov, N., (2007) "The Place of Peers in Learning & Assessment". In Boud, D., Falchikov, N. (eds.), *Rethinking assessment in higher education*, pp. 128-143. London: Routledge. Cited in Spiller (2012).

Fox, J., (2017) "Using Portfolios for Assessment/Alternative Assessment". In Shohamy, E., Or, I.G., May, S. (eds.)*, Encyclopedia of Language and Education* (3rd edition),

Vol. 7 (*Language Testing and Assessment*), pp. 135-147. New York: Springer International Publishing.

Freeman, M., (1995) "Peer assessment by groups of group work", *Assessment and Evaluation in Higher Education*, Vol. 20, No. 3, pp. 289–300. Cited in Peng (2010).

Gan, J.S.M., (2011) *The effects of prompts and explicit coaching on peer feedback quality* (Doctoral thesis), The University of Auckland, Auckland, New Zealand. Cited in Hattie (2012).

García, G.E., Pearson, P.D., (1991) "The role of assessment in a diverse society". In E.F. Hiebert (ed.), *Literacy for a diverse society*, pp. 337- 391. New York: Teachers College Press. Cited in Huerta-Macías (2002).

Gipps, C., (1994) *Beyond testing*, Brighton: Falmer. Cited in Novello (2012).

Griffiths, C., (2008) "Strategies and good language learners". In Griffiths, C. (ed.) *Lessons from Good Language Learners*, pp. 83-98. Cambridge: Cambridge University Press.

Gronlund, N.E., (1998) *Assessment of student achievement*, 6th edition. Boston: Allyn and Bacon. Cited in Brown (2004).

Han, A., (2011) "Gaining insights into student reflection from online learner diaries", *Journal of the European Confederation of Language Centres in Higher Education (CercleS)*, Vol. 1, No. 1, pp. 195-210. DOI: https://www.researchgate.net/publication/274885723_Gaining_insights_into_student_reflection_from_online_learner_diaries (last visited: 14/04/2020).

Hansen Edwards, J.G., (2013) "Peer Assessment in the Classroom". In Kunnan, A.J. (ed.) *The Companion to Language Assessment*, Vol. 2, Part 6, pp. 730-750. Hoboken, NJ: Wiley-Blackwell.

Harris, D.P., (1968) *Testing English as a Second Language*, New York: McGraw Hill. Cited in Hughes (2003).

Hattie, J., (2012) "The flow of the lesson: the place of feedback". In Hattie, J. (ed.) *Visible Learning for Teachers. Maximizing impact on learning*, pp. 129-154. London/New York: Routledge.

Herman, J.L. et al., (1992) *A Practical Guide to Alternative Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Holec, H., (1981) *Autonomy and foreign language learning*. Oxford: Pergamon. Cited in Novello (2014).

Huerta-Macías, A., (2002) "Alternative Assessment: Responses to Commonly Asked Questions". In Richards, J.C., Renandya, W.A. (eds.) *Methodology in Language Teaching. An anthology of Current Practice*, pp. 338-343. Cambridge: Cambridge University Press.

Hughes, A., (2003) "Reliability". In *Testing for Language Teachers*, 2nd Edition, pp. 36-52. Cambridge: Cambridge University Press.

Hughes, A., (2003) "Testing Writing". In *Testing for Language Teachers*, 2nd Edition, pp. 83-112. Cambridge: Cambridge University Press.

Huttunen, I., (2003) "Planning learning: the role of teacher reflection". In Little et al. (eds.) *Learner autonomy in the foreign language classroom: teacher, learner, curriculum and assessment,* pp. 122-134. Dublin: Authentik Language Learning Resources Ltd.

James, M., Pedder, D., (2006) "Professional learning as a condition for assessment for learning". In Gardner, J. (ed.) *Assessment and learning*, pp. 27-44. London: Sage Publications. Cited in White (2009).

Jung, M., (2016) "Peer/Teacher-Assessment Using Criteria in the EFL Classroom for Developing Students' L2 Writing", *Journal of Pan-Pacific Association of Applied Linguistics*, Vol. 20, No.1, pp. 1-20. DOI: https://www.semanticscholar.org/paper/Peer%2FTeacher-Assessment-Using-Criteria-in-the-EFL-Jung/3f4a313396e868f896a78234a88642167947a0c8 (last visited: 19/05/2020).

Keefe, J., (1979) *Student Learning Styles: Diagnosing and Prescribing Programs*. Reston, VA: National Association of Secondary School Principals. Cited in Brown (2000).

Keller, J.M., (1983) "Motivational design of instruction". In Reigelruth, C.M. (ed.) *Instructional Design Theories and Models: An Overview of Their Current Status*, pp. 383-434. Hillsdale, NJ: Lawrence Erlbaum Associates. Cited in Brown (2000).

Keller, S., Westfall-Greiter, T., (2014) “How to give effective Feedback”, *Der Fremdsprachliche Unterricht Englisch*, Vol. 48, No. 130, pp. 8-11.

Krashen, S., (1981) *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press. Cited in Brown (2000).

Kroll, B., (1991) “Teaching writing in the ESL context”. In Murcia, C. (ed.) *Teaching English as a second language*. New York: Newbury House. Cited in Peñaflorida (2002).

Kunnan, A.J., (2013) *The Companion to Language Assessment*, Vol. 2, Part 6. Hoboken, NJ: Wiley-Blackwell.

Lewis, M., (1997) *Implementing the Lexical Approach*, Language Teaching Publications, Hove, England. In Keller and Wesfall-Greiter (2014).

Little, D., (1991) *Learner autonomy (1): Definitions, issues and problems*. Dublin: Authentik Language Learning Resources Ltd. Cited in Novello (2014).

Little, D. (2002) “The European Language Portfolio: Structure, origins, implementation and challenges”, *Language Teaching*, Vol. 35, No. 3, pp. 182–189. DOI: https://www.researchgate.net/publication/231870127_The_European_Language_Portfolio_Structure_origins_implementation_and_challenges (last visited: 17/04/2020).

Little, D., (2003) “Learner autonomy and public examinations”. In Little et al. (eds.) *Learner autonomy in the foreign language classroom: teacher, learner, curriculum and assessment,* pp. 223-233. Dublin: Authentik Language Learning Resources Ltd.

Little, D., Ridley, J., Ushioda, E., (2003) *Learner autonomy in the foreign language classroom: teacher, learner, curriculum and assessment.* Dublin: Authentik Language Learning Resources Ltd.

Liu, N., Carless, D., (2006) “Peer feedback: The learning element of peer assessment”, *Teaching in Higher Education*, Vol. 11, No. 3, pp. 279-290. DOI: https://www.researchgate.net/publication/253967244_Peer_feedback_The_learning_element_of_peer_assessment (last visited: 13/05/2020).

Lynch, B.K., (2001) “The ethical potential of alternative language assessment”. In Elder, C. (ed.) *Experimenting with uncertainty: Essays in honour of Alan Davies*, pp. 228-239. Cambridge: Cambridge University Press. Cited in Brown (2004).

Maslovaty, N., Kuzi, E., (2002) “Promoting motivational goals through alternative or traditional assessment”, *Studies in Educational Evaluation*, Vol. 28, No. 3, pp. 199–222. Cited in Fox (2017).

Mendonça, C.O., Johnson, K.E., (1994) “Peer Review Negotiations: Revision Activities in ESL Writing Instruction”, *TESOL Quarterly*, Vol. 28, No. 4, pp. 745-769. Cited in Topping (1998). DOI: https://www.jstor.org/stable/3587558?seq=1#metadata_info_tab_contents (last visited: 17/05/2020).

Meyer, C.F. (2002) *English Corpus Linguistics: An introduction*. Cambridge: Cambridge University Press.

Novak, J., Gowin, D., Kahle, J., (1984) *Learning How to Learn*. Cambridge: Cambridge University Press.

Novello, A., (2012) “Motivare alla valutazione linguistica”, EL.LE, Vol. 1, No. 1, pp. 91-110. DOI: https://edizionicafoscari.unive.it/it/edizioni/riviste/elle/2012/1/motivare-alla-valutazione-linguistica/ (last visited: 16/04/2020).

Novello, A., “Riflessioni metodologiche: Pianificare e progettare la valutazione formativa”, *ANILS*, 2012, 8-9, pp. 38- 41. DOI: https://www.academia.edu/7608088/Progettare_e_pianificare_la_valutazione_forformat (last visited: 14/04/2020).

Novello, A., (2014) *La valutazione delle lingue straniere e seconde nella scuola. Dalla teoria alla pratica*, Edizioni Ca’ Foscari.

Oxford, R.L., (1990) *Language Learning Strategies: What Every Teacher Should Know*. Heinle & Heinle Publishers. Cited in Griffiths (2008).

Patri, M., (2002) “The influence of peer feedback on self- and peer-assessment of oral skills”, *Language Testing*, Vol. 19, No. 2, pp. 109–131. Cited in Peng (2010).

Pellegrino, J.W., Chudowsky, N., Glaser, R. (2001) *Knowing what students know: the science and design of educational assessment*. Washington: National Academy Press. Cited in Double et al. (2019).

Peñaflorida, A.H., (2002) "Non-traditional Forms of Assessment and Response to Student Writing: A Step Toward Learner Autonomy". In Richards, J.C., Renandya, W.A. (eds.) *Methodology in Language Teaching. An Anthology of Current Practice*, pp. 338-343. Cambridge: Cambridge University Press.

Peng, J., (2010) "Peer Assessment in an EFL Context: Attitudes and Correlations". In Prior, M.T., Watanabe, Y., and Lee, S.K. (eds.), *Selected Proceedings of the 2008 Second Language Research Forum: Exploring SLA Perspectives, Positions, and Practices,* pp. 89-107. Cascadilla Proceedings Project, Somerville, MA, USA. DOI: https://www.semanticscholar.org/paper/Peer-Assessment-in-an-EFL-Context%3A-Attitudes-and-Peng/4c6db58154d91825596a9fc36c6f3c4f30dca5e6 (last visited: 21/05/2020).

Richards, J.C., Schmidt, R., (2002) *Dictionary of language teaching and applied linguistics*, 3rd edition, London: Longman. DOI: https://www.academia.edu/37077225/Dictionary_of_Language_Teaching_and_Applied_Linguistics (10/04/2020).

Richards, J.C., Renandya, W.A. (2002) *Methodology in Language Teaching. An Anthology of Current Practice*, Cambridge: Cambridge University Press.

Römer, U. (2004) "Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching". In Aston, G., Bernardini, S., Stewart, D. (eds) *Corpora and Language Learner*, pp. 151-168. Amsterdam: John Benjamins.

Shohamy, E., Or, I.G., May, S*., Encyclopedia of Language and Education*, 3rd edition, Vol. 7 (*Language Testing and Assessment*). New York: Springer International Publishing.

Slavin, R.E. (1990). *Co-operative learning: Theory, research and practice*. Englewood Cliffs, NJ: Prentice Hall. Cited in Topping (2005).

Slavin, R., (2003) *Educational psychology: Theory and practice*. Boston: Allyn and Bacon. Cited in Brown (2000).

Spiller, D., (2012) “Assessment matters: Self-assessment and Peer assessment”, *Teaching Development*, pp. 1-19. Hamilton, New Zealand: University of Waikato. DOI: https://www.academia.edu/9882320/Assessment_Matters_SelfAssessment_and_Peer_Assessment_Teaching_Development_W%C4%81hanga_Whakapakari_Ako_Assessment_Matters_Self_Assessment_and_Peer_Assessment_Introduction (last visited: 14/05/2020).

TeachingEnglish | British Council. DOI: https://www.teachingenglish.org.uk/ (last visited: 26/04/2020).

Thomsen, H., (2003) “Scaffolding target language use”. In Little, D., Ridley. J., Ushioda, E. (eds.) *Learner autonomy in the foreign language classroom: teacher, learner, curriculum and assessment*, pp. 29-46. Dublin: Authentik Language Learning Resources Ltd.

Topping, K.J., (1998) “Peer assessment between students in colleges and universities”, *Review of Educational Research*, Vol. 68, No. 3, pp. 249–276. DOI: https://www.jstor.org/stable/1170598?readnow=1&seq=1#page_scan_tab_contents (last visited: 16/05/2020).

Topping, K.J., (2005) “Trends in Peer Learning”, *Educational Psychology*, Vol. 25, No. 6, pp. 631-645. DOI: https://www.re searchgate.net/publication/252332862_Trends_in_Peer_Learning (last visited: 16/05/2020).

Vygotsky, L.S., (1978) *Mind in society: The development of higher psychological processes*. Cole, M., John-Steiner, V., Scribner, S., Souberman, E. (eds.). Cambridge, MA: Harvard University Press.

Weiss, C.H., (1972) *Evaluation Research: Methods for Assessing Program Effectiveness*. Englewood Cliffs, NJ: Prentice-Hall. Cited in Dam and Legenhausen (2010).

White, E., (2009) “Student Perspectives of Peer Assessment for Learning in a Public Speaking Course”, *Asian EFL Journal*, Vol. 33, pp. 1-36. DOI: https://www.asian-efl-journal.com/pta_January_09.pdf (last visited: 28/05/2020).

Wood, D., Bruner, J.S., Ross, G., (1976) “The role of tutoring in problem solving”, *Journal of Child Psychology and Psychiatry*, Vol. 17, pp. 89-100, Pergamon Press. DOI: https://www.researchgate.net/publication/228039919_The_Role_of_Tutoring_in_Problem_Solving (last visited: 11/03/2020).

Appendix A: Questionnaire on peer assessment for second-year master students

My name is Chiara Colombo and I am a second-year master student of foreign languages (English and German). Some of you know me, but some of you don't, because I spent the winter semester 2019-2020 in Germany as an Erasmus student. The focus of my master thesis is on peer assessment, and for my Chapter Three I am carrying out a study on peer assessment with first-year students. Nonetheless, I would like to also include a section in this chapter where I discuss the position of more experienced students — that would be you — toward peer assessment. I know that during the course *Lingua, Linguistica E Traduzione Inglese 2* with Professor Ackerley you have worked on peer assessment. Therefore, I would like to record your observations, opinions and suggestions on your experience with peer assessment. This questionnaire is anonymous. Please, be as honest as possible.

Thank you very much for filling it in.

1) Age *
- o 18-23
- o 24-29
- o 30-...

2) Sex *
- o Female
- o Male

3) Are you interested in becoming a schoolteacher in the future? *
- o Yes
- o No
- o Not sure

4) If your answer to question 3 was "yes" or "maybe", did you find assessing your peers useful? Would you ever propose a similar activity to your future students? Please motivate your answer.

5) During the course with Professor Ackerley you were asked to assess your peers' presentations among other things. Had you ever done any kind of peer assessment before? If so, was it at University or in school? *

6) Were you happy about assessing your peers' work? *
- o 1 (Not happy at all)
- o 2
- o 3
- o 4
- o 5
- o 6
- o 7
- o 8

- 9
- 10 (Very happy)

7) Please motivate your answer. *

8) Were you happy about having your peers assess your work? *
- 1 (Not happy at all)
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 (Very happy)

9) Please motivate your answer. *

10) Were you happy with the results you got? Think about both the grade that you received (not the final grade, but the grade from your peers on your presentation) and the feedback (if you went to the office hours). Please, be as honest as possible. *

11) Please motivate your answer. Be as honest as you can. *

12) Did you feel prepared and adequate in your role as a "teacher" (i.e. when you were asked to prepare an evaluation grid and assess your peers' work)? *

13) Would you rather have already had an evaluation grid and simply filled it out? *

14) What do you think are the benefits of designing your own evaluation grid? Think, for instance, about the freedom you were given on which aspects of your peers' presentation to focus on (i.e. intonation, pronunciation, mechanics...). *

15) Given your recent experience as an assessor, do you think that peer assessment is objective? *
- Yes
- No
- Not sure

16) Please motivate your answer. *

17) Do you think that your judgement was biased when assessing your peer's presentation? Pleased be as honest as possible. *
- Yes
- No
- Not sure

18) Do you think that your peers' judgement was biased when they assessed your presentation? Please be as honest as possible. *

- Yes
- No
- Not sure

19) Please motivate your answers to question 13 and 14. Which do you think are the main causes for biases? (e.g. friendship, dislike...). You may also want to provide suggestions on how to avoid biases. *

20) General observations, suggestions, opinions... *

Please write your observations, suggestions or opinions on peer assessment. For example, you may want to write if you liked/did not like the activity, or if you found it useful.

Thank you for your time and help!
Chiara Colombo

Appendix B: Pre-activity questionnaire on peer assessment for first-year bachelor students

My name is Chiara Colombo and I am a second-year master student of foreign languages (English and German). The focus of my master thesis is on peer assessment, and for my Chapter Three I am carrying out a study on peer assessment with first-year students (you). Therefore, I would like to record your observations, opinions and suggestions on peer assessment.
This questionnaire is anonymous and will not be used other than in my dissertation. Please, be as honest as possible.

Thank you very much for filling it in.

1) Age *
- o 18-20
- o 21-29
- o 30-39
- o 40-…

2) Sex *
- o Female
- o Male

3) Level of written English * (Please make reference to any language certification you might have, or to the pre-course test at the beginning of the year: A2, B1, B1+, B2...).

4) Do you know what peer assessment is? *
- o Yes
- o No
- o Not sure

5) Have you ever been asked to assess (with or without grades) your peers' work? * (You should think about both providing feedback and providing grades to any peer's performance, written or spoken).
- o Yes
- o No
- o Not sure

6) If your answer to question 5 was 'yes', how did you feel about assessing your peers' work? Please motivate your answer. (e.g. happy, very prepared, excited, uncomfortable, scared, inadequate...).

7) If your answer to question 5 was 'no', how would you feel about assessing your peers' work? Please motivate your answer. (e.g. happy, very prepared, excited, uncomfortable, scared, inadequate...).

8) Peer assessment consists in students mutually assessing their peers' performances. Given this definition, do you think that peer assessment is as objective as traditional teacher assessment? *

- 1 (Not objective at all)
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 (Equally objective)

9) Please motivate your answer. *

10) Do you think that friendship can be a cause for biased assessment (= valutazione non oggettiva)? *

- Yes
- No
- Not sure

11) Do you think that anonymous peer assessment is a good solution to avoid biased assessment (= valutazione non oggettiva)? *

- Yes
- No
- Not sure

Thank you for your time and help!
Chiara Colombo

Appendix C: Post-activity questionnaire on peer assessment for first-year bachelor students

This questionnaire is anonymous.
I will record your answers, observations and suggestions and I will use them for my dissertation. Please be as honest as possible.

Thank you for taking part in this study.

1) Age *
- o 18-20
- o 21-29
- o 30-39
- o 40-…

2) Sex *
- o Female
- o Male

3) How did you feel about assessing your peer's work? *
- o 1 (Very uncomfortable)
- o 2
- o 3
- o 4
- o 5
- o 6
- o 7
- o 8
- o 9
- o 10 (Very comfortable)

4) How did you feel about having another student assess your work? *
- o 1 (Very uncomfortable)
- o 2
- o 3
- o 4
- o 5
- o 6
- o 7
- o 8
- o 9
- o 10 (Very comfortable)

5) Why? Please motivate your answer to question 4. *

6) Are you satisfied with how your peer assessed your work? * (Think about both the feedback and the grade that you received).
- o 1 (Not satisfied at all)

- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 (Very satisfied)

7) Why? Please motivate your answer. *

8) Would you rather have simply provided feedback to your peer's work or were you happy with also assigning a grade to his/her work? *
- Just feedback
- Just grade
- Feedback and grade

9) What was the hardest? Providing feedback or assigning a grade? *
- Feedback
- Grade

10) Why? Please motivate your answer to question 6 and 7. *

11) Did you find assessing your peer's work difficult? *
- 1 (Not difficult at all)
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 (Very difficult)

12) Do you think that having guidelines on how to assess your peer was helpful and made assessing your peer's performance easier? Please motivate your answer. (Think about both the useful phrases for feedback and the evaluation grid which you were provided.)

Thank you for your time and help!
Chiara Colombo

Appendix D: Questionnaire on peer assessment and peer feedback for university English teachers

My name is Chiara Colombo and I am a second-year master student of foreign languages (English and German).
The focus of my master thesis is on peer assessment, and for Chapter Three I am carrying out a study on peer assessment with first-year bachelor students of languages. As suggested by my supervisor, Professor Katherine Ackerley, I would like to also include a section in my dissertation where I discuss the position of teachers of English toward peer assessment.

This questionnaire is anonymous.
Thank you very much for your help with my project.

1) Which classes have you taught? *
- o 1st-year bachelor
- o 2nd-year bachelor
- o 3rd-year bachelor
- o 1st-year master
- o 2nd-year master

2) Have you ever carried out a peer assessment or peer feedback activity with your students? *
- o Yes
- o No
- o Not sure

3) Have you ever asked your students to give a grade to a peer's performance? *
- o Yes
- o No
- o Not sure

4) Which students do you think can provide more helpful feedback to their peers? * (Choose between very helpful, not so helpful, and not sure).
- o 1st-year bachelor
- o 2nd-year bachelor
- o 3rd-year bachelor
- o 1st-year master
- o 2nd-year master

5) Do you think that students' language proficiency affects their ability to provide helpful feedback to their peers? *
- o Yes
- o No
- o Not sure

6) Which level of students would you consider as more suited for peer assessment? *
- o A1-A2

- B1-B2
- C1-C2

7) In your experience, how successful have the peer assessment activities been? *
- 1 (Not successful at all)
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 (Very successful)

8) If you have ever done peer assessment with your students, what kind of activities have you asked them to do?

9) Further comments about peer assessment? *

10) Are you willing to further discuss this topic with me in a Zoom meeting? *
(My e-mail address is: chiara.colombo.3@studenti.unipd.it).
- Yes
- No

Thank you very much for your time and help.
Chiara Colombo

Riassunto

La valutazione tra pari (*peer assessment*)

La mia tesi tratterà principalmente della valutazione tra pari (in inglese *peer assessment*), un particolare tipo di valutazione che coinvolge in prima persona lo studente a cui viene chiesto di valutare l'esecuzione di un compito di uno o più compagni di classe. La valutazione tra pari può essere eseguita a coppie ma anche in piccoli gruppi, come spesso avviene nelle classi molto numerose. Prima di addentrarmi nella spiegazione della valutazione tra pari, è bene che io chiarifichi alcuni concetti, primo tra tutti quello di 'valutazione'.

Definizioni preliminari

Con il termine valutazione si intende l'interpretazione, generalmente da parte dell'insegnante, dei dati che vengono raccolti con la somministrazione di verifiche. Secondo Novello (2014, p. 9), la valutazione scolastica permette di aprire tra insegnante e discente un canale di comunicazione "attraverso il quale vengono scambiate informazioni basilari per un buon procedere del processo di insegnamento e apprendimento". L'insegnante, infatti, tramite la valutazione, ha sia l'opportunità di osservare e monitorare l'evolversi del processo di apprendimento degli studenti, sia di rilevare l'efficacia o meno del metodo di insegnamento e/o di verifica utilizzato e, quindi, se necessario, apporre dei cambiamenti *in itinere*. Lo studente, dal canto suo, può "comunicare i propri progressi, le proprie difficoltà, i bisogni" (Novello 2014, p. 9).

È bene ora fare un'importante distinzione tra valutazione formativa e valutazione sommativa.

Con valutazione formativa si intende quel tipo di valutazione prettamente di natura qualitativa che permette all'insegnante di ottenere un *feedback* sul processo di insegnamento e apprendimento. La valutazione formativa, infatti, avviene durante il corso e accompagna il processo di insegnamento e di apprendimento. Il *feedback* è generalmente un commento dettagliato, che viene fornito all'allievo sulla sua prestazione. Può consistere, per esempio, in suggerimenti, critiche costruttive e precisazioni sugli errori più importanti che l'insegnante ha individuato.

Al contrario, la valutazione sommativa vuole misurare quantitativamente quanto uno studente ha appreso e, pertanto, avviene solitamente alla fine di un corso o di un'unità didattica.

Faccio ora un'ulteriore precisazione. Molto spesso, i termini 'valutazione' e 'verifica' vengono usati come sinonimi. Tuttavia, vi è tra di essi una differenza di significato, seppure lieve. Brown (2004) definisce la verifica come un procedimento formale volto, appunto, a verificare le conoscenze e le competenze degli studenti (cosa uno studente sa o non sa e cosa uno studente sa o non sa fare). Inoltre, Novello (2014) ritiene che la verifica permetta di comparare i risultati della prestazione degli studenti con gli obiettivi del corso e, in particolare, di verificare se tali obiettivi sono stati raggiunti, in quale misura, e da chi. La valutazione, invece, come già spiegato, è un processo che comprende ben più di questo: permette di formare dei giudizi sui discenti, basandosi sull'interpretazione dei dati raccolti nelle verifiche e permette agli insegnanti di 'raddrizzare il tiro' e prendere delle decisioni riguardo al metodo di insegnamento e di verifica adottati fino a quel momento.

Il processo valutativo

Molti studiosi ed esperti concordano sulla necessità che la valutazione sia motivante per gli studenti. Imparare una lingua straniera richiede, infatti, molti sforzi e sacrifici e la motivazione gioca un ruolo essenziale nel processo di apprendimento. Tuttavia, un'importante parte dell'apprendimento riguarda la valutazione, da cui il discente può ricavare importanti informazioni per migliorarsi. Pertanto, anche motivare lo studente alla valutazione si rivela essenziale. Se, infatti, gli studenti percepiscono che il loro apprendimento potrà trarre dei benefici dalla valutazione, saranno più motivati a sottoporsi periodicamente a verifiche e a ricevere *feedback* frequenti. La valutazione, quindi, non deve essere vista dallo studente come un giudizio di cui è oggetto passivo, bensì come un'occasione per migliorarsi e per raggiungere i propri obiettivi.

Inoltre, gli esperti dell'insegnamento delle lingue straniere consigliano anche di avere sempre dei criteri di valutazione ben definiti e precisi e, soprattutto, di condividere tali criteri con gli studenti prima della valutazione. Tale trasparenza, infatti, aiuta questi ultimi ad accrescere il loro senso di fiducia nei confronti del processo valutativo.

Per ottenere valutazioni il più oggettive e specifiche possibili vengono spesso utilizzate delle scale valutative (*rating scales*). Queste consistono in descrizioni

(*descriptors*) più o meno approfondite delle diverse competenze auspicabili nei diversi livelli di apprendimento e servono a guidare la valutazione dell'insegnante (ma, come vedremo più avanti, anche degli studenti). Tali scale possono essere olistiche (*holistic rating scales*), quindi consistere in un unico voto finale dato dall'impressione globale di una data *performance*, o analitiche (*analytic rating scales*), quindi consistere in diversi voti, poi sommati, sui diversi aspetti della *performance* che sono stati presi in considerazione. Per esempio, quando si vuole valutare un saggio, si possono considerate diversi aspetti di quest'ultimo, come il contenuto, la correttezza grammaticale, la pertinenza del lessico utilizzato, mentre nel caso di una presentazione orale possono essere considerati altri aspetti, come l'intonazione, il linguaggio non verbale, e le abilità riassuntive dello studente.

La valutazione alternativa: definizione e tipologie

Tornando al tema della mia tesi - la valutazione tra pari - è da ricordare che essa fa parte di quel tipo di valutazione detta 'alternativa' o 'non tradizionale' (*alternative assessment*, *non-traditional assessment*). Si tratta di un tipo di valutazione piuttosto recente che racchiude diverse pratiche, il cui nome indica una contrapposizione con la valutazione tradizionale, il *teacher assessment*, dove è l'insegnante l'unico responsabile della valutazione dei suoi allievi. Con il termine valutazione alternativa, si intende, infatti, quel processo valutativo che si avvale di pratiche non-tradizionali per valutare l'intero processo di apprendimento dei discenti (e non più solo il prodotto del loro apprendimento).

Numerosi sono i benefici che vengono riconosciuti alla valutazione alternativa, primo tra tutti il permettere allo studente di essere il vero protagonista del suo apprendimento e di venire indirizzato verso l'autonomia (*learner autonomy*) grazie a pratiche che puntano alla riflessione e all'auto-responsabilità degli studenti nei confronti del proprio percorso. Infatti, come si vedrà in seguito, la valutazione alternativa incoraggia gli studenti a porsi degli obiettivi e a raggiungerli ciascuno con il proprio passo. Tra queste pratiche di valutazione alternative troviamo, per citarne alcune, i diari dello studente (*learner diaries/journals*), i dossier (*portfolios*), l'autovalutazione (*self-assessment*) e la valutazione tra pari (*peer assessment*). In breve, i *learner diaries* consistono in voci che vengono costantemente aggiornate dagli studenti, nei quali essi possono registrare i loro pensieri, idee o progressi nei confronti di quegli obiettivi di

apprendimento che si sono prefissati. La scrittura di diari può essere un potente strumento pedagogico, in quanto può essere uno spunto per l'autoriflessione con cui gli studenti si prefissano obiettivi e monitorano il loro progresso e il raggiungimento di questi ultimi. I *portfolios*, invece, consistono in raccolte di stralci di produzione scritta (ma anche orale, come nel caso dei dossier elettronici o *e-portfolios*) che, teoricamente, dovrebbero rappresentare il progresso dello studente nella scrittura. Nel campo dei *portfolios*, un importantissimo strumento nell'insegnamento delle lingue straniere è il *Council of Europe's European Language Portfolio* (*ELP*). Questo strumento è costituito da tre componenti fondamentali: un 'passaporto linguistico', una 'biografia linguistica' e un dossier, e può essere usato dagli studenti per valutare la propria competenza linguistica e ha, pertanto, un riconosciuto valore per l'auto-valutazione. Il *self-assessment*, invece, come suggerisce il nome stesso, consiste in una valutazione che gli studenti fanno delle loro stesse prestazioni, di qualsiasi natura esse siano. Si tratta, pertanto, di una valutazione 'auto-amministrata'.

Tra i benefici più evidenti dell'autovalutazione vi è, per esempio, l'autonomia. Gli studenti autonomi (*autonomous learners*) "sono in grado di operare delle scelte su vari aspetti del loro apprendimento. Queste scelte possono coinvolgere, per esempio, la pianificazione di obiettivi di apprendimento e l'adozione di strategie per raggiungere tali obiettivi, il monitoraggio dell'apprendimento personale e persino la valutazione obiettiva delle proprie capacità" (Little 1991, p. 1). Come osserva un altro studioso, Gardner (1999, p. 51), "gli studenti autonomi decidono cosa imparare, come imparare e quando imparare. Si assumono la responsabilità del proprio apprendimento e questo riguarda anche il monitoraggio dell'apprendimento" tramite, appunto, l'autovalutazione.

Nonostante i vantaggi e i benefici delle pratiche di valutazione alternative, esse sono anche oggetto di numerose critiche, che riguardano principalmente la loro oggettività (*objectivity*), affidabilità (*reliability*) e validità (*validity*). L'oggettività di una valutazione descrive il suo grado di imparzialità e di indipendenza da influenze esterne. L'affidabilità, invece, si riferisce alla coerenza in un test, ovvero alla costanza di risultati attraverso somministrazioni successive. Infine, la validità di un test o di una valutazione considera quanto un test misuri effettivamente ciò che si propone di misurare (es. un test di grammatica deve misurare solo la grammatica, un test di lessico solo il lessico e non anche la grammatica). Come si può facilmente comprendere da queste definizioni, poiché

il *peer assessment* e il *self-assessment* vengono eseguiti dagli studenti stessi, molto spesso si ritiene che essi non siano in grado di valutarsi o di valutare altri studenti in modo oggettivo, affidabile e valido a causa, principalmente, della loro ovvia inesperienza nel campo valutativo. Ciononostante, come puntualizza White (2009, p. 24), "ogni tipo di valutazione può essere soggetta a giudizi errati, indipendentemente dalla natura della valutazione". Inoltre, come osserva Peng (2010, p. 89), "raggiungere un'alta validità e affidabilità non è l'obiettivo ultimo della valutazione tra pari" (né di tutte le altre tipologie di valutazione alternativa).

La valutazione tra pari: benefici e limiti

Vediamo ora più nel dettaglio la valutazione tra pari, i suoi benefici, ma anche i suoi aspetti negativi.

Il *peer assessment* è un tipo di valutazione piuttosto recente (generalmente si fa risalire agli anni '80) e, pertanto, ancora in sviluppo. Come accennato all'inizio, con valutazione tra pari si intende, generalmente, quella valutazione che avviene tra due o più studenti. Bisogna qui fare un'importante precisazione: 'peers' non significa studenti coetanei, quanto piuttosto studenti che abbiano un livello di apprendimento uguale o comunque molto simile (Topping 1998). Anche Jung (2016, p. 10) sostiene che il *peer assessment* debba avvenire solo "tra discenti con capacità e progressione linguistiche equivalenti" per garantire equilibrio ed efficacia al procedimento valutativo. In caso contrario, infatti, gli studenti potrebbero sentirsi giudicati e, di conseguenza, manifestare sentimenti negativi verso questa pratica valutativa. Pertanto, è più probabile che uno *status* di apprendimento uguale o molto simile porti a risultati migliori.

Per riprendere un concetto già menzionato, quello di valutazione formativa, il *peer assessment* sembra più adatto ad assolvere a scopi formativi piuttosto che sommativi, poiché, per esempio, può aiutare ad identificare per tempo gli errori, evitando che questi si fossilizzino nella mente dello studente. Inoltre, può spingere gli studenti a sviluppare delle strategie per colmare le loro lacune.

La valutazione tra pari ha numerosi vantaggi. Hansen Edwards (2013) li riassume in una tabella, prendendo in considerazione gli aspetti che possono trarre particolare beneficio da questo tipo di valutazione. Tra questi, per esempio, Hansen Edwards menziona le competenze cognitive e metacognitive, in quanto lo studente si soffermerà più a lungo sul compito per revisionarlo, avrà una maggiore comprensione del processo

valutativo e lo 'farà proprio', nonché svilupperà la propria autonomia di apprendimento; il *feedback*, in quanto verrà ricevuto in quantità maggiore e più velocemente; le competenze sociali, in quanto lo studente dovrà collaborare e 'negoziare' coi compagni; lo sviluppo linguistico, poiché lo studente dovrà sviluppare le sue competenze comunicative per comunicare ai compagni le sue osservazioni, nonché avrà una maggiore opportunità di utilizzare la lingua straniera (il *feedback*, infatti, va trasmesso al compagno nella lingua che si sta studiando).

Tuttavia, come già affermato in precedenza, è possibile individuare anche molti aspetti negativi, la maggior parte dei quali sono gli stessi svantaggi di tutte le altre procedure di valutazione alternative che sono state individuate prima, vale a dire la validità, l'affidabilità e l'oggettività. Oltre a questi, tuttavia, la valutazione tra pari pone diversi altri problemi, soprattutto a causa della sua natura reciproca e collaborativa. Infatti, mentre altre pratiche alternative si concentrano principalmente sul singolo studente (come nel caso dell'autovalutazione, dei diari e dei dossier), la valutazione tra pari si concentra in particolare sull'interazione tra studenti, oltre a concentrarsi sulle prestazioni del singolo. In effetti, nella valutazione tra pari, gli studenti sono tenuti a valutare il lavoro dei loro pari e a condividere le loro osservazioni e, spesso, questo fatto solleva alcuni problemi, come il possibile "imbarazzo sociale" (White 2009, p. 8), in particolare per quanto riguarda l'identificazione delle debolezze nel lavoro dei coetanei, nonché la potenziale faziosità della valutazione tra pari (*biased evaluation*), soprattutto quando si tratta di giudicare la prestazione di un amico (o di un nemico). Inoltre, un altro problema che viene spesso sollevato a sfavore della valutazione tra pari riguarda l'inesperienza e l'inadeguatezza degli studenti nel valutare i propri compagni. In effetti, come possono degli studenti che sono ancora in fase di apprendimento essere in grado di apportare delle valutazioni accurate e complete delle prestazioni dei loro coetanei?

A questo proposito, vari esperti di didattica e di valutazione hanno proposto alcune soluzioni per ovviare agli svantaggi di questo tipo di valutazione.

Per esempio, Brown (2004) suggerisce agli insegnanti di fornire agli studenti delle linee guida ben precise per valutare i compagni.

Per quanto riguarda le obiezioni sull'affidabilità, sulla validità e sull'oggettività del *peer assessment*, Peng (2010) osserva che le funzioni e i ruoli della valutazione tra pari e della valutazione degli insegnanti sono molto diversi, e che il primo dovrebbe

funzionare semplicemente come uno strumento di supporto al secondo, in modo da tranquillizzare gli studenti e 'liberarli' di una responsabilità così grossa. Infatti, Peng (2010, p. 90) osserva che la valutazione tra pari è un "metodo di valutazione supplementare" che serve a coinvolgere e responsabilizzare gli studenti piuttosto che una vera e propria sostituzione della tradizionale valutazione degli insegnanti. Questa visione è condivisa anche da Topping (1998, p. 262), il quale sostiene che la valutazione tra pari dovrebbe essere vista come un valore aggiunto al processo di apprendimento. Infatti, aggiunge, incorporare il *peer assessment* al metodo di valutazione tradizionale può portare a risultati buoni tanto quanto la semplice valutazione degli insegnanti.

Invece, per superare i comprensibili problemi legati all'inesperienza degli studenti nel campo della valutazione, Keller e Westfall-Greiter (2014, p. 9) sostengono che è molto importante fornire agli studenti frasi utili (*prompts*) per scambiarsi un *feedback* efficace, nonché per svolgere delle valutazioni più approfondite e dettagliate sulla prestazione dei loro pari. Inoltre, Hattie (2012) sottolinea che tali suggerimenti possono essere utilizzati dagli studenti anche per monitorare e riflettere sul proprio apprendimento e non solo su quello dei loro coetanei e che, pertanto, possono essere dei potenti strumenti per favorire l'autonomia dello studente di cui si è già parlato sopra.

Come accennato in precedenza, uno dei principali svantaggi delle forme alternative di valutazione è generalmente ritenuta l'oggettività. Infatti, non solo i docenti, ma anche gli studenti stessi sono spesso scettici riguardo al *peer assessment* poiché ritengono che le valutazioni fatte dagli insegnanti siano più accurate rispetto a quelle fatte da loro. Tuttavia, come già sottolineato, tutte le valutazioni possono essere soggette a giudizi errati, indipendentemente dalla loro natura (*teacher assessment*, *peer assessment*, o *self-assessment*). In breve, molti esperti (come Hattie 2012, Spiller 2012, Keller e Westfall-Greiter 2014), notano come gli studenti abbiano semplicemente bisogno di una formazione adeguata nella valutazione tra pari per raggiungere l'imparzialità, formazione che può essere sviluppata grazie alla pratica, ma anche grazie a scale di valutazione precise che li guidino durante le loro valutazioni.

Infine, per quanto riguarda le preoccupazioni circa la potenziale faziosità nella valutazione tra pari, molti studi (Donato 1994; Topping 1998; White 2009; Peng 2010; Gan 2011; Spiller 2012; Jung 2016) hanno dimostrato che le valutazioni fatte dall'insegnante e quelle fatte dagli studenti sono spesso molto simili. Quindi, pare che gli

studenti siano in grado "di fornire supporto ai loro coetanei durante le interazioni collaborative usando la L2 in modo analogo a quello degli esperti" (Donato 1994, p. 51), dimostrando, così, che essi sono spesso più adatti a questo tipo di attività di quanto pensino.

La letteratura sul *peer assessment*

Negli ultimi anni sono stati fatti molti studi sul *peer assessment*. La maggior parte di essi, tuttavia, si concentra su due questioni in particolare: le percezioni che gli studenti hanno sul *peer assessment* e quanto problematiche siano, effettivamente, l'affidabilità, l'oggettività e la validità di queste valutazioni in contesti scolastici reali.

Nel mio secondo capitolo ho esaminato alcuni di questi studi, che ho presentato in ordine cronologico. Per ciascuno di essi ho spiegato perché e in che misura sia stato utilizzato nella parte sperimentale della mia tesi sul *peer assessment*. In particolare, ho preso in considerazione gli studi di Mendonça e Johnson (1994), Azarnoosh (2013), e Jung (2016) per quanto riguarda le abilità scritte (*writing skills*) che costituiscono anche l'oggetto su cui si è focalizzata la mia ricerca, mentre ho utilizzato quelli di White (2009) e Peng (2010), che riguardavano le abilità orali, per la costruzione dei questionari. Inoltre, mi sono servita degli studi di DeNisi et al. (1983) e di Mendonça e Johnson (1994) per l'analisi del *feedback* scambiato tra gli studenti (*peer feedback*), e degli studi di Falchinov e Goldfinch (2000), White (2009), Peng (2010), Azarnoosh (2013) e Jung (2016) per l'analisi della valutazione tra pari.

Ricerca sperimentale sulla valutazione tra pari: coinvolgimento di studenti e docenti

Nella parte sperimentale della mia tesi mi sono soffermata ad analizzare le percezioni della valutazione tra pari in due gruppi di studenti: un gruppo della laurea magistrale in lingue e letterature europee e americane (LMLLA) presso l'Università degli Studi di Padova e l'altro del primo anno di laurea triennale in lingue, letterature e mediazione culturale (LTLLM) sempre presso l'Università degli Studi di Padova. Successivamente ho preso in esame le percezioni dei docenti di inglese (sia dei moduli frontali, che di lettorato) dei corsi di lingue sia triennali che magistrali.

Studenti di laurea magistrale

Con il primo gruppo di studenti, l'attività di *peer assessment* è stata messa in pratica durante il corso di *Lingua, linguistica e traduzione inglese 2* e, in particolare, durante il modulo frontale di *Spoken English*, tenuto dalla professoressa Ackerley, nonché mia relatrice, durante il semestre invernale 2019-2020. Ho fornito un questionario con venti brevi quesiti, a cui hanno risposto dodici studenti su ventiquattro frequentanti totali.

Essi, durante il corso, dovevano preparare, individualmente o a coppie, una presentazione di 10 minuti su uno degli argomenti trattati durante le lezioni. Essi venivano, poi, valutati sia dalla professoressa che da un compagno (*peer assessment*) utilizzando una scala di valutazione, sia analitica che olistica basata sul Quadro comune europeo di riferimento per le lingue (QCER), da essi progettata: potevano decidere liberamente i parametri, i criteri di valutazione e gli aspetti delle prestazioni dei loro coetanei. Tutte le rubriche sono state pubblicate in *Moodle* alcuni giorni prima dell'inizio delle presentazioni, in modo tale che gli studenti esaminati sapessero in anticipo sia chi li avrebbe valutati sia con quali criteri.

Come dimostrano i risultati del questionario, tutti gli studenti hanno espresso delle opinioni positive sulla valutazione tra pari e l'hanno trovata utile per migliorare le proprie competenze orali (*oral skills*) e per avere una migliore comprensione di ciò che la valutazione in classe comporta. Solo un paio di studenti sembravano imbarazzati a dover effettuare delle valutazioni sulle prestazioni dei loro coetanei, sia per il fatto di dover valutare degli amici sia di dover valutare studenti notoriamente 'bravi', cioè studenti che di solito ottengono buoni voti. Quest'ultimo aspetto è stato menzionato da alcuni studenti anche come una possibile causa di valutazione non oggettiva (*biased evaluation*). In effetti, valutare degli studenti che di solito ottengono voti alti potrebbe essere difficile per i loro coetanei che si possono percepire inadeguati in questo ruolo.

Nel complesso, questo particolare gruppo di studenti si è dimostrato consapevole delle difficoltà che la valutazione, e in particolare la valutazione tra pari, comporta ed è stato in grado di fornire alcuni suggerimenti su come raggiungere l'oggettività in questo tipo di valutazione. Alcuni, ad esempio, hanno notato che le opinioni o i pregiudizi che avevano precedentemente sui loro coetanei avevano influenzato inizialmente le loro valutazioni; tuttavia, hanno anche affermato di aver cercato di fare del loro meglio per evitare che la loro soggettività impedisse di fornire valutazioni oggettive.

La mia ricerca ha diversi limiti, primo fra tutti il numero esiguo di partecipanti al questionario rispetto al numero totale dei partecipanti al corso che non può quindi essere considerato rappresentativo dell'intera classe e, di conseguenza, rende difficile il trarre delle conclusioni. In particolare, poi, le domande relative all'età, al sesso e alla professione auspicata non hanno mostrato differenze significative nelle risposte. Infine, le risposte degli studenti sono state raccolte solo attraverso un questionario e diversi mesi dopo la fine del corso.

Studenti di laurea triennale

Per quanto riguarda il progetto con gli studenti del primo anno di laurea triennale, solo sei studenti hanno partecipato. Ho formato tre coppie equilibrate: la prima composta da due studenti di livello B1, la seconda da due studenti con B2 e C1 e la terza composta da due studenti di livello B1+. Il mio lavoro con gli studenti del primo anno è stato più completo rispetto a quello con gli studenti della magistrale: ho preparato due questionari anonimi (un questionario a cui sottoporli prima dell'attività, per capire la loro esperienza sulla valutazione tra pari e la loro percezione su questo tipo di valutazione e un questionario di fine attività per capire se la loro percezione sul *peer assessment* fosse cambiata e se avessero tratto dei benefici dall'attività in vista dell'esame), un compito scritto (*writing task*), ovvero un *listening report*, una griglia di valutazione analitica da usare per il *peer assessment* e una lista di frasi utili per il *feedback* (tratte da Keller e Westfall-Greiter, 2014). Il compito consisteva nell'ascoltare un'intervista con la regista Angelina Jolie sul suo film 'Unbroken' e nello scrivere un *report* di circa 300 parole riassumendone il contenuto, annotando eventuali passaggi difficili e indicando come si era proceduto per superarli. Per quanto riguarda la griglia di valutazione, non è stato richiesto agli studenti di esprimere un voto numerico, ma soltanto un giudizio (*poor*, *mediocre*, *good*, *outstanding*) basato su cinque indicatori: *spelling* e punteggiatura, correttezza grammaticale, lessico, coesione e realizzazione complessiva del compito.

L'anonimità è stata garantita agli studenti durante tutto il progetto, sia nei questionari anonimi, che nell'attività di *peer assessment*, in quanto fungevo da intermediaria: gli studenti si sono limitati a completare il compito e ad inviarmelo via e-mail, io poi ho provveduto a scambiare i *report* perché venissero reciprocamente corretti. Inoltre, ho fornito le mie correzioni e il mio *feedback* a tutti gli studenti in modo che, alla

fine del progetto, ricevessero due correzioni e due *feedback* diversi (uno del compagno e uno mio).

Tutti gli studenti hanno apprezzato l'attività, sentendosi a proprio agio sia a valutare il compito del compagno, sia a far valutare il proprio compito. Inoltre, tutti hanno riconosciuto l'utilità di disporre di linee guida precise per il *feedback* e di una scala di valutazione che li guidasse nella scelta del giudizio finale. Nel complesso, tutti gli studenti sono stati in grado di fornire al proprio compagno dei consigli utili e precisi su come migliorare le proprie abilità scritte in preparazione dell'esame finale di giugno, prendendo dai testi esempi concreti di frasi sbagliate e correggendole.

I giudizi degli studenti mostrano un accordo con i giudizi da me assegnati: in particolare, riguardo al *feedback* non ci sono differenze significative, nonostante il mio commento fosse più dettagliato.

Riguardo alla valutazione numerica (anche se non richiesta), invece, quella degli studenti della prima coppia è stata leggermente più alta rispetto alla mia, mentre una studentessa della terza coppia ha assegnato un voto leggermente inferiore al mio alla sua compagna.

Purtroppo, il numero ridotto di studenti che ha aderito al progetto rende i risultati statisticamente poco significativi.

Coinvolgimento dei docenti

Infine, presento i risultati del questionario sulla valutazione tra pari e sul *feedback* tra pari, distribuito a quattordici professori di inglese del Dipartimento di Studi Linguistici e Letterari (DISLL) presso l'Università dei Padova. Di fatto, nove professori hanno risposto al questionario. Di questi, la maggior parte ha insegnato sia a studenti della triennale che a studenti della magistrale e tutti hanno fatto delle attività di valutazione tra pari con i propri studenti.

Nella prima domanda chiedevo di indicare quali studenti, per fascia di livello, ritenessero più capaci di dare un *feedback* utile ai propri compagni. La maggior parte ha risposto che gli studenti più adatti fossero quelli di livello B1-B2 e C1-C2; soltanto un docente ha incluso gli studenti di livello A1-A2. Nella seconda domanda veniva chiesto se la competenza linguistica potesse influenzare la capacità degli studenti di dare un *feedback* preciso e utile ai compagni: in questo caso, la maggioranza (6) ha risposto affermativamente, mentre solo un docente ha dato una risposta negativa.

Nell'ultima domanda ho chiesto di scrivere commenti o osservazioni sul *peer assessment*. Le risposte si sono rivelate molto utili e significative per la mia tesi. In particolare, diversi docenti hanno notato che gli studenti hanno bisogno di istruzioni e linee guida molto chiare su come operare nella valutazione tra pari, confermando quanto afferma anche la letteratura scientifica sull'argomento.

Un docente ha evidenziato il fatto che gli studenti che formano le coppie o i mini-gruppi per il *peer assessment* dovrebbero avere un livello di competenza linguistica simile - osservazione anch'essa presente molto spesso nella letteratura sulla valutazione tra pari.

Un altro docente ha sottolineato l'importanza di spiegare agli studenti in che cosa consista la critica costruttiva, perché spesso essi tendono a 'farsi complimenti' piuttosto che aiutarsi a migliorare.

Infine, alcuni professori hanno spiegato i motivi per cui svolgono delle attività di valutazione tra pari con i loro studenti: soprattutto per incoraggiare il pensiero critico, aumentare la consapevolezza degli studenti e promuovere la loro autonomia.

Come menzionato nel primo capitolo, uno dei molti vantaggi dell'utilizzo di forme alternative di valutazione è che queste possono migliorare l'autonomia del discente e la valutazione tra pari sembra essere un metodo eccezionale per questo scopo, dato che gli studenti, giudicando il lavoro di un compagno, possono diventare più consapevoli delle proprie capacità e migliorare così le proprie prestazioni.

Conclusioni

Con la mia ricerca, ho voluto studiare le diverse percezioni della valutazione tra pari da parte di tre diversi gruppi: un gruppo di studenti di magistrale, uno di studenti di triennale e un gruppo di docenti di inglese. Più specificamente, ho voluto soffermarmi sulle differenze più evidenti tra studenti della magistrale – soprattutto gli aspiranti insegnanti – e della triennale. Infine, per quanto riguarda gli insegnanti, ho voluto confrontare i loro commenti sulla valutazione tra pari con ciò che dice la letteratura sul *peer assessment*.

Confrontare i risultati dei miei studi con gli studenti di magistrale e di triennale risulta difficile per una serie di motivi. Innanzitutto, l'attività di valutazione tra pari svolta dal primo gruppo era molto diversa dall'attività svolta dal secondo gruppo. In effetti, al primo gruppo di studenti è stato chiesto di progettare una scala di valutazione per le abilità orali e di valutare delle presentazioni orali, mentre al secondo ho fornito io stessa una

griglia progettata appositamente per valutare le abilità scritte. Inoltre, il mio studio con il primo gruppo è basato su un questionario, mentre il mio progetto con gli studenti di triennale è più completo (consiste in due questionari, un compito di *listening report* e nella valutazione tra pari di tale compito). Infine, anche il numero di partecipanti ai due studi è molto diverso.

Detto questo, in questi due gruppi può essere comunque osservata, verso la valutazione tra pari, una tendenza simile.

In primo luogo, entrambi hanno affermato che la valutazione tra pari li ha aiutati in diversi modi. La cosa principale che ho notato è che alcuni aspiranti insegnanti hanno affermato che essa li ha aiutati a capire i principali problemi coinvolti nella valutazione e che avere questo ruolo da insegnante è stata una notevole opportunità in vista della loro futura professione. Invece, gli studenti del primo anno hanno dichiarato che la valutazione tra pari li ha aiutati ad identificare errori che, altrimenti, avrebbero tralasciato e ha dato loro l'opportunità di confrontare il loro lavoro con quello di un pari di livello simile.

In secondo luogo, gli studenti del secondo gruppo hanno concordano sul fatto che l'anonimato nella valutazione tra pari possa garantire delle valutazioni eque ed obiettive. Ciò è stato suggerito anche da uno degli studenti di magistrale.

In terzo luogo, in entrambi i gruppi alcuni studenti non si sono sentiti all'altezza del compito, suggerendo così che la valutazione tra pari necessita di formazione, pratica e istruzioni chiare. Come notato anche da alcuni degli insegnanti intervistati, gli studenti devono essere istruiti anche su cosa siano le critiche costruttive, perché possano fornire un *feedback* utile.

Dati i risultati positivi dei miei studi con gli studenti e i commenti positivi degli insegnanti sulla valutazione tra pari, sembra che essa sia un ottimo modo per coinvolgere gli studenti nel processo di valutazione, spronarli a riflettere sul proprio compito osservando il compito di un compagno e a farsi carico del proprio apprendimento. Di conseguenza, la valutazione tra pari ha anche delle ripercussioni positive sull'autovalutazione, promuovendo così l'autonomia degli studenti. Ultimo ma non meno importante, la valutazione tra pari può anche garantire un *feedback* più immediato e diversificato rispetto alla valutazione degli insegnanti.

Date queste premesse, la valutazione tra pari dovrebbe essere intrapresa anche nelle scuole, in modo che gli studenti possano impegnarsi maggiormente nel proprio apprendimento già in giovane età.