Exploration of peer-assessment practices in a Diploma Programme English class in an international school in Spain and their effects on classroom talk.

Mayo 2023

Exploración de las prácticas de coevaluación en una clase de inglés en el Programa de Diploma en un colegio internacional en España y sus efectos en la comunicación en el aula.

Exploration of peer-assessment practices in a Diploma Programme English class in an international school in Spain and their effects on classroom talk.

Autora: María Dolores Salmerón Carmona

Directoras: Dra. María Enriqueta Cortés de los Ríos
Dra. Sagrario Salaberri Ramiro

PROGRAMA DE DOCTORADO EN EDUCACIÓN

Línea de investigación: Didáctica e Innovación para una Sociedad Inclusiva

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors Dr. Maria Sagrario Salaberri Ramiro and Dr. Maria Enriqueta Cortés de los Ríos. Their expert guidance was key to the completion of this thesis. I would also like to express my gratitude to Dr. Jane Jones, who encouraged me to write this PhD and continuously pushed me to make it better. She has been a role model for me since I met her at Kings College London over twelve years ago. I would also like to thank Jessica Davenport for her help – her input helped me tell a story with this thesis and her enthusiasm has been a great source of motivation. Finally, I would like to thank my family for their guidance, support and understanding (especially during the countless hours that I spent glued to the computer screen).

INDEX

LIST OF OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

A Level: Advanced Level

AERA: American Educational Research Association

AFL: Assessment for Learning

APA: American Psychological Association

AS Level: Advanced Subsidiary Level

BOE: Boletín Oficial del Estado

CAS: Creativity, Action and Service

CEFR: Common European Framework of Reference for Languages

CLIL: Content and Language Integrated Learning

CP: Career-related Programme

CPE: Centros de Profesorado

DP: Diploma Programme

EACEA: European Education and Culture Executive Agency

EBAU: Evaluación de Bachillerato para el Accesso a la Universidad

EEC: European Economic Community

EFEPI: Education First English Proficiency Index

EvAU: Evaluación para el acceso a la Universidad

EC: Emerging Categories

EFEPI: Education First English Proficiency Index

EFL: English as a Foreign Language

ESL: English as a Second Language

ESO: Educación Secundaria Obligatoria

ESOL: Examinations for Speakers of Other Languages

EU: European Union

GCSE: General Certificate of Secondary Education

HE: Higher Education

HL: Higher Level

IA: Internal Assessment

IBO: International Baccalaureate Organization

IB: International Baccalaureate

IELTS: International English Language Testing System

INECSE: Instituto Nacional de Evaluación y Calidad del Sistema Educativo

LOCE: Ley Orgánica de Calidad de la Educación

LOE: Ley Orgánica de Educación

LOGSE: Ley Orgánica de Ordenación del Sistema Educativo

LOMCE: Ley Orgánica para la Mejora de la Calidad Educativa

LOMLOE: Ley Orgánica de Modificación de la LOE

MA: Master of Arts

MEC: Ministry of Education

MFL: Modern Foreign Languages

MYP: Middle Years Programme

NCME: National Council on Measurement in Education

OECD: Organization for Economic Co-operation and Development

PA: Peer Assessment

PAU: Pruebas de Acceso a la Universiad

PEDLA: Plan Estratégico de Desarrollo de las Lenguas en Andalucía

PEVAU: Pruebas de Evaluación de Bachillerato para el Acceso a la Universidad

PISA: Programme for International Students Assessment

PYP: Primary Years Programme

QDA: Qualitative Data Analyses

RQ: Research questions

SEK: San Estanislao de Kostka

SIIU: Sistema Integrado de Información Universitaria

SUEE: Spanish University Entrance Examination

TOEFL: Test of English as a Foreign Language

UNED: Universidad Nacional de Educación a Distancia

UK: United Kingdom

ABSTRACT

Teachers in Spain are greatly aware of the importance of learning a foreign language, especially English. However, many find it hard to engage their pupils in classroom dialogues and to find an appropriate form of assessment of oral skills. Teachers do not give sufficient importance to the speaking assessment in order for students to prioritise oral skills and actively participate in classroom discussions. Using peer assessment (PA) on a regular basis to assess students' oral skills in English language learning is a strategy that allows them to be participants in their own learning and become more autonomous learners. This motivates them to do better, first, because they are more likely to listen to their peers' comments and second, because they are aware of a constant assessment process. Students tend to understand each other's comments better than those of the teacher, due to their honesty and the fact that they converse in 'their own language'.

In addition, it is important to create an atmosphere of trust in the classroom to encourage conversation between the assessor and assessed, thereby getting students to deepen their knowledge about English language learning. The objective of this study is to gain insight into how students engage with peer assessment in an English lesson, and the impact of this form of assessment on the classroom.

The researcher undertook a study over a five-month period at the only international school in Almería, Spain, with thirteen students who performed an oral task in pairs, for which they had to give each other feedback. The task was divided into four stages to allow students the opportunity to understand progressively the process of peer assessment. The data was collected and then coded from semi-structured interviews, firstly, of the students with their classmates and then, of the students with their teacher. The move to peer assessment is not simple for teachers and students and building a trusting atmosphere in the classroom is essential to its success.

The findings show that it is possible to conduct peer assessment in secondary school English classrooms and that providing pupils with more agency and autonomy in their learning can lead to huge benefits beyond those available from the task in hand. Some of these benefits include

greater self-confidence, improved critical thinking skills as well as strengthened belief in self-assessment and independent learning.

RESUMEN

Los profesores en España son conscientes de la importancia de aprender una lengua extranjera, especialmente inglés. Sin embargo, a muchos les resulta difícil conseguir que sus alumnos participen en las conversaciones en clase y encontrar la forma apropiada de evaluar las destrezas orales. No le dan la suficiente importancia a evaluar como hablan los alumnos para que se esfuercen en priorizar esta destreza y participen activamente en la clase. El uso de la evaluación por pares de forma regular para evaluar las destrezas orales de los alumnos en inglés es una estrategia que les permite ser partícipes de su propio aprendizaje y convertirse en alumnos más autónomos. Esto les motiva a mejorar, primero, porque escuchan más los comentarios que sus compañeros puedan darles, y segundo, porque son conscientes de que están siendo evaluados constantemente. Los alumnos tienden a comprender los comentarios entre ellos mejor que los del profesor debido a su honestidad y el hecho de que hablan 'el mismo lenguaje'.

Además, es importante crear un ambiente de confianza en la clase que anime a los alumnos a hablar entre ellos, el que evalua y el que es evaluado, y de este modo profundicen en su conocimiento del inglés. El objetivo de este estudio es obtener información sobre cómo los alumnos se involucran en la evaluación por pares en una actividad en inglés y el impacto que tiene esta forma de evaluar en la clase.

La investigadora llevó a cabo un estudio durante cinco meses en el único colegio internacional en Almería, España, con trece alumnos que realizaron una tarea oral en parejas para la que tenían que darse 'feedback' los unos a los otros. La tarea se dividió en cuatro etapas para dar a los alumnos la oportunidad de comprender el proceso de evaluación por pares progresivamente. Se recogieron los datos y después se codificaron de entrevistas semi-estructuradas, primero, de los estudiantes con sus compañeros de clase y después, de los estudiantes con su profesora. El paso hacia la evaluación por pares no es sencillo para los profesores ni para los alumnos y crear un clima de confianza en la clase es esencial para que se pueda realizar con éxito.

Los resultados muestran que es posible llevar a cabo la evaluación por pares en las clase de inglés en secundaria y que proporcionar a los alumnos más agencia e independencia en su aprendidaje puede conducir a enormes beneficios más allá de los disponibles de la tarea en cuestión. Algunos de esos beneficios incluyen una mayor confianza en sí mismos, una mejora de sus destrezas de pensamiento crítico además de una mayor creencia en la autoevaluación y en el aprendizaje independiente.

CHAPTER 1: INTRODUCTION

1. BACKGROUND AND PURPOSE OF THE STUDY

Regardless of the recent efforts in Spain to improve the teaching of foreign languages, when trying to have an authentic conversation students' oral skills in the main continue to be poor. According to Arnaiz Castro (2017), in relation to the level of English the country continues to be ranked towards the botton of the European states. The last European Comission (2012) report revealed that just 22% of the over eighteens claimed to have a good level of English, with only a 7% reporting having a high level. In contrast, in The Netherlands 90% of the population had a high level, whilst in Sweden and Denmark the figure stood at 86%. Some students fail to gain access to some universities abroad owing to their inability to communicate orally in English effectively. In sum, many Spanish students lack the ability to reap the benefits of having proficient oral English language skills.

Palacios-Hidalgo et al., (2021) have noted how European countries have been developing quality programmes aimed at helping their citizens to become competent in English communication. In Spain, in the 2000-2001 academic year, students started to encounter English at six years old and since the academic year 2006-2007, they have done so at three. They continue to study it until the end of their secondary schooling (including when undertaking Bachillerato). Spain is the only country apart from the German speaking area of Belgium where English is compulsory from kindergarten onwards. That is, most EU countries do not start teaching English until children are six, whilst in Finland this does not happen until they are nine (Arnaiz Castro, 2017).

With the aim of making their citizens more compentent in the English language, Andalusia and the Community of Madrid became pioneers in developing bilingual education (BE) programmes that involved studying curriculum subjects (such as science, history and geography) in non-native language. These were introduced in Andalusia in 2005 through the 'Plurilingualism Promotion Plan'[1], with the goals of reinforcing modern foreign languages (MFL) learning and communicative proficiency (Junta de Andalucía, 2005). Subsequently,

[1] Plan de Fomento de Plurilingüismo, language policy document regulating language teaching and learning in Andalusia (Spain) for the period 2005-2015.

the 'Strategic Plan for the Development of Languages in Andalusia (PEDLA)[2]' was implemented, which updated the earlier guidelines, by focusing more on interculutural competence (Junta de Andalucía, 2016). At the start of the 2019/20 academic year, 25.31% of public schools in the automous region of Andalusia were provided BE in English, French or German (Junta de Andalucía, 2019), with most studying in the foremost language.

Despite these developments, studies undertaken by researchers such as Serrano and Muñoz (2007), have elicited that the number of hours students in bilingual schools study in a different language does not correlate positively with their learning it better. Moya Guijarro and Ruiz Cordero (2018) undertook a comparative study with students in the 4th ESO (year 5) on the bilingual programme and those who were not. It emerged that, whilst the students following the bilingual programme got slightly higher marks, in particular, in oral communication, it would have been desireable that they had obtained even higher ones.

Despite the Spanish education policymakers aiming to respond to Europe's suggestions and guidelines for delivering the competency in at least two MFLs (Council of Europe, 2018), BE programmes in Spain have been increasingly criticised by parents, policymakers and even teachers (Burgos, 2017, Madrid Fernández et al., 2019). This could be attributed to the BE schools being mislabelled as 'bilingual schools', when the reality is that they are content-based programmes with varying levels of exposure to the MFL and in essence, most 'have little to do with bilingualism' (Roldán-Tapia, 2012, p. 71). In spite of all the time spent learning English and the resources put into bilingual education, many students struggle to hold a basic conversation in the language (Rubio Alcalá and Martínez Lirola, 2008) and the evidence suggests that little heed is being paid to this fact.

In today's globalised world, English is the dominant language of communication and on the Internet, which provides opportunities for developing proficiency in both formal and informal settings (Education First, 2018). However, as Seedhouse and Jenks (2015) have pointed out,

[2] The acronym stands for Plan Estratégio de Desarrollo de Lenguas en Andalucía, a language policy document setting strategic obectives for language teaching and learning in Andalusia (Spain) for the period 2016-2020.

most students' only contact with English is at school and the use they will make of it once they leave school will depend on what they have learnt in the classroom.

Recently, scholars have recommended teaching in the second language (L2) 90% of the time (Hlas, 2016). Amengual-Pizarro (2019) found that some teachers experience high levels of anxiety in the classroom, which can be attributed to their lack of English proficiency. Amir (2015) contended that it is crucial that teachers have mastery of their subject whilst Arnaiz and Castro (2017) stressed the importance of special teacher training for those involved in BE.

Shin et al., (2020) compiled a review of L1 usage in foreign language classrooms from empirical literature ranging from 2011 to 2018. They concluded that teachers' speech in L2 needs to take into consideration the students' understanding and level of language proficiency. Morever, giving students sufficient time for the internalisation of a teacher's speech is essential. This could be followed up by asking them whether they understand through retelling the teacher's speech with their peers in L1 or asking them to take notes about this speech and giving them time to check it. In their review, most students were in favour of the use of using L1 in the classroom as a resource and not as a medium.

In addition to using the L2 to communicate with the teacher, students need to interact with their peers in class, which has been shown to foster the co-construction of language and learning (Tognini et al., 2010). When referring to the judidicious usage of L1, this should be assessed according to its purpose, content and task (Shin et al., 2020).

Spain continues to invest considerable resource in bilingual education, with the aim of students becoming more proficient in English (Arnaiz Castro, 2017), however, there remains the absence of oral skills in the Selectividad University Entrance Exam (SUEE). Consequently, English teachers in Bachillerato dedicate most of the time to just teaching to the test. That is, they instruct their students in how to succeed in their SUEE exam and the skills that they will be evaluated on receive the most attention. According to Bueno Alastuey and Luque Agulló (2012) and Amengual-Pizarro and Méndez García (2012), oral skills are omitted from the format of the exam, which has a detrimental impact on the teaching of these.

When students are not assessed, neither teachers nor students give these skills a high priority, whilst some students do not make any effort to speak, especially the shyer ones, afraid of making mistakes and that their classmates might laugh at them (Aleksandrzak, 2011; Fauzan, 2016). Teaching and learning are greatly affected by testing methods (Alderson and Wall, 1993; Clarke, 1991; Sanmartí, 2010; Taras, 2008; Wiliam, 2010), so perhaps a more formative assessment by their peers can make them accept their mistakes without being afraid of speaking. Moreover, this formative assessment, as suggested by Fauzan (2016), enables students to assess themselves as well as their peers, involving them in the assessment process, which helps their language learning and makes them more autonomous learners (Grenfell and Harris, 1999; Ion et al., 2019; Salaberri, 1995; Sanmartí, 2010; Webb and Jones, 2009).

The researcher's experience with assessment as a student, first and then, as a teacher have encouraged her to write this thesis. Firstly, reflections on the researcher's personal experiences are provided, both when she was at school with regard to the assessment of oral skills in Modern Foreign Languages (MFL) and then, her professional background as a teacher and researcher at an international bilingual school in Andalusia (Spain). Secondly, the background of oral assessment practices at schools in Spain, especially formative assessment and more specifically peer assessment practices are provided. These practices are the focus of this research. Then, the difficulties teachers face when assessing oral skills in languages are explored as well as the impact that the lack of proper assessment can have on students' futures.The different types of educational assessment will be investigated giving special emphasis to those in which the students become participants of their own learning. This type of assessment might help teachers to assess students in a meaningful and fair way in the language classroom and hence, enable them to enhance their progress.

1.1. The researcher experience with assessment

The researcher's first experience regarding educational assessment as a student was in the 1970s, in Spain. At that time, it was viewed as a means to promote or fail a pupil based on attained grades. The grades in all subjects were numerical, from 1 to 10. The report only

provided numerical grades– it provided no written feedback on the student's progress (See Appendix 1). The following scale indicates the meaning behind each numerical grade:

0-4 Insuficiente
5 Suficiente
6 Bien
7-8 Notable
9-10 Sobresaliente

In the reports, there was no precedent or tradition for providing comments and receiving two failures (insufficient) would mean the student would fail to progress into the next academic year. This assessment practice based on failure was widespread and willingly accepted at the time, although with hindsight, it is clear that it failed to motivate slow or reluctant learners. This system did little to guide students towards improvement – it served merely to highlight attainment and non-attainment. In addition, it stigmatised failure, was based on negative rather than positive reinforcement and did little to provide any positive experiences to learning. Given that learning is most productive in an environment with positive reinforcement (Pressey, 1962; van Duijvenvoorde et al., 2008), this type of assessment would have been detrimental to many youngsters already having difficulty early on in their educational journey.

The assessment methods used at the time did not prioritise the students' positive classroom experience, nor encourage them to take responsibility for their learning. This assessment practice promoted students to the next stage as a result of having achieved the required grades to secure a pass and hence, this process was summative. It provided an assessment on performance to an extent and was used to promote pupils into the next higher class, and for managerial purposes (Tanner and Jones, 2003).

This was about forty years before the time of writing this thesis. Since then, educational research into teaching and learning has expanded significantly and there has been a considerable amount of research on AFL, especially in the UK (Black and Wiliam 1998a; Jones and Wiliam, 2008; Leathy and Wiliam, 2015). However, in Spain research has been scarce on formative assessment and even though research has become internationalised, most teachers lack clear guidelines about what type would be the most suitable for developing new

approaches in their classrooms. This study was aimed at exploring formative oral assessment practices, specifically peer assessment, in an international bilingual school in Spain and to discover to what extent they can improve students' oral skills in the MFL classroom.

1.2. Purpose of the study

A great many modern foreign language teachers find it hard to engage their pupils in classroom dialogues to become better at talking in the target language and even more so, to assess them (Aleksandrzak, 2011; Chou, 2018; Correia, 2016). Whilst teachers are aware of the importance of learning how to speak foreign languages proficiently, especially English (Amengual-Pizarro, 2009), some do not give enough weight to the assessment of speaking, so that students can give it a high priority and become interested in participating in the classroom (Hornero et al., 2013).

Teachers in Spain sometimes feel frustrated by the Spanish University Entrance Examination (SUEE)[3], for which there is not an oral exam (Rubio and Tamayo, 2012). This study takes place in an Andalusian school and according to the current Andalusian educative law, the exam of the Second Foreign Language in SUEE only assesses the student's reading comprehension and written expression, as the official state document asserted on the 26th of December (Real Decreto 105/2014 de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato). Even though teachers know that their responsibility is to teach their students to communicate verbally, a considerable amount of them tend not to do so and teach them to pass the test, thus failing to improve their communicative competence (Bolívar, 2015). This may have implications in the students' future personal and vocational lives (Cribb, 2011). According to Gleeson and Knights (2006, p. 2), some teachers might feel like 'de-professionalized victims', who, as is highlighted by Jiménez Raya (2007), wish they could be allowed to be more autonomous in their practices to be accountable for their pupils' needs.

[3] Prueba de acceso a la Universidad (PAU)

In order to have effective conversations in the class, students need to feel comfortable and trust their teacher, who must be supportive, always giving them enough time to think before they speak (Black and Jones, 2006; Hodgen and Marshall, 2005). It is necessary to create an atmosphere of trust (To and Panadero, 2019), where students are interested in what their classmates think for the collaboration needed in peer assessment (Black et al., 2003; Liu et al., 2018). Some students might be reluctant to give or receive feedback, because they believe they speak worse than most of the students in the class and want to avoid looking bad in front of their classmates (Hamzah and Ting, 2009) or because they feel they are not ready for peer learning (To and Panadero, 2019). This is closely related to the competitiveness which there is sometimes among them since some weaker students do not want to show their lower level of English in front of their peers (Cestone et al., 2008).

This research is rooted in formative assessment, which has been widely recognised as fostering learning (e.g., Black and Wiliam, 1998b; Fuentes-Diego and Salcines-Talledo, 2018; Hattie, 2012; Hortigüela Alcalá et al., 2016; Jones and Wiliam, 2008; Leahy and Wiliam, 2015; López-Pastor and Pérez Pueyo, 2017; Sanmartí, 2010), although this is unclear to others, like Willis (2011), who did not find a conclusive relationship between Assessment for Learning (AFL) and learner autonomy. Many researchers have extensively studied how assessment is the engine of learning, and what and how we teach depends on it (Fuentes-Diego and Salcines-Talledo, 2018; López Pastor and Pérez Pueyo, 2017; Sanmartí, 2010). This chimes with one of the approaches to teaching in the Language B acquisition guide of the International Baccalaureate Organization (IBO) for the Diploma, which is a publication produced to support the Diploma Programme (DP)[4].

Formative assessment is a tool or process that teachers can use to improve student learning; it is about assessment for learning, rather than simply assessment of learning. “In the language acquisition classroom, there should be opportunities for both teacher-led formative and summative assessment as well as peer assessment and opportunities for students to reflect on their own learning and performance on assessment instruments in order to develop their language skills” (International Baccalaureate Organization, 2018a, p. 58).

[4] The Diploma Programme is a rigorous pre-university course of study designed for students in the 16 to 19 age range.

Pupils must participate in their own learning in order to make progress, giving feedback to their classmates, their teachers and assessing themselves as part of everyday teaching (Hattie and Timperley, 2007; López Pastor and Pérez Pueyo, 2017). As Sanmartí (2010) suggests, the purpose of assessment is to comprehend why students do not understand a particular task and their mistakes can give us a great amount of information to help them change their strategies. Black et al., (2003) indicate that teachers should share or even co-create success criteria with their pupils and use peer, self-assessment, and summative tests as formative, giving them no marks, but rather, comments or feedback for improvement. This might help students internalise the criteria, understand them better as well as activating their learning strategies (Fraile et al., 2017).

The focus of this study is peer assessment, which is common practice in MFL lessons and a valuable formative assessment tool, despite some limitations or the perceptions of some students, which might not always be positive, since, as the work of To and Panadero (2019) shows, they sometimes do not have trust in what they are doing. Students might not make an objective assessment of their peers' work depending on the classmate they are assessing (Foley, 2013). Goodman (2011) asserts that some pupils do not like the idea of getting more formative comments and fewer marks. Jones and Wiliam (2008) have observed that students need to take part actively in their own learning and teachers need to give them strategies to become independent. We must be mere guides they consult as a last resort and teach them self-management skills (Boekaerts, 1995; López- Pastor and Pérez Pueyo, 2017; Sanmartí, 2010). Collaborative learning and social interaction can help students to self-regulate their own learning since they can talk in groups about their own learning process (Coyle, 2007). Moreover, pupils who are engaged in peer assessment can develop into more independent, motivated, and successful language learners (Jones, 2014) since they are able to use self-management skills to build their own learning. Despite there being a considerable amount of research about formative assessment, especially in Higher Education, which will be useful to explore in the literature, research is scarce on Secondary Education and in the IB classroom.

Based on the above and in response to those teachers who want to practise assessment in a fair and meaningful way, despite the constraints that might be found, this reseach pertains to peer assessment of the productive (speaking and talking) and receptive oral skills in an International Baccalaureate (IB) classroom. The students do a presentation, interview each

other and finally they give feedback to each other following certain set criteria. This hopefully could help a great number of students with language learning in any MFL classroom. Finally, it is important to consider the researcher's background and interests to justify this research.

1.3. Reasons motivating the development of this study

After twenty-five years teaching, the researcher has participated in and led oral skills assessments at a variety of levels, ranging from year 7 (11 years old) to year 13 (18 years old) and with a range of objectives. Especially in the last decade, educational assessment has become an area of particular interest for her. This is due to her own professional development as well as her conviction that oral assessment, if used for formative purposes and with specific feedback on ways to progress, can help students improve their oral skills and thus, achieve their goals. There is a wide variety of educational assessment practices that are used for different purposes, which will be discussed in the second chapter of this thesis. Each form of assessment can provide helpful information to reflect learning or performance, including summative assessment, which is used for entrance purposes into some schools and careers. However, the focus of this study is assessment with regard to foreign language and it will be argued that oral formative feedback has a positive impact on the summative results following an exam taken at a specific stage in the student's education.

After completing her studies, the researcher began teaching English language, and in 1999 joined SEK[5]-Alborán International School in Almería, where she has bee working for over 22 years. Spending the bulk of her career at the school was a wonderfully enriching experience. She joined the year the school was founded, with a limited number of students and the aura of

[5] San Estanistao de Kostka (SEK) Education Group was founded in 1892. There are SEK International Schools in Madrid, Barcelona, Almería, Pontevedra, Dublin, Les Alpes, Doha and Riyadh. Camilo José Cela University was also founded in the year 2000 as part of SEK Educational group. The hallmarks of its education model are innovation, interdisciplinarity and international approach. https://www.sek.es/en/sek/about-us/

a start-up and witnessed its impressive development to become the leading school it is now. Some of her contributions include being Head of the English department and coordinating the early implementation of the IB programme (2008) – one of the first in the region (even before the creation of a framework to harmonise it with national universities). It was not until the academic year 2020-21 that a framework to harmonise the IB programme with the national universities was created. In the previous twelve years, students had to take both Baccalaureates: National and International. Also, they had to do the SUEE, which is the way they still do it in some schools. This is due to the fact that the IB is a very challenging programme and it is harder to get the highest marks in the DP than in the SUEE. The evaluation of the DP is based on a combination of final exams and assignments that are assessed internally, but they are also moderated by the IB. Subjects exams are held on the same day all over the world (sometime in May for schools in the North Hemisphere).

Oral skills are not given enough weight in Secondary Education assessment in Spain (García Laborda and Fernández Alvarez, 2011; Rubio Alcalá and Martínez Lirola, 2008) even though all teachers are aware of the importance of learning how to speak foreign languages proficiently. Teaching at the only IB School in the city, SEK-Alborán International School, one in which teachers are always encouraged to try new practices that improve their students' learning, has made possible to pursue this study. Formative assessment allows students to take responsibility for their own learning, which is essential, not just in school, but in life in general (Carless and Boud, 2018; Stobart, 2008; Wanner and Palmer, 2018). Reflection, self-assessment, self-managing are central pillars of Assessment for Learning (AFL), encouraging students to become more independent in their learning. AFL is an approach to formative assessment in which feedback is continually incorporated in the classroom practices to guide the future learning of students. The fact that they have to assess themselves as well as their peers can help foster students' autonomy as students learn how to learn (Heitink et al., 2016).

The researcher was motivated to understand to what degree an international school, which promised to maximise each students' full potential, made use of formative assessment. Many authors link the skill of self and peer assessment with effective life-long learning (Boud and Dawson, 2021; Molloy et al., 2020; Stobart, 2008; Tai et al., 2017). According to Tai et al., (2017), for feedback processes to be effective, students must be able to make evaluative judgements on the quality of their own work and that of others. In terms of their own

performance, they need frequent opportunities to engage in self-evaluation, thereby strengthening its accuracy over time (Boud et al., 2015). When learners come to realise that they need to change what they are doing, they are more likely to do so (Boud and Molly, 2013). Peer feedback is an effective way of getting students to share evaluation of their own and the work of others (Carless, 2015). It can facilitate self-assessment of their own production by allowing for comparisons with that of others (McCologue, 2015). Giving feedback to peers requires significant cognitive engagement covering, such matters as the application of set criteria, diagnosis of problems and putting forward solutions (Nicol et al., 2014), all of which can be considered lifelong learning skills.

A dilemma emerged in this researcher's teaching, which triggered the exploration of the connection between formative assessment, especially peer assessment, and the improvement of oral skills in a MFL DP classroom in Spain. The students who did not participate in class did not show much improvement in oral skills and of those who did, some did not take the exercise seriously as they were not formally assessed. This posed a question in the teacher's mind: can peer assessment be beneficial for the students' oral skills in a MFL classroom? Teaching a 13-student class, all of whom were keen to participate in a peer assessment activity, made this research possible.

2. OBJECTIVES AND RESEARCH QUESTIONS (RQ)

2.1. General and specific objectives

The general objective of this study is to raise teachers' awareness of how best to use peer assessment formatively during the teaching of oral skills in Diploma Programme language acquisition.

From this general objective described above, the following specific objectives arise:

1. To explore how a group of 1st year Diploma IB (International Baccalaureate) students engage in peer assessment in an English as a foreign language (EFL) classroom in an Independent School in Spain and the impact it has on classroom talk.

2. To study the students' emotional responses towards peer assessment of oral skills.

3. To examine the teacher's perceptions regarding the indicators of improved student performance.

4. To investigate the benefits of peer assessment for the assessed and the assessor.

5. To look into how the practice of peer assessment can help students become more independent learners, with the ability to build their own learning.

2.2. Research questions

From the general objective the following research question is derived: How can we best raise teachers' awareness of how best use peer assessment formatively during the teaching of oral skills in the Diploma Programme Language? From the specific objectives the following research questions are derived:

RQ1: To what extent do peer-assessment practices in an EFL classroom in an independent school in Spain improve students' talk?

RQ2: What are the students' emotional responses to practising peer-assessment in oral skills?

RQ3: What are the teachers' perceptions about indicators of improvements in student performance?

RQ4: How do the peer assessed or the assessor benefit from peer assessment?

RQ5: To what extent does peer assessment help students to become more independent learners, able to build their own learning?

The data which will be used to answer every objective and RQ are indicated below:

Objective and RQ1 are addressed by selecting sections of the transcripts with the feedback that students give to each other in an oral activity. Aspects such as affective language and praise are analysed to this end.

Objective and RQ2 are tackled by analysing sections of the transcripts in which students show their emotional responses when giving feedback to their peers. Positive and critical comments are selected to exemplify the students' reactions to the task.

Objective and RQ3 are addressed analysing the data in which students give comments using straightforward language and criteria language.

Objective and RQ4 are to be answered by analysing the different types of feedback that students give to each other.

Objective and RQ5 are addressed by analysing the data in which students assess themselves and use forms of self-regulation of their own learning and/or critical thinking.

The hypothesis of this study is that the use of peer assessment has the potential to improve students' development of oral skills in EFL and to make them more independent. Taking an active role in assessment and providing feedback for their classmates may help students to build their own learning.

3. THESIS OUTLINE

This thesis is organised into eight chapters: 1. Introduction, 2. Theoretical framework, 3. Methodology, 4. Data analysis, 5. Results, 6. Conclusions 7. References and 8. Appendices. The content of each chapter is described below.

The first chapter introduces the background and purpose of the study. It explains the rationale behind its development, outlines the general and specific objectives, research questions and hypothesis as well as sets out the structure of the thesis.

In the second chapter, the literature of the central current debates on the subject is reviewed. Firstly, this chapter covers the difference between evaluation and assessment. Then, it deals with educational assessment, including why learning is assessed and the different modes of assessment, paying special attention to formative assessment and self-regulation in languages teaching. Secondly, it provides a brief contextualisation of EFL assessment and description of the International Baccalaureate (IB), both in Spain and abroad. Then, some issues and problems of assessment in the context of secondary schools, both in Spain and abroad are explored. Finally, this chapter introduces the different ways of assessing oral skills, especially peer assessment. The constraints and possibilities in assessing oral skills are also covered.

Chapter three provides the methodology of the study. It is divided into several sections describing the context of research, the participants, the student's voice, the research approach, the research plan and the instruments and procedures involved in data collection. It also explores the constraints and possibilities when assessing oral skills as it can be time-consuming, whilst it is also being challenging to assess reliably and students are sometimes reluctant to perform oral tasks. Finally, some ethical considerations are contemplated and a brief conclusion is provided to summarise all the previous issues raised.

Chapter four presents the data analysis and is divided into several sections, according to the four emerging categories (EC) identified during data coding: 1. Students' emotional responses to practising peer-assessment, 2. Benefits for the peer assessed and for the assessor, 3. Construction of their own learning and 4. The researcher's perceptions. These sections, in turn,

include several subsections, where some subcategories are identified and explored in the students' transcripts.

Chapter five introduces and discusses the research findings. Following critical discussion, this chapter provides responses to the objectives and the research questions. In Chapter six, the contributions to knowledge in the field and in methodology as well as future research paths are discussed. The researcher summarises the main findings and some limitations of the study are identified and explained. The references of the bibliography used in this thesis are in Chapter 7. Finally, Chapter 8 incorporates the appendices used in this thesis.

CHAPTER 2: THEORETICAL FRAMEWORK

1. INTRODUCTION

As assessment is at the heart of this study, various aspects regarding it will be explored in this chapter, through the critical discussion of a range of relevant literature relating to the broad and specific aims of the investigation and to the research questions. Firstly, the difference between evaluation and assessment is considered. These two concepts are by some viewed as synonymous, whilst others see them as being distinct. Secondly, this chapter explores educational assessment, which as Gordon (2020, p. 72) states, "can and should be used to inform and improve teaching and learning processes and outcomes", why learning is assessed, the different purposes of assessment in English as a foreign language as well as the distinction between the summative and formative forms.

This chapter also explains how the power of formative assessment lies in merging five strategies in which the teacher works together with the students to reveal how they are progressing towards achieving their objectives and what further might need to be done to help them achieve these objectives with success.

The focus of this study is on formative assessment, also known as Assessment for Learning (AFL) by classroom teachers, which has been found to accelerate learning through the provision of valuable feedback. It is about the meaning and ways of assessing oral skills, in this case in the IB classroom, with a focus on peer assessment.

Then, the context of EFL assessment and the neglect of oral skills which has been revealed in the academic literature (Alonso, 2014; Luque Agulló et al., 2016; Plo et al., 2014) is discussed. Despite more importance having been given to oral skills in the last few years and more teachers starting to use the communicative method to teach EFL, there is still a large number of language teachers who follow a very traditional approach (Santos Estrada, 2016, García Sampedro, 2021) and do not assess oral skills in the classroom (Rubio Alcala y Tamayo Rodríguez, 2012). In Spain, this is partly influenced by the absence of an oral exam in the SUEE (Bueno-Alastuey and Luque-Agulló, 2012). Spanish students applying for the university only take a written English exam and, as a result, teachers focus on writing for the last two

years of schooling, providing few opportunities for oral skills practice (Rubio Alcalá and Martínez Lirola, 2010).

Firstly, it will be necessary to differencitate between Evaluation and Assessment. As aforementioned, the school in which this research has taken place is an international school that follows the International Baccalaureate (IB) curriculum. Therefore, it is necessary to describe what the IB is, how it started as well as its legislation, origin, and history. Given the school is in Spain, it is also necessary to explain when the IB was implemented in that country and more specifically, in the school where this research was undertaken (SEK-Alborán International School). There is also discussion on the IB methodology, its criteria, curriculum, and different IB programmes. Then, the isssues and problems of formative assessment in the language class will be is explored. Finally, the different ways of assessing oral skills as well as their constraints and possibilities, especially in peer assessment, are investigated.

2. EVALUATION

As stated by Sanmartí (2018), to evaluate is a process based on collecting information through different instruments, analysing it, and giving it a value to take a later decision. Harris and McCann (1994) explain that evaluation "involves looking at all the factors that influence the learning process, such as syllabus objectives, course design, materials, methodology, teacher performance and assessment" (p. 2). On the one hand, evaluation can be summative, where the aim is to certify before students, parents, and society, in general, the level of knowledge a student has acquired at the end of a unit or a learning stage. On the other hand, evaluation can be formative and pedagogic in character, with the aim of identifying whether the objectives and activities planned are the right ones for a specific group. Evaluation is performed to decide whether the student meets a predetermined criterion, and it evaluates instruction. For instance, systemic evaluation is to determine how effectively an education system is functioning and assessment is an essential aspect of any such evaluation, being the way of judging the quality of content of the programmes provided to students (Ghaicha, 2016). However, assessment has an individual focus as it assesses learners' needs and records their improvement towards meeting objectives (Kizlik, 2012).

3. ASSESSMENT

Broadly speaking, the terms assessment and evaluation are used interchangeably by many practitioners. However, research suggests that they do not have the same meaning. The term 'assessment' implies an objective process of measurement of students' learning (Drummond, 2011; Linn and Gronlund, 2000) and of obtaining information (Desforges, 1989; Rowntree, 1987). Assessment gives us information for making decisions, as noted by Bachman and Purpura (2008). It must be designed and used in a way that maximises "the chances for fair and equitable treatment of individuals and groups in terms of their access to opportunities based on merit" (Bachman and Purpura, 2008, p. 461). According to Stobart (2008), the strength of assessment can be appreciated in everyday life, as he states that "Assessment in the broad sense of gathering evidence in order to make a judgement, is part of the fabric of life" (p. 5). It is now well established from a variety of studies (Boud and Soler, 2016; Broadfoot and Black, 2004, among others) that assessment plays an important role in education, for we must assess what our students have learnt to have evidence of how well or badly they are reaching their learning goals (Baird et al., 2017). That evidence not only helps us take decisions to improve our pupils' learning, but also, to reflect on our practices so that we can change them, if they are not working as expected (Suskie, 2018). It is, therefore, essential to explore the definition of assessment and its purposes.

Broadfoot and Black (2004) suggest that "Assessment must be understood as a social practice, an art as much as a science, a humanistic project with all the challenges this implies and with all the potential scope for both good and ill in the business of education" (p. 8). Therefore, the process of assessment is complex and varied, as it applies to different approaches to gathering evidence. They also highlight that the link between evaluation and assessment is due to the fact that assessment provides teachers with the most valuable information about what is taking place in a learning environment. Regarding the assessment of oral skills, Harris and McCann (1994) point out that it is better to do it in groups or in pairs practising peer assessment, despite the damage that criticism from own classmates can sometimes cause.

At present, there are researchers on assessment, in general, such as Sanmartí (2020), who claim that to assess and to learn are the same process. To learn is to be able to identify what your ideas are, understand the reasons that explain them and take decisions in order to change

them, if necessary. As a result, learning is assessing (Sanmartí, 2020). Chiming with this, Suskie (2018) clearly demonstrates the link between assessment and learning, stating that "Assessment is deciding what we want our students to learn and making sure they learn it" (p. 8). Hence, assessment is successful when the students' learning is understood and used to enhance the curriculum (Drummond, 2011; Weeden et al., 2002). This broader view of assessment gives the student a more active role in trying to find out what they know, understand, and can do to improve and accelerate learning, as Cauley and Macmillan (2010) observe.

There is evidence that a specific test can be summative and formative, depending on how we use the information provided by the test (Bennet, 2011; Irons, 2007). Research has indicated that identical test questions can be used for varied and distinct purposes (Brown, 2004; Gipps and Stobart, 1993; Harlen and James, 1997; Webb, 1992). There is a great number of types of assessment used for a wide variety of purposes.

3.1. Formative and summative assessment

Assessment and learning are at the heart of the educational experience. This can be seen in the tension between the functions of formative and summative assessment, i.e. assessment in support of learning and as that geared towards validation and accreditation. Black (1998) does not see a big difference between the summative and formative assessments and concludes by saying that it is how you use it that matters, as this quote indicates.

> Some have laid stress on the differences between the formative and summative purposes and have argued that the assessment instruments and procedures needed for the one are so different from those for the other that neither can flourish without clear separation. On the other side, it can be argued [as Black does] that the two functions are two ends of the same spectrum and that there is no sharp difference, and that if the two functions are separated, then teachers' assessment work will be devalued. (p. 34)

Black (1998) suggests that the difference between formative and summative assessments is their purpose and function. However, he also states that both types of assessment sit within the same "spectrum of practice in school-based assessment rather than two isolated and completely different functions" (p. 35).

Formative assessment is defined by the Standards as "an assessment used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning with the goal of improving students' achievement of intended instructional outcomes" (American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME), 2014, p. 219). Dunn and Mulvenon, (2009) highlight the consensus on the definition of formative assessment and the limited empirical evidence concerning its impact in education. Wiliam and Leathy (2016) explain what formative assessment is using the following definition:

> Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited. (Black & Wiliam, 2009, p. 9)

The term formative assessment is often misunderstood as it is seen as a tool rather than a process. Wiliam emphasises the importance of process with the use of right tools. Leenknecht et al., (2021) refer to formative assessment as follows:

> When looking at assessments as a formative practice, the roles of the teachers, the students themselves and their peers are recognised, and the developmental role of the assessment is highlighted. In this practice view, formative assessment is seen as a cyclical programme of high and low-stake tasks in which students are actively involved (as assessee and/or assessor). (p. 2)

Cizek et al., (2019) propose the following definition:

> As part of a planned assessment system, formative assessment supports teachers' and students' inferences about strengths, weaknesses, and opportunities for improvements in learning. It is a source of information that educators can use in instructional planning and students can use in deepening their understandings, improving their achievement, taking responsibility for, and self-regulating, their learning. Formative assessment includes both general principles, and discipline-specific elements that comprise the formal and informal materials, collaborative processes, ways of knowing and habits of mind particular to a content domain. (p.14)

The definition above incorporates four elements: inference, self-regulating learning, disciplinary specificity and being part of a planned system. However, only the two of them of relevance to this study will be explained below: 1. inference and 2. self-regulating learning.

1. Inference: the educator's or students' judgements about the response provided by students in the oral task performed generally go beyond the response itself. The response is an indicator of what the student knows and can do with respect to some task class, and it is that knowledge and capability we want to enhance. However, this response must be interpreted in context, taking into account the conditions surrounding the task performance, the student's background and past performances. The particular inference that is made forms the basis for the feedback given to the student and for the next instructional steps considered (Cizek el al., 2019).

2. Self regulating learning refers to when learners set goals, followed by their engaging in cognitive, affective, and behavioural practices and that facilitates the achievement of these goals (Zimmerman and Schunk, 2011). The association between self- regulated learning and formative assessment has been recognised (Andrade and Brookhart, 2019; Black and William, 1998b; Nicol and Macfarlane-Dick, 2006), with the key common theme being relating to students' being at the centre of the construction of their own learning (Penuel and Shepard, 2016). Formative assessment stimulates students drawing upon their cognitive and motivational capacities and orients them towards considering their learning goals. It also

facilitates their monitoring and reflection on their learning as well as providing feedback and strategies that can help them in achieving their goals. In sum, formative assessment can help students regulate their learning while at the same time fostering the development of salient metacognitive and learning management skills (Panadero, Andrade, and Brookhart, 2018). There has been research on various aspects of formative assessment, including self-assessment (Panadero, Jonsson, and Botella, 2017), and peer assessment (Panadero, Jonsson, and Strijbos, 2016; Reinholz, 2016).

Summative assessment is about ascertaining whether predetermined learning outcomes have been delivered in accordance with planned programmed objectives and/or the set requirements have been met to the extent that that accreditation or certification can be awarded.

> It is a form of information gathering about students that is conducted primarily for the purposes of making judgements about the status of individual learners or determinations about the effectiveness of educational programmes or systems. (Cizek et al., 2019, p. 6)

This gathering of data and information often rests on the assumption that objective truths can be revealed through some examples of assessment such as testing. However, Hanson (1993) goes further, noting that while testing is often considered "a means of measuring qualities that are already present in a person", the reality is that "tests to a significant degree produce the personal characteristics they purport to measure. The social person in contemporary society is not so much described or evaluated by tests as constructed by them" (p. 4). It would be superficial to consider assessment as only being a device for measurement without forming part of the learning process.

Summative assessment occurs at the end of an instructional programme, whereas formative assessment occurs alongside, or even as part of instruction. Summative assessments have the goal of making evaluative judgements about individuals or programmes, whilst the role of formative assessment is providing feedback and information aimed at adjusting ongoing teaching and learning. Stiggings (2005) distinguishes the difference as being "assessment of learning" versus "assessment for learning".

Miller et al., (2016, p. 288) explain the difference between summative and formative assessment as, in Stake's words: "When the cooks taste the soup, that's formative: when the guests taste the soup, that's summative". Cizek et al., (2019) hold that assuming some soup remains in the pot, or the chef plans to make it again, the guests' suggestions would constitute a formative deployment of summative information, which means that summative tests can bring forth formatively relevant information. However, using impressions obtained from a series of formative assessments for informing teachers' judgements in regard to allocating grades or recording achievement could lead to the formative assessment process being corrupted and hence, such assessment should not be used for summative purposes. In other words, whilst summative assessment data can be used formatively, the latter type of assessment should only be used to serve that end.

Both forms of assessment, formative and summative, are necessary to understand what students know, what they need to know, what they are learning or have learnt and how to adjust our teaching to make the most of students' learning. Teachers are expected to use them both when teaching languages (Taylor, 2017).

3.2. Feedback

Feedback is a crucial part of assessment which allows learning to progress and it forms a key part of formative assessment (Clarke, 2001; Panadero and Lipnevich, 2021; Stobart, 2008; Weeden et al., 2002). Feedback occurs during regular classroom interaction and can take many different forms, including orally as well as through a written response to the students' work. Gipps, in the context of primary education helpfully categorises feedback into two distinct groups: *evaluative feedback*, focussed on reward and punishment, including conveying approval and disapproval; and *descriptive feedback*, explaining to students whether they are correct or incorrect, why an answer is right or wrong, what they have and have not achieved, and finally, providing advice on how to improve (Tunstall and Gipps, 1996) (see Table 1).

EVALUATIVE FEEDBACK	*A1: Rewarding*	*B1: Approving*
	A2: Punishing	*B2: Disapproving*
DESCRIPTIVE FEEDBACK	*C1: Specifying attainment*	*D1: Constructing achievement*
	C1: Specifying improvement	*D2: Constructing the way forward*

Table 1: Typology of teacher feedback

Adapted from Tunstall and Gipps (1996, p. 389)

Various studies have assessed the efficacy of feedback and formative assessment to support learning (Nicol and Macfarlane-Dick, 2006). They argue that students in higher education already evaluate and provide feedback on their own work. They also offer some examples of feedback strategies that higher education institutions can use to build on this existing ability. Nicol and Macfarlane's (2006) positive impact on the concept of formative assessment can be attributed to their sound pedagogical recommendations, anchored in self-regulated learning. Their work proposed seven principles for "good feedback practice": (1) clarify what good performance is, (2) facilitate self-assessment, (3) deliver high quality feedback information, (4) encourage teacher and peer dialogue, (5) encourage positive motivation and self-esteem, (6) provide opportunities to close the gap, and (7) use feedback to improve teaching. Hence, their contribution to the topic provides comprehensive guidelines to educators. In sum, their model builds upon the work by Sadler by putting forward clear pedagogical recommendations.

However, the impact of feedback is not so simple (Graham et al., 2015; Perrenoud, 1998), for if teachers only give minimal feedback, then this does not really 'boost' pupils to the next stage. Pupils need regular and effective feedback on specific tasks that gives them opportunities to talk about their own learning (Brookhart et al., 2009; Nicol and Macfarlane-Dick, 2006). With regard to how to provide useful feedback, research suggests including approval and commendation (Grimm, 1986; Nilson, 2003; Saddler and Andrade, 2004) as well as

localisation. As Nilson (2003) suggests, “Instructors should present peer feedback items that ask students to identify or to personally react to defined parts of the paper, speech, or project” (p. 37). However, it is important to note that, while praise may motivate students, more specific feedback may be more useful. Rather than just telling the students what is right or wrong, teachers should give them examples, set targets, and establish criteria. In so doing, students can understand their goals and, with the feedback provided by their peers, can make changes and regulate their own learning, as suggested by Boud (2000):

> Unless feedback is applied and used to demonstrate improvement, there is no way to tell if it has been effective. In sustainable assessment learners need not only to be recipients of such feedback, but they need to be able to arrange it for themselves and know when it is complete. (p. 11)

There has been an ongoing debate about formative assessment and as Burner (2016) observes, there are a lot of contradictions and there is a need for mutual understanding between “teachers’ and students’ perceptions and experiences of formative assessment of writing” (p. 641). Burner notes that, while students welcome recurrent writing practice and helpful teacher feedback, they find negative feedback and lack of support from the teacher challenging. Despite teachers having limited time, it is important for them to devote some of this to greater interaction with the students during the different stages of writing. “They need to talk more about AFL, and thus develop a shared language”, Burner (2016, p. 641) explains. It is also suggested that some explicit teaching about formative concepts is needed so that students can learn this specific behaviour in the classroom, as DeLuca et al., (2018) indicate in their study. They point out the fact that students value their teachers’ feedback; however, they do not value their peers’ comments and when they do not like them they look for those of the teacher. The differences in grades are sometimes very significant.

An alternative would be to provide their peers feedback with no mark, which is a recommended practice in formative assessment (Hattie and Timperley, 2007; Sadler, 1989; Taylor, 2017). Moreover, Black et al., (2004, p.14) argue that “students may accept criticisms of their work from one another that they would not take seriously if the remarks were offered

by a teacher". Students are likely to be tough on one another, even more so than most teachers (Wiliam, 2017). Webb and Jones (2009) point out that peer assessment is sometimes difficult, because some students would rather work on their own and receive comments from their teacher than from their classmates. As it is indicated in their study, "some students prefer to work individually and to receive feedback from the teacher rather than from other students, but children do adapt well so the major challenge may be for teachers to support them at both individual and group level in making this change" (Webb and Jones, 2009, p. 177). This study focuses on peer assessment in speaking, which is believed to increase the motivation and self-reliance of the students. As Fauzan (2016) has indicated, "debate and peer assessment enable enhancing the students' confidence as well as their motivation" (p. 55).

In their article, Hattie and Timperley (2007) stated that successful feedback must respond to three important questions asked by the pupil or the teacher: "Where am I going (What are the goals?), How am I going? (What progress is being made toward the goal?) and where to next? (What activities need to be undertaken to make better progress?)" (p. 86). The key word is feedback as this has been demonstrated to accelerate learning. Eventually, the focal point is to receive information and act upon it. This feedback must be clear, and it must give students an opportunity to close the gap and demonstrate that they have learnt what they had to learn for a specific task (Panadero and Lipnevich, 2021). Moreover, they also argue that the feedback message should be of good quality, whereby the information provided to the learner is clear and accessible, and such that it can be used by them. Boud (2000, p. 158) writes that "unless students are able to use the feedback to produce improved work through, for example, re-doing the same assignment, neither they nor those giving the feedback will know that it has been effective".

Recent literature (Alben et al., 2019; Lechermeier and Fassnacht, 2018; Panadero and Lipnevich, 2021) has proposed a new model that focuses on the centrality of the student during the feedback process, rather than the feedback itself and its effects on learning. According to Panadero and Lipnevich, students' individual differences must be taken into account if the learner is to be situated in the very centre of the feedback process and it is essential to "provide feedback that is better aligned with learners' needs and that facilitate their effective use of feedback, because in the end what matters is what the learner does with the feedback" (p.14).

3.3. Different purposes of assessment

Assessments are performed for different reasons. Black (1998) distinguishes three "main purposes of assessment, as explained below: 1. Accountability. 2. Certification, progress, and transfer. 3. Support of learning" (p. 24). He defines the three purposes more explicitly as: "1. Summative for accountability to the public, 2. Summative for review, transfer, and certification and 3. Formative to aid learning" (p. 35).

3.3.1. Assessment for accountability

Assessment for accountability relates to the responsibility of educational establishments to the public and government given the financial support received (Black, 1998). The importance of assessment for accountability is undeniable. Accountability, in the broad sense, pertains to the "responsibility for one's actions to someone or to multiple parties as a result of legal, political, financial, personal or simply morally based ties" (Zumeta, 2011, p. 133). Gordon et al., (2014) advocate a balance between formative assessment and assessment for accountability. Each serves a different but complementary function in education. Whilst this accountability can have an adverse impact, it is necessary for assessment to guarantee that governments' educational departments can compare their education system with others. This information is key and without it, national education systems would not have the solid foundation needed to make evidence-based decisions and to prepare students for their futures in the global labour market (Altbach et al., 2019).

An example of this type of Assessment is the Programme for International Students Assessment (PISA). This assesses to what extent students in their last year of secondary education in the Organization for Economic Co-operation and Development (OECD)[6] countries have acquired the knowledge and abilities necessary in today's society. It is a comparative study of school systems, which assesses the knowledge and skills in terms of what skills are likely to be essential for the future. Rieckmann (2017) points out that PISA does not assess understanding of literacy or more holistic educational objectives (as defined by the

[6] https://www.oecd.org/pisa/35070367.pdf

United Nations Education, Scientific and Cultural Organization), but rather, focusses on maths, science and reading. It is therefore judged that PISA is a limited measure of educational success. In addition, according to Steward (2013):

> Any attempt to achieve a fair comparison between more than 50 different education systems operating in a huge variety of cultures, that allows them to be accurately ranked on simple common measures was always going to be enormously difficult to deliver in a way that everyone agrees. (p. 2)

Steward (2013) adds that even though culturally biased questions may have been omitted, questions in different countries can still vary.

3.3.2. Assessment for certification, progress and transfer

Assessment for certification, progress, and transfer needs to be understood in both the institutional and individual sphere. In order to be of use for further education or employment, programmes and qualifications must be certified and recognised by accreditation bodies (Altbach et al., 2019). On an institutional level, certification is crucial, because it serves as a recognition by the accreditation body that a qualification meets its requirements. Individually, certification is required to endorse ability of specific skills and knowledge. This can then be used by the student to enter the next grade, level of learning or employment.

There are tests that can tell us if students are ready to pursue academic studies with score-based information, which certifies the achievement of a particular level of English language. There are a considerable number of examples of these, for example, the Cambridge Examinations for Speakers of Other Languages (ESOL), International English Language Testing System (IELTS) and Test of English as a foreign language (TOEFL). These are called high-stakes tests or, as Davies (2010) suggests, proficiency tests, the content of which is not based on that of the languages courses at their schools that the students sitting those tests might have followed. They are usually taught at language schools or in private schools, in which not

only may they have a considerable impact on the way students are taught and on the way they learn, for they are also used to judge teacher and school effectiveness (Brown and Gao, 2015).

In Spain, ESOL is the most common of these tests[7], which can be taken in 120 schools and over 8,000 preparation centres. Approximately four million candidates in over 130 countries sit the Cambridge tests, which celebrated their hundredth anniversary in 2013. Of the four million, 250,000 are Spanish. As Cruz (2016) points out, students sit these tests because they result in a certificate, which most schools and universities require, and for which a great number of credits are obtained at university. Therefore, obtaining a high score can have significant benefits, not least the pride felt by both students and parents, as well as the advantage of potential increased employability and mobility (Altbach et al., 2019; Cruz, 2016). Certification is often required for job applications. In addition, when students move to a different country or education level, this often requires a transfer of credits and recognition of previous learning to the new country or educational institution (Garnett and Cavaye, 2015). In 2011, nine Andalusian universities signed a Convenio de Colaboración (2011). This agreement set the standards for regulating acknowledgement of foreign language levels within Andalucía. The minimum B1 level had already been agreed in 2010 by the General Direction of Universities of Andalusia (Dirección General de Universidades Andaluzas).

There is another kind of assessment that also informs decisions about whether or not students are ready for further education, such as the SUEE, General Certificate of Secondary Education (GCSE), A Levels and Diploma Programme (DP).The purpose of these assessments, as well as end of the unit or end of the course assessments undertaken at schools, known as summative assessments, is accountability (Brown and Gao, 2015). But do these assessments have an impact on learning? How should learning be defined? These exams sometimes have a damaging effect on the students' attitude towards learning, triggering feelings of anxiety, boredom, and lack of motivation (Tsagari, 2011). Moreover, these examinations have often been criticised for encouraging memorisation (Ramsden, 1992) and it is possible that there is not much connection between these assessments and actual learning, since these have more to

7 https://www.cambridgeenglish.org/es/news/view/mas-de-7-millones-de-personas-realizaron-examenes-de-cambridge-english-en-2019/

do with grading (Bloom, 1956). Desforges (1989) states that, "rather than looking at one school in comparison with another, the key question to ask is whether the school is making the most of the pupils it gets" (p. 120). Furthermore, there is a strong body of opinion (MacDonald and Boud, 2003; Sadler, 1989; Stobart and Gipps, 1997) that standards can be raised through the use of formative assessment and significant gains can be achieved (Black, 2005; James and Gipps, 1998).

3.3.3. Assessment to support learning

Early examples of research into assessment already claimed that there was an excess of summative assessment (Black and Wiliam, 1998b, Crooks, 1988). As Stiggings (2005) points out, it was believed that the pressure to get high marks in tests would motivate students to make a greater effort, with him stating that "assessment has served as the great intimidator. Pressure to get high test scores and good grades, it was believed, would motivate greater effort and thus more learning" (p. 324). Language learning practice was dominated by the use of these standardised language tests, as Dung and Dieu Ha (2019) observe: "For a long time, the use of standardized language test has dominated the language assessing practice"(p. 1). Some researchers, such as Broadfoot and Black (2004), in the light of the importance of assessment as a policy tool, started to identify some weaknesses in these assumptions since policies also encouraged 'lifelong learning' and this seemed to be incompatible with the use of high stakes testing to raise 'standards':

> Tensions have re-emerged recently in the more subtle guise of the apparent incompatibility between policies which lead to an increasingly 'test-driven' educational culture in many parts of the world alongside an explicit policy commitment to encouraging 'lifelong learning'. (Broadfoot and Black, 2004, p. 11)

An objective of the modern languages Advanced Subsidiary (AS) level and Advanced (A) level in the United Kingdom (UK) is "to equip [students] with transferable skills such as autonomy, resourcefulness, creativity, critical thinking, and linguistic, cultural and cognitive

flexibility that will enable them to proceed to further study or to employment" (Department for Education, 2015, p. 3). The Organic Law 3/2020 that modifies the Organic Law 2/2006 (LOMLOE) also aims to "Desarrollar el espíritu emprendedor y la confianza en sí mismo, la participación, el sentido crítico, la iniciativa personal y la capacidad para aprender a aprender, planificar, tomar decisiones y asumir responsabilidades" (p. 66). This view has been reiterated by a considerable number of authors (Baird et al., 2017; Fisher and Frey, 2007; Jones, 2014; Zhao et al., 2016). They state that it would be beneficial for students to take part in their own assessment and engage in peer assessment to expand their understanding and self-regulating skills, such as resourcefulness or taking responsibility for their own learning (Clark, 2012a; Claxton, 1995).

In the International Baccalaureate (IB) programme, one of the five approaches to learning includes developing self-management skills and teachers are required to undertake differentiated teaching informed by assessment (Boud, 2000; International Baccalaureate Organization, 2018a). This way of learning is valued by students because it helps to develop their lifelong learning skills, necessary for future professional careers, and enhances their understanding of the assessment process, according to Malan and Stegmann (2018).

It is necessary to discuss in this thesis the terms formative assessment and Assessment for Learning (AFL), which some authors (Black and Wiliam, 2003; Stiggings, 2005; Swaffield, 2011) construe as rather different. AFL makes clear the importance of developing student self-regulation and autonomy practising assessment in an informal way, as part of a constant process and everyday practice (Brookhart, 2007; Cizek, 2010; Jones and Wiliam, 2008; Klenowski, 2009; Popham, 2008). In contrast, formative assessment is almost any kind of assessment (Swaffield, 2011), which provides some feedback on performance (Taras, 2008) and which is usually in preparation for summative assessment (Stobart, 2008).

Gordon et al., (2014) support a balance between both types of assessment, formative and summative. Despite being distinct, these different forms of assessment are nonetheless complementary. The link between formative, which requires modified teaching and learning, and summative, focussed on monitoring improvement and progress, is not fully clear (Harlen and James, 1997; Stobart, 2008). Criticism of the fixation with assessment for learning, rather

than providing measurement, provoked Broadfoot (1996, p. 46) to ask, “to measure or to learn: that is the question”.

However, in education, which is fundamentally focussed on learning, improving students’ understanding and development is a priority. The design of assessment is not necessarily the most important aspect, but rather, how the assessor uses the outcome of the test (Brown, 2004; Popham, 2008). A formative judgement is being made when the assessor uses the test outcome to support the student. They may do this by pointing out ways in which to progress and explaining what is required to achieve a higher grade (Black et al., 2003; Popham, 2008). By using information from summative assessments to provide feedback to students, they can be used formatively (Carless, 2012). According to Rowntree (1987), a helpful way to understand assessment as either formative or summative is by identifying its objectives, rather than its form.

Data from UK primary schools have demonstrated that, although desirable, formative assessment is not necessarily straight-forward for teachers to conduct (Hall and Burke, 2007; Torrance and Pryor, 2001). Studies on secondary teachers identified four features of formative assessment that were successfully applied and resulted in students’ learning (Black et al., 2003), these being: “questioning; feedback through marking; peer and self-assessment by students; and the formative use of summative tests” (Black et al., 2003, p. 31). These activities are embedded in the following framework (Table 1) developed by Black and Wiliam.

	Where is the learner going	Where the learner is right now	How to get there
Teacher	Clarifying learning intentions and criteria for success*	Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding*	Provide feedback that moves learners forward
Peer	Understanding and sharing learning	Activating students as instructional resources for one another	

	intentions and criteria for success	
Learner	Understanding learning intentions and criteria for success	Activating students as the owners of their own learning*

* Indicates five key strategies in FA

Table 2: Aspects of formative assessment (Black and Wiliam, 2009, p. 5)

In Table 2, the teacher begins by sharing their learning criteria, aims and intentions with the students. This helps to promote the students' autonomy, requiring them to take on some of the responsibility for learning. Throughout the lesson, students are encouraged to question, provide feedback, and undertake peer assessment as well as self-assessment. Feedback is centred around improvement, rather than on student performance. In order to perform peer and self-assessment, students are provided with rubrics and asked to identify priorities for review. The teacher allows more time for the students to answer her questions, which require a deeper understanding. By actively involving the students and asking effective questions, i.e.,'Socratic questioning' (Magee, 2000, p. 16), the students are encouraged to review their thinking and may reach contrasting conclusions. This kind of interaction is central to AFL (Alexander, 2006).

However, Bennett (2011) claims the methodology and approaches of formative assessment must be conceptualised to get the maximum benefit from it. Furthermore, other researchers found that by carrying out formative assessment regularly in their lessons, over time, teachers were able to understand how to use it better, thereby being able to engage students more in their learning and increasing their achievement (Brookhart et al., 2009; Washington-Clark, 2019). It is necessary for teachers to engage in professional development to ensure they have a good understanding of this type of assessment (Chang et al., 2012; Greenstein, 2010; Popham, 2003; Wilson-Thompson, 2005).

There is another kind of assessment for progress called ipsative, which was defined by Cattell (1944) as "a convenient term for designing scale units relative to other measurements on the person himself" (p. 294). This assessment is believed to be highly motivating for

students, especially for those who start from a disadvantaged position, as they can compare their current and previous performance. According to Hughes (2014, p. 6), because ipsative assessment is "longitudinal and focused on learners, it is a powerful approach that allows those who begin from a disadvantage position due to recognition and reward progress, and not only achievement". As there is no comparison with external standards, and it is rarely part of qualifications, students are rewarded for progress, not only for achievement. This is compatible with Assessment for Learning since students become less dependent on the teacher as they better understand the assessment process.

3.4. A focus on formative assessment in languages teaching at school

In the last twenty years there has been a shift from teacher-centred to more student- centred methodologies in language teaching (Araujo, 2020). There is a need for students to control their own learning "in pursuit of the lifelong learning competencies that are essential in the new economy" (Clark, 2012a, p. 24). The future professionals need to be trained in all the skills needed for lifelong learning, and formative assessment can help students to develop some of those skills through practices, such as self-evaluation and peer-evaluation (Araujo, 2020; Black and Wiliam, 1998a; Boud, 2000; Boud and Soler, 2016). Boud and Soler's (2016) research on higher education also suggests that it is necessary to be able to judge the quality of your own work as well as that of others for learning throughout life. Through judging one's own work and that of others in different assessment activities, students can become better learners in and out of school, according to Ajjawai et al., (2020). The term evaluative judgement comes from Sadler (1989), and it has been taken up recently in education as a cognitive ability needed for life-long learning. In Spain, this is identified by Panadero (2017) and Sanmartí (2020), among others, who suggest that students need to build their own personal systems in order to learn and improve progressively, with the aim of becoming independent learners. She has observed that:

> Promover que el alumnado sepa autoevaluarse tiene, entonces, la finalidad de construir conocimientos clave de manera significativa, es decir, competencial, en vez de

> memorística, que le posibiliten continuar aprendiendo a lo largo de la vida y en espacios diferentes de los escolares" (Sanmartí, 2020, p. 3).

These skills can be obtained through the feedback the students use to make changes to regulate their own learning and are called self-regulating skills (Wiliam, 2011). But how is this related to formative assessment in languages teaching? The last decade has seen a rise in literature and practice on how formative assessment practices have an effect on self-regulated learning (Allal, 2016; Panadero et al., 2018), with some empirical evidence that self-assessment and some particular feedback given by teachers can influence self-regulated learning (Brown and Harris, 2013; Panadero et al., 2017). Further research is needed on how formative practices can influence self-regulated learning, especially in languages teaching. Whilst there has been some research on formative assessment in EFL (Araujo, 2020; Baran-Lucarz, 2019; Barbosa and Beserra, 2015; Burner, 2016; Cheng et al., 2004; Dung and Ha, 2019; Ketabi and Ketabi, 2014; Ortega and Minchala, 2017; Rea-Dickins and Gardner, 2000; Taylor, 2017), further investigation is needed, especially on peer assessment in the IB context.

As mentioned above, formative assessment is closely related to feedback, which needs to be meaningful and incentivise students to progress and improve (Radford, 2014; Taylor, 2017; Wiliam, 2011). According to Black (2003), it is not the amount of feedback that matters, but the nature of it. Black and Jones (2006) observe that mindful comments every two or three weeks can be more convenient than a more frequent mark on every activity. Hattie asserts that it is not good practice to provide students a comment accompanied by a mark, because they are likely to read just the mark and ignore the comments (Hattie, 2012).

What seems to be motivating for students, according to Wiliam (2011), is to give them feedback on their performance with an explanation and specific activities to make them understand what they have done wrong in order to improve. Black and Jones (2006) indicate how, for example, language transformation, whilst being challenging, can be good practice to make students think and use what they know in other contexts. Moreover, as pointed out by Hattie and Timperley (2007), students need to be given some time to work on the feedback received.

Xiao and Yang (2019) observed in their study in English language learning that formative assessment can foster the ability of students to self-regulate, allowing them to take responsibility of their own learning through self-assessment, setting their own learning goals and using feedback to take action for improvement. This has been supported by other researchers (Lam, 2015; Purpura, 2016; Salaberri, 1995; Tseng et al., 2015). Therefore, as noted by Ortega and Minchala (2017), in order to assess students authentically it is important that EFL/ESL teachers use formative assessment in the classroom, in particular to collect evidence of their students' knowledge and abilities. Teachers can also use these assessments to modify their own teaching, based on the requirements of the students (Dung et al., 2019).

Peer assessment (PA) is "an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal status learners", according to Topping (2009, p. 20). Through PA, students understand that they can take part in everything that happens in the classroom, becoming more responsible for their own learning and the learning of their classmates (Ozogul and Sullivan, 2009). Studies done so far show a very positive attitude from pupils (De Grez et al., 2012) and highlight that PA is an effective way of promoting the learning process (Chen and Warren, 2005; Hattie, 2009), and developing self-regulating skills (Falchikov, 2005) through regular and relevant feedback (Black and Wiliam,1998; Ortega and Minchala, 2017).

4. THE CONTEXT OF EFL ASSESSMENT IN SPAIN

Taking a foreign language is one of the three common subjects that every student must study to access university. According to the data published annually by the Sistema Integrado de Información Universitaria (SIIU, 2017), in the last five years more than 97% of the students chose English as a foreign language.

None of the Autonomous Regions complies with the established guidelines in the curriculum (Real Decreto 1105/ 2014) nor in the Council of Europe. In both documents, the assessment of four skills is required to obtain any of the levels set by the Council of Europe (2001). However, none of the seventeen Autonomous Communities assess the oral production

and only Galicia and Catalonia assess the listening comprehension. This lack of the oral production was mentioned by some authors as well its negative consequences (Erreyes and García, 2017; Palomeque, 2015; Tamayo-Beritán and Fernánez-Terry, 2017). Teachers give more relevance to the contents that will be in the SUEE so the oral skills lose relevance. As Ruiz Lázaro et al., (2021) explain, students access the university with a lower level of English than expected. Moreover, the heterogeneity of the exams among communities does not give the same equality of opportunities to all the students when accessing to the Spanish Universities and so, a common exam should be created for all the seventeen Autonomous Communities (Ruiz Lázaro et al., 2021).

4.1. The neglect of oral skills in Spain

The recommendations made by the Common European Framework of Reference for Languages (Council of Europe, 2001) point out that MFL should focus on five skills (listening, speaking, talking, reading, and writing). Many classes will involve a combination of these skills. In an English class at school, for instance, a learner may be required to listen to a teacher's explanation, read a textbook, silently or out loud, interact with fellow pupils in a group or project work, write exercises or an essay, and even to mediate, whether as an educational activity or in order to assist another pupil (Council of Europe, 2001, p. 57). However, the academic literature has revealed considerable neglect of oral skills in EFL classrooms in Spain in comparison to the other skills (Alonso, 2014; Mur-Dueñas et al., 2013). As a result, and according to Chela (2008), many EFL pupils are better at writing and reading than maintaining a conversation.

In 2002, the European Council met in Barcelona for the second time on 15th and 16th of March in its yearly meeting schedule to discuss the environmental, economic, and social situation in the European Union. In this meeting, they called for further action "to improve the mastery of basic skills, in particular by teaching at least two foreign languages from a very early age" (Presidency conclusions, Barcelona European Council, 2002, p. 19). Spain is one of the countries embracing this policy and where EFL is taught at the earliest age. According to the data on the report by Eurydice about teaching languages at school in Europe, in 2014, in

Spain, almost all students studied at least one foreign language in primary education, but in 2016 the amount of instruction time dedicated to foreign languages was still small compared to the time devoted to the whole primary curriculum (Education and European Commission/EACEA/ Eurydice, 2017). Moreover, Spain has been one of the first European countries to develop Content and Language Integrated Learning (CLIL) programmes, which are also offered in some other countries in Europe. In this programme, subjects, such as maths and the sciences are taught in English (Caraker, 2016). CLIL is defined as "a teaching method which provides additional teaching in foreign languages without increasing the overall instruction time or taking away lessons from other curriculum subjects" (European Commission/EACEA/Eurydice, 2017). CLIL was introduced in Europe in the 1999s, as part of an exercise in re-thinking how languages are learnt. It was promoted in Finland as a form of open-source framework. According to Marsh (2002) "CLIL is any dual-focused educational context in which an additional language, thus not usually the first foreign language of the learners involved, is used as a medium in the teaching and learning of non-language context" (Marsh, 2002, p. 5). However, as observed by Pavón and Ellison (2013), content and subject teachers must work together to "mutually support and learn *from* and *with* each other" (p. 76). For instance, the content teacher may learn how to prepare communicative tasks for students and the language teacher could learn about what students are expected to learn in classes given in their mother tongue. Every teacher who has to be part of a CLIL programme must accept that they will have to change their practices and grow professionally.

In the last few years, there have been changes in the Spanish education system that have given more importance to oral skills, as the official State document (Real Decreto 1105/2014 26 December) asserted on the 3rd of January 2015 (in relation to the learning of foreign languages:

> La lengua castellana o la lengua cooficial solo se utilizarán como apoyo en el proceso de aprendizaje de la lengua extranjera. Se priorizarán la comprensión y la expresión oral. (Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria y del Bachillerato)

An increasing number of teachers have been adopting a more communicative method of teaching English, with students being more exposed to the language inside and outside the classroom, according to Plo and Pérez-Llantada (2010). However, most teachers still follow a very traditional approach to EFL teaching, devoting most of their time to grammar and only a small amount of time to oral skills (Calvo Benzies, 2016). Nevertheless, schools in Andalucía have been hiring English teaching assistants to help students improve their oral skills in the classroom (Plo et al, 2014). García Laborda et al., (2018) suggest that the fact that the SUEE still does not include an oral component and has had few revisions during the last twenty years pressurises teachers to focus primarily on writing, reading and grammar, thus discouraging them from paying due attention to oral skills. According to García Laborda (2012), "there is a dissociation between Spanish teaching methods and models and the test construct" (p. 18).

A study carried out by Calvo Benzies (2016) at preuniversity, and university levels suggests that Spanish EFL students think that oral skills are undervalued, and that teaching materials and assessments focus primarily on grammar. One of the students in Calvo's study gave the following answer when asked about assessment:

> I believe that the skills that were truly assessed were, apart from the proper contents within each subject, writing and grammar since (...) nearly all of the exams or projects were written-based, and this did not leave any room for testing the rest of the skills. Moreover, (...) on the occasions in which we did talk, as was the case with oral presentations, our way of speaking was not normally assessed (fluency, pronunciation, intonation, communication strategies...); the only aspect assessed was the actual content of the presentations and (...) the written versions of these oral presentations. (Calvo Benzies, 2016, p. 11)

There are 17 autonomous regions in Spain operating under the same Organic Law 8/2013 (Ley Orgánica 8/2013, de 9 diciembre, para la mejora de la calidad educativa), and each community adapts the law through decrees to suit the requirements of their region. The SUEE exams are also different in each community (Caretaker, 2016). The new law (Ley orgánica

3/2020, de 29 diciembre, de modificación de la ley orgánica 2/2006, de 3 mayo (LOMLOE), will be implemented in the 1st Bachillerato in the year 2022-23, hence students will not take the SUEE until 2024.

Despite the reforms, oral skills remain the most overlooked skills in the modern foreign language (MFL) classroom in Spain (García Laborda and Litzler Jerman, 2015). Teachers still devote only a relatively small amount of time to oral skills as they prepare students for the SUEE test. There is also pressure from the students and parents, because the former are more concerned with doing well in the exams than improving their English (Lafuente et al., 2020). This, Plo et al., (2014) infer, could be the reason for the low ranking of Spanish students in MFL competence on international evaluations, and why Spain still has one of the lowest levels of English proficiency in Europe (Carataker, 2016).

Adding an oral skill component to the SUEE would probably increase the amount of time spent practising oral skills in the MFL classroom. Plo and Pérez- Llantada (2010, p. 316) conducted a survey on Baccalaureate English teachers' views on the assessment of oral skills in the SUEE, with most teachers confirming that, "if the SUEE exam concentrates on reading and writing skills, these skills will go on being a priority in the English classroom". They also suggest that either a listening test or an oral interview in the SUEE would "indeed lead to a greater interest in the spoken skills and to devoting more time and effort in practising and acquiring those skills" (Plo and Pérez-Llantada 2010, p. 324). Consequently, if no effort is made to make those tests as close as possible to communicative teaching practices, it is unlikely that students will become more proficient at speaking. In short, if we consider that speaking must be taught in the classroom, it needs to be assessed and this assessment needs to be embedded in the teaching approach and be meaningful to the students (Leathy and Wiliam, 2015). Teachers could practise formative assessment of their students' oral skills on a regular basis for example through interactive tasks. These tasks do not have to be very long and they make all the students in the class talk as they have to selfassess and assess their peers.

4. 2. Overall view of the Spanish University Entrance Examination in Spain (SUEE)

The SUEE is a written test that all students who hope to be accepted into a Spanish University must sit. This examination is part of the Evaluación de Bachillerato para el Acceso a la Universidad (EBAU), also called Pruebas de Acceso a la Universidad (PAU) in some regions.

Before 2010, the SUEE consisted of at least 20 written tests. There were four common subjects that every student had to study as well as some specific subjects, which the students chose depending on what they were going to study at university.

Common subjects:

1. Spanish Language and Literature.
2. Foreign language (the most common was English, but students can also study French, German, Italian and Portuguese).
3. History of Philosophy or History of Spain (They could choose which one to do).
4. Galician, Catalan, Valencian, and Basque were also compulsory in Galicia, Catalonia, Valencia and the Basque Country.

Specific subjects related to the following groups:

1. Arts.
2. Letters.
3. Technical Sciences.
4. Health Sciences.

Since the academic year 2009/2010, the new model to get into a Spanish University has been the following.

General Phase (compulsory)

1. Spanish Language and Literature.

2. Foreign language: Students can choose among: German, French, Italian, English and Portuguese.
3. History of Spain or History of Philosophy (They choose one to do in the SUEE).
4. A subject of the modality of the second year of the Bachillerato and the student can choose from among the following: Mathematics, Mathematics Applied to Social Sciences, Latin and Fundamentals of Art.
5. Co-official Language and Literature in case of Basque, Catalan, Valencian and Galician.

Specific Phase: There are five modalities of the Bachillerato:

1. Arts.
2. Technology.
3. Social Sciences.
4. Health Sciences.
5. Humanities.

The exams are about subjects of the modality the students choose, and they can choose how many exams they do, with a maximum of four and only two will count for their final mark (the ones in which students get the highest mark). The subjects chosen should be related to the degree they want to study for and where they want to do it.

The SUEE (Selectividad) has existed as we know it for 20 years. It was established in 1974 through the law Esteruelas (established by the minister Cruz Esteruelas). The first exams were held in 1975 and consisted of a summary of a conference, a text analysis, and questions of language, mathematics, and other optional subjects. However, as García Laborda (2012) explains, the foreign language was not introduced in the exam until 1989 and since then, this exam has been almost unchanged except for some small variations in every Autonomous Community. Even though some authors have questioned the validity of the exam (Amengual-Pizarro, 2006; Soler, 1999), their studies have not had any impact and have been ignored. Whilst the SUEE should value the four skills, the oral skill is not included in the exam, except in Catalonia and Galicia, where there is a multiple-choice activity to test oral comprehension (Pizarro, 2006). This SUEE has had different names through history: First, Selectividad, then

PAU (Prueba de acceso a la Universidad) and now it is called EvAU (Evaluación para el Acceso a la Universidad) or, in some Autonomous Communities, EBAU (Evaluación de Bachillerato para el Acceso a la Universidad). In 2006, as García Laborda (2006) observes, the foreign language exam had the same elements as it had in 2021:

1. A reading text with comprehension questions.
2. Some grammar questions.
3. A piece of writing.
4. An oral comprehension task (only in Catalonia and Galicia). This was an important inclusion.

4.3. The SUEE in Andalusia

In Andalusia the SUEE is called PEVAU (Pruebas de Evaluación de Bachillerato para el Acceso a la universidad y/o Pruebas de Admisión) (Distrito único andaluz, 2022). The exam consists of two phases: Access Phase and Admission Phase.

1. Access Phase:

- Spanish Language and Literature.
- Foreign language: Students can choose among: German, French, Italian, English and Portuguese.
- History of Spain or History of Philosophy, to choose (They choose one to do in the SUEE).
- A subject of the modality of the second year of the Bachillerato that the student can choose among the following: Mathematics, Mathematics Applied to Social Sciences, Latin and Fundamentals of Art.

2. Admission Phase: In this phase students can do from one to four subjects of the subjects they have studied in the second year of the Bachillerato:

- Core subjects of the modality: Fundamentals of Art II, Latin II, Mathematics Applied to Social Sciences II and Mathematics II, different to the core subjects of the Access Phase.

- Optional Core subjects: Performing Arts, Audiovisual Culture II, Biology, Technical Drawing, Business Economics, Geography, Design, Physics, Greek II, Geology, History of Art, Chemistry.
- Specific compulsory: History of Philosophy.
- A second foreign language (German, French, English, Italian or Portuguese) different to the language they took in the Access Phase.

The foreign language exam has had the same structure almost since its creation (Laborda, 2012). There are three blocks: reading comprehension, use of language and writing. There has never been any assessment of oral skills[8], despite the recommendations made by the Common European Framework of Reference for Languages (Council of Europe, 2001).

5. THE INTERNATIONAL BACCALAUREATE (IB) IN SPAIN

The International Baccalaureate Organization (known as the IB) is a non-profit foundation that offers four challenging world-class education programmes for students aged 3-19. It collaborates with schools and governments from all over the world, including Spain. It is characterised by demanding educational programmes and rigorous assessment methods. It is a very prestigious programme, which was not well recognised in Spain until very recently. However, it has become more and more popular and in Spain and Andalusia eight high schools have been internationalised. The purpose is to improve the quality of teaching and learning and to contribute to the development of people, communities, institutions, nations and the world in general (Resolución del 18 de febrero de 2022). According to the Order of January 15, 2021, "La evaluación se convierte, en sí misma, en un proceso educativo que considera al alumnado

8 https://www.juntadeandalucia.es/economiaconocimientoempresasyuniversidad/sguit/examanes_anios_anteriores/selectividad/sel_2020-2021-Orientaciones_ingles.pdf

como centro y protagonista de su propia evolución. El carácter formativo de la evaluación puede contribuir al desarrollo de los centros" (p. 225).

5.1. Origin and history of the IB

The IB was founded in Geneva in 1968 to develop inquiring, knowledgeable and caring young people, who would help to create a better and more peaceful world through intercultural understanding and respect (see Appendix 2). In the 1960s, some progressive educational trends started to become popular (see Appendix 3). Some of the key influential educationalists were John Dewey, A.S Neill, Jean Piaget and Jerome Bruner (see Appendix 4), whilst some early IB influencers were Robert Leach, John Goormaghtigh, Alec Peterson and Kurt Hahn (see Appendix 5).

The Primary Years Programme (PYP) concerns students aged 3-12. It focuses on helping students to become inquirers in and out of the classroom. Pupils study six transdisciplinary, very important global themes, in six subjects: language, mathematics, social studies, arts and personal, social and physical education[9]. The learners participate in their own learning and in that of others, appreciating the benefits of working with others and developing an international mindset.

The Middle Years Programme (MYP) concerns students aged 11-16. It focuses on encouraging pupils to provide connections between what they study and the real world. Pupils study eight subject groups: language acquisition, language and literature, individuals and societies, sciences, mathematics, arts, physical and health education and design. All the teaching and learning happens in context. Students have to complete a personal project about a topic they want to learn about. They also have to complete some hours of creativity, action and service (CAS)[10].

[9] https://www.ibo.org/globalassets/digital-toolkit/brochures/pyp-programme-brochure-en.pdf

[10] https://www.ibo.org/programmes/middle-years-programme/curriculum/

The Career-related Programme (CP) concerns students aged 16-19. It focuses on giving students who want to pursue a career-related education an international education with the IB values. Students take at least two IB DP courses, which provide them with the theory they need. The CP study provides the students with real world approaches to learning and the CP core provides them with the skills they will need to become lifelong learners.[11]

The first IB school was founded in 1962 and the 'International Baccalaureate' name was given to it. In 1968, the students took their first IB Diploma Programme exams. The IB programmes were recognised in many countries. A summary of its history can be found below.

Year	Event
1962	*Atlantic College founded International Schools Association (ISA) conference of teachers of social studies recommends that International Passport to Higher Education to be called "International Baccalaureate", following Bob Leach report*
1968	*IB Organization registered in Geneva*
1970	*First official IB Diploma Programme exams taken by students from 12 schools from 10 countries- including two state schools (Germany and France)*
1970-80	*Recognition of the IB programmes negotiated in many countries*
1975	*North America Regional Office opens in New York*
1976	*First inter-governmental IB Conference, The Hague; Piet Gathier*
1977	*Heads Standing Conference, Palais des Nations, Geneva*
1978	*Standing Conference of Governments London; 32 countries represented*
1982	*Regional offices open in Buenos Aires, London and Singapore*
1983	*IB Diploma programme becomes available in Spanish*
1994	*Middle Years Programme introduced*
1997	*Primary Years Programme introduced*
1998	*IB develops mission statement*
2006	*IB develops Learner Profile*
2010	*Opening of Washington D.C Global centre*
2011	*Opening of The Hague Global Centre*
2012	*Career-related Programme (CP) introduced and opening of Singapore Global Centre*
2016	*Opening of Cardiff IB Global Centre*
Adapted from IB (2011)	

[11] https://www.ibo.org/programmes/career-related-programme/

In 1968, the first programme offered by the IB, the Diploma Programme, was established. It sought to provide a challenging, yet balanced, education that would facilitate geographic and cultural mobility by providing an internationally recognised university entrance qualification that would also serve the deeper purpose of promoting intercultural understanding and respect (International Baccalaureate Organization, 2019, p. 1). This programme is known for its rigorous assessment, which gives IB students access to the most important universities in the world.[12] Currently there are 5,400 schools all over the world, of which 174 are in Spain and just 3 in Almería [13].Within the DP, students are expected to complete both internal and external assessments. The latter are marked by IB examiners, while the former are marked by teachers and later 'externally moderated by the IB' (International Baccalaureate Organization, 2018a, p. 27). All schools that teach the Diploma must be authorised first, which means that they take part in a rigorous evaluation process to ensure the quality of the teaching and learning. Hence, the Diploma is classified as a high stakes curriculum, which is an important concept in the field of assessment and refers to 'testing for tracking, promotion and graduation' (Huebert and Hauser, 1999, p. 3).

In Spain, the first school to teach the DP was the International School SEK El Castillo in 1977 and the International School SEK-Alborán started to do so in 2007. The DP was recognised, but there was the limitation that students still had to complete the selectividad exams (Universidad Nacional de Educación a distancia (UNED) version for foreigners). Also, the equivalence of grades was not very favourable compared to the selectividad. After 2010, the new arrangements for DP students improved significantly the access to Spanish Universities as the DP grades were used directly (no selectividad exams needed), with much fairer equivalence of grades. A large number of private DP schools are now offering Baccalaureate students a choice of IB DP or Spanish Baccalaureate, thus removing the combined DP/Spanish Baccalaureate.

[12] https://ibo.org: The IB is an organization which works with schools, governments and international organizations to develop challenging programmes of international education and rigorous assessment.

[13] https://www.ibo.org/about-the-ib/facts-and-figures/

The negotiated validation of the IB marks has improved, but not as much as expected. As a result, most DP students still do both DP and SUEE exams (Resnik, 2019), and in many schools students do a combined IB programme with the national tests (Thoilliez and Rappoport, 2018). This is really challenging for the students as well as their teachers, especially for those who need a very high mark to access university.

In 1994, the Middle Years Programme (MYP) was introduced and in 1997, the Primary Years Programme (PYP). With all three programmes: PYP, MYP and Diploma, there was a continuum of international education for students aged 3 to 19. In 2012, the continuum was enriched with the introduction of the Career-related Programme (CP) for students from 16 to 19 years old. Schools can have all four programmes or they can implement them independently. "The aim of all IB programmes is to develop internationally minded people who, recognizing their common humanity and shared guardianship of the planet, help to create a more peaceful world" (International Baccalureate Organization, 2019, p. 5). International mindedness is fostered:

- by helping pupils consider not only their own perspective, but also, other people's perspectives by engaging with different beliefs
- through multilingualism as students have to study more than one language in the IB
- by engaging globally and doing service for their own community

5.2. The IB learner profile

IB learners need to strive to be inquirers, knowledgeable, thinkers, communicators, principled, open-minded, caring, risk-takers, balanced and reflective. The IB proponents believe that these attributes can help individuals and groups become responsible members of local, national and global communities (International Baccalaureate Organization, 2019, p. 5).

In the public schools in Andalusia, the DP is studied simultaneously with the National Baccalaureate, giving the students the opportunity to achieve a double title: The IB and the Spanish Baccalaureate titles. Two titles give students more options to continue their higher

studies in the best universities of the world. Moreover, the students obtain a rigorous and balanced preparation. They are taught to think in a critical way, to understand other cultures and experiences and to apply what they learn to different contexts and disciplines. The programme has been taught for at least 40 years in some private schools and from the year 2020-2021, students could study the Diploma in six public schools in Andalusia[14]. This programme is very rigorous and demanding. It uses a creative and cooperative methodology and students must have a high motivation as it implies longer hours at school, with a high commitment to working hard and meeting deadlines.

5.3. The IB teacher's methodology

The IB offers programmes in a wide variety of schools all over the world. The curriculum is reviewed regularly so that IB students receive the best possible education. Teachers are an important part in developing the programmes. Teachers need to centre on the learner, develop the approaches to teaching effectively and to support students as they develop the approaches to learning. Teachers must teach students to become lifelong learners, who can work independently as well as in collaboration with others.[15]

The approaches to teaching in the IB programmes are six. However, teachers are allowed to use their own strategies depending on their own students' needs. The approaches are based on inquiry, focused on conceptual understanding, developed in local and global contexts, focused on effective teamwork and collaboration, designed to remove barriers to learning and informed by assessment. The approaches to learning are based on the belief that students need to learn how to learn. The IB works in the development of five skills that play an important part in the developing of the confidence students "need to thrive and make a lasting difference: thinking skills, research skills, communication skills, social skills and self-management skills".[16]

[14]https://www.juntadeandalucia.es/organismos/educacionydeporte/areas/ensenanzas-escolarizacion/bachillerato/paginas/bachillerato-internacional.html

[15] https://www.ibo.org/benefits/the-ib-teaching-style/

[16] https://www.ibo.org/benefits/

5.4. Programmes and curriculum of the Diploma

The IB educational system is fully aligned with that of the National system in Spain at all stages. The PYP is for students aged 3-12, (Primary) the MYP for those 11-16 (Secondary) and the DP for those 16-19 (Bachillerato). This study is focused on the DP, as a pre-university programme and the most popular one. It comprises six subject groups, which must be studied over two years. The groups are:

1. Language and Literature.
2. Language Acquisition.
3. Individual and Societies.
4. Sciences.
5. Mathematics.
6. Arts.

Students have to choose a subject of each of the first five groups and either a subject of the group 6 of Arts or a second subject of the groups 1-5. Every student must study a minimum of three subjects (and a maximum of four) at higher level and the rest at standard level[17].

1. Three subjects at advanced level (240 hours). Students are required to obtain more knowledge and comprehension of the subject as well as more abilities.
2. Three at standard level (150 hours).
3. Theory of knowledge essay to develop critical thinking reflecting about the nature of knowledge and the way in which we know what we claim we know.
4. 4,000-word extended essay.
5. A project related to Creativity, Activity and Service.

[17] https://www.ibo.org/es/programmes/diploma-programme/curriculum/

5.5. Assessment in the Diploma

At the end of the programme, students have to sit written external exams marked by IB examiners (the researcher of this thesis is an examiner of English B higher level). They also do some internal assessments (works that are, firstly, marked by their own teachers and then, moderated by IB examiners). Subjects are marked from 1-7 and students can obtain three more points for Theory of Knowledge and the Extended Essay.

The Diploma is achieved by those students who obtain at least 24 points, if they have complied with all the IB requirements, and they have participated in the component of Creativity, Activity and Service. The highest mark students can obtain is 45. The assessment is criterion based. One of the most interesting things about the Diploma is the wide variety of subjects that it offers[18], although not every school offers all the subjects. SEK-Alborán school was a pioneer in offering the three programmes of the IB in Almería. They started with PYP and MYP in 1999 and introduced the Diploma in 2007. Initially, it was integrated with the national curriculum but since 2020 students only study for the Diploma.

6. ISSUES AND PROBLEMS OF FORMATIVE ASSESSMENT IN THE LANGUAGE CLASS

Formative assessment refers to assessment that evaluates students' work, arriving at decisions aimed at maximizing the effectiveness of the teaching and learning process through the provision of constant and timely feedback. The specific context and that of the individual student are recognised, but no grade is conferred (Hamodi et al., 2017). Some of the problems of using formative assessment in the classroom in Spain are, firstly, the lack of clear guidelines that language teachers have (Mur-Dueñas and Hornero, 2013). As a consequence, teachers give different weightings to the different skills. Secondly, the oral assessment is still overlooked due to the fact that these are not assessed in the SUEE, as a result, despite the fact that the use of

[18] https://www.ibo.org/es/university-admission/support-students-transition-to-higher-education/course-selection-guidance/

formative assessment has been demonstrated to help the necessary skills for their future professional practice, some teachers use formative assessment only occasionally (Palacios and López, 2013). Finally, teachers are trained at the university and later there is continuing teacher training development, which helps them to implement all the changes in the subsequent different assessment laws (Hijano Trujillo, 2021). Teachers are trained in "Centros de Profesorado (CEP)", where all their needs are met aligned with the strategic lines marked out annually by the Education Department. There is a high percentage of teacher participation in the formative activities carried out in the CEP, as stated by Alfageme-González and Nieto Cano (2017). However, these activities are sometimes irrelevant and too general and do not consider the particular contexts (Escudero and Portela, 2015). As a result, it is up to teachers to bring what they have learnt to their classrooms if they do not have the opportunity to choose what exactly they want to be trained in and if they are not going to have additional practice and follow up in their work place (Alfageme González and Nieto Cano, 2017).

The the only way to galvanise change in teachers' practices is making them work together for a long period to solve the common problems identified in their lessons. In short, teachers value the activities in CEP, but they do not think that what they learn is applied in their practices and there is no follow up. Hence, it is what teachers do in the classroom that can have the most direct impact on the students' learning (Robinson and Timperley, 2007).

6.1. Lack of homogeneous guidelines in Spain for the assessment of oral skills

In the UK, where there has been a considerable amount of research on Assessment for Learning (Black and Wiliam 1998a; Jones and Wiliam 2008; Leathy and Wiliam, 2015) and school accountability assessment (Bradbury, 2014; Brown and Harris, 2009), most schools expect lesson plans to include some kind of formative assessment (Black and Jones, 2006; the education inspection framework (Ofsted, 2019). However, in Spain research is scarce on formative assessment and teachers lack clear guidelines about what type of assessment would be the most suitable for developing new approaches in their classrooms. Consequently, according to Mur-Dueñas, Plo and Hornero (2013), every teacher or MFL department makes their own decisions about the assessment practices they want to perform in their schools to

make the assessment meaningful for them. Even in the same region teachers assess their students differently across different schools.

Fernández- Sanjurjo et al., (2019) found very varied practice amongst teachers. Some teachers created their own rubrics for summative assessments and provided the students with feedback in an effort to be as objective and fair as possible. Panadero and Brown (2017) found that some teachers occasionally used more formative assessment, such as peer-assessment or self-assessment. All teachers were giving different official weighting to the different skills. Research seems to indicate that teachers need to be given the opportunity to develop and share homogeneous guidelines and standards in Spain, especially for oral skills. These remain the most overlooked and are not assessed in the SUEE (Amengual Pizarro, 2010; García Laborda et al., 2018). Given the complexity of the assessment and anxiety of students aiming for high grades, Mur-Dueñas et al., (2013) suggest that as well as being trained in strategies to pass the SUEE, students also need different activities to be included in their lessons to avoid boredom and monotony and in order to achieve the B1 level required by the CEFR.

Finally, Spanish teachers need some common guidelines, which are the same in every community, to provide assessments, particularly of oral skills in the classroom. In short, there is a need for a match between teaching, assessment and policy. In Spain, policy change in EFL assessment has been confusing and contradictory and has lacked the understanding of and participation by teachers. As Jiménez Raya (2007) argues, teachers must understand the need for a change in their teaching. So, asking teachers' opinions and enabling them to participate in the elaboration of new policies would help them to teach and assess their students' differently, thus raising the possibility of improving their students' oral abilities

> (…) people are active by nature and driven to persistently adapt to changes in the environment in order to be successful. Unless teachers retain a sense of agency about why and how they might teach differently, the call for innovation in modern language teaching will likely ring hollow. (Jiménez Raya, 2007, p. 32)

6.2. The need for teacher training in curriculum development

In 1990, the Ley Orgánica de Ordenación General del Sistema Educativo (LOGSE) was implemented, and through a constructivist approach, the promotion of formative assessment was initiated, further promoted by the following law: Ley Orgánica de Calidad de la Educación (LOCE, 2002), which was never implemented. Coll and Martin (2014) observe the importance of these laws introducing for the first-time practices such as peer and self-assessment. However, the last two reform acts, the Ley Orgánica de Educación (LOE, 2006) and the Ley Orgánica para la Mejora de la Educación (LOMCE, 2013) have seemingly promoted more summative practices: "las evaluaciones externas de fin de etapa constituyen una de las principales novedades de la LOMCE con respecto al marco anterior" (Preámbulo Ley Orgánica para la Mejora de la Calidad Educativa, 2013, p. 6). The organic Law 3/2020 that modifies the Organic Law 2/2006 only refers to formative assessment when it talks about secondary education "La evaluacion del proceso de aprendizaje del alumnado de la educación de secundaria obligatoria será continua, formativa e integradora" (p. 71). There is no mention of formative assessment in the primary or bachillerato, which leaves it up to autonomous communicties, schools and teachers how much use of it they want to make. The different autonomous communities organise courses on formative assessment, for example the government of La Rioja (Centro Riojano de Innovación Educativa) organised a course titled " Instrumentos de evaluación formativa y competencial para LOMLOE" in 2011[19].

Despite the innovative nature of the ideas promoted, it seems that a lot of teachers are still trained in implementing the traditional approach to assessment and they do not have any mandatory training in new ways of assessment. As highlighted by López Bertomeo and González Olivares (2018), the LOE followed a constructivist approach that has become marginalised and has left the schools having to choose their own theory in education and teachers continuing to focus on the external exams. Until the SUEE includes an oral exam, neither teachers nor students will take seriously the need to develop oral skills (Ruiz Lázaro et al., 2021).

[19]https://g1.edurioja.org/web/activida/documentos_externos/5485Instrumentos_Eval_Competencial_LOMLOE-Feliu.pdf

In Spain, in the last twenty years, there have been five educational laws (LOGSE, LOCE, LOE, LOMCE and LOMLOE). Professional training for teachers was introduced in 'Teachers' Centres' to help them improve their practices. However, teachers cannot be expected to change their practices every time a new educational law is passed. Teachers usually do not understand why they have to change their current methodology, if it is working in preparing their students for the SUEE exam. Bolivar reveals that the changes in the curricula have not always been accompanied by teacher training. As stated in the World Bank report on secondary education in the world: "There is a profound mis-match between learning needs, competencies, and skills demanded from students in the knowledge society and the teaching skills of secondary teachers after their passage through teacher training colleges and in-service training programs" (Bolivar, 2015, pp. 357-358).

The work of Jiménez Raya (2008) indicates that educational policies must facilitate effective practices, but when they are introduced without giving teachers time and training in order to reflect, understand and develop the new approach, they usually fail.

Hamodi et al., (2017) carried out a study with university students focused on their training to become teachers to ascertain whether having experience of formative assessment during their teacher education could influence their future practice. They found that formative assessment was rarely used. They also observed that significant barriers to change and introduce innovative techniques exist in primary and secondary school education. Methods requiring collaboration with colleagues (e.g., peer assessment) were viewed negatively by the graduates who participated in the study. The authors came to the conclusion that formative assessment should be introduced at university to prospective teachers. Moreover, it should be included in continuing professional development, such that, in future, school teachers find solutions to challenges arising in their schools when seeking to implement these assessment systems.

7. ASSESSING ORAL SKILLS

The use of a foreign language is an action that pupils perform and for which they need communicative competence (CEFR, 2001). The CEFR (2001) differentiates between speaking (spoken production) and talking (spoken interaction). Speaking is when "the language user produces an oral text which is received by an audience of one or more listeners" (CEFR, 2001, p. 58). On the other hand, talking is when the language user acts alternately as speaker and listener, with one or more interlocutors so as to construct conjointly, through the negotiation of meaning following the co-operative principle, conversational discourse" (CEFR, 2001, p. 73).

Some members of the MFL community of practice use speaking to account for talking. This is seen in the work of a great number of researchers, who point out that 'in order to communicate effectively, speaking is one of the most essential skills' (Bahadorfar and Odmivar, 2014; Fauzan, 2016; Leong and Ahmadi, 2017; Ur, 1996; Zyoud, 2016). In addition, students often assess their progress in a language by noting improvements to their speaking abilities (Leong and Ahmadi, 2017). For many, the first impression of a person's language ability is based on the use of productive and receptive oral skills – oral fluency and comprehension (Bahadorfar and Omidvar, 2014). For the purposes of this research, it has been decided to use productive oral skills, which include both speaking and talking. However, sometimes a reference to the varied terminology as used in the various contexts will be made.

In order to develop fluency, oral skills need frequent practice at all stages and levels of learning (Zyoud, 2016). Chrichton (2009) writes that teachers can use real-life situations to weave talking into the classroom. For example, they can ask their students about their holidays, especially if there was something in the written task related to the holiday theme. That is, teachers can respond to the possibilities and stimuli that a classroom provides (van Lier, 1996). Salaberri (1995) also points out that teachers need to take advantage of all the situations and activities in the classroom to generate language.

According to Harmer (1991), students revert to their mother-tongue when speaking about a topic on which they do not hold sufficient knowledge. Moreover, unless teachers firmly

encourage their students to speak in the target language (TL), they will return to their mother-tongue to talk to their classmates (Leong and Ahmadi, 2017). Not only do students need practise in the TL, but also, there needs to be significant exposure to listening to the target language, as Conti and Smith (2019) observe. They point out that we learn to speak through the aural medium. According to Renukadevi (2014), around 45% of language competence is obtained through listening. Moreover, students need to have linguistic knowledge to be able to deal with contingencies (Perrenoud, 1998). They also need to think on their feet and undertake retrieval of language from medium- and longer-term memory.

The IB programme specifies the need for students to develop productive and interactive skills, which are assessed in an individual oral interview. In this oral interview "Students demonstrate the ability to verbally interact in the target language in response to a literary stimulus" (International Baccalaureate Organization, 2018, p.16). Then, there is "A conversation with the teacher, based on an extract from one of the literary works studied in class, followed by discussion based on one or more of the themes from the syllabus" (International Baccalaureate Organization, 2018, p. 30). In short, speaking a foreign language in the IB is a skill students need to learn in order to be able to express ideas, maintain a conversation and communicate more generally.

7.1. Ways to assess oral skills

The Diploma Programme (DP) focuses on criterion-related summative assessment and students need to be recorded by the teacher in the second year of the course of study. However, the IB programme encourages teachers to undertake formative assessment practices throughout the course, as indicated in the Standards and Practices document for authorised schools (International Baccalaureate Organization, 2014) and in the Language B guide (International Baccalaureate Organization, 2018). The latter requires that:

> in the language acquisition classroom, there should be opportunities for both teacher-led formative and summative assessment as well as peer assessment and opportunities for students to reflect their own learning and performance on assessment instruments

> in order to develop their language skill. (International Baccalaureate Organization, 2018, p. 58)

As a result, PA is a common practice to enable communication between students.

As aforementioned, the Spanish Curriculum, despite including oral competence and the need to assess it, does not include an oral exam in the SUEE. Instead, each teacher creates their own criteria and activities to assess oral skills, and, in some cases, does not assess it at all, as discussed in Chapter 1. This means that oral skills are the most overlooked skill in the MFL classroom in Spain, as García Laborda and Litzler Jerman (2015) point out. Teachers devote only a small amount of time to oral skills as they prepare students for the SUEE exam. Students are more concerned about doing well in the exam, rather than improving their English (Lafuente et al., 2020). Plo et al., (2014) suggest that this could be the reason for the low rank of Spanish students with MFL competence in international evaluations (Hornero et al., 2013).

Adding an oral component to the SUEE would probably increase the amount of time spent practising oral skills in the MFL classroom. If no effort is made to ensure the tests mirror as closely as possible communicative teaching practices, it is doubtful whether students can become more proficient at oral communication. In short, if it is a requirement that speaking be taught in the classroom, then it needs to be assessed. This assessment needs to be blended into the teaching approach and be meaningful to the student speakers.

Oral skills could be formatively assessed on a regular basis, for example, through peer assessment. Students can be given rubrics with the criteria, perform the task and then, give feedback to each other. PA techniques can be used throughout a programme, being the most salient issue to ensure that action is taken on peer feedback. Regular routines can be followed, aimed at drawing on the feedback that involve clear structures for students to act upon. Leathy et al., (2005, p. 3) hold that ‘activating learners as instructional resources for one another’ is a key strategy. Students see the benefit in doing peer assessment and have become more aware of what constitutes impactful feedback. If the giving of oral feedback is seen by students to be of a high standard and beneficial, then they will become more willing to engage in the process, as observed by Nicol et al., (2014).

7.2. Assessing productive oral skills in a peer assessment framework

This study focuses on peer assessment in oral skills, which is believed to increase the students' motivation and self-reliance, for as Fauzan (2016) states "debate and peer assessment enhance the students' confidence as well as their motivation" (p. 55). Topping (1998) devotes a section of his review of the PA literature to the assessment of oral presentation skills. He provides evidence that it leads to confidence in speaking, improvements in grades and appraisal skills. Through PA, students understand that they can take part in everything that happens in the classroom, becoming more responsible for their own learning and that of their classmates (Jones and Wiliam, 2008; Ozogul and Sullivan, 2009). Studies done so far show a very positive attitude from pupils (De Grez et al., 2012) and highlight that PA is an effective way of promoting the learning process (Chen and Warren, 2005; Hattie, 2009), and developing self-regulating skills (Falchikov, 2005) through regular and relevant feedback (Black and Wiliam, 1998b). Various studies have assessed the efficacy of feedback and formative assessment to support learning (Coffield et al., 2004; Jones, 2014; Nicol and Macfarlane-Dick, 2006). They agree that formative assessment and feedback can be useful tools to allow students to self-regulate. Self-regulation relates to the extent to which students during learning can regulate elements of their thinking, behaviour, and motivation (Pintrich & Zusho, 2002).

Jones and Wiliam (2008) assert that oral assessment can be undertaken continuously during the lesson, asking students the right questions: at the beginning, to check prior learning; in the middle, to thoroughly secure learning, to check understanding or any misconceptions, with other pupils being potentially used to help those who are struggling; and at the end, to see if the students have achieved the objectives. Some kind of formal assessment is also needed, because "if the students are not assessed formally, they may claim that they have not had the same opportunities to show their competence in oral speaking as their peers" (Bachman and Palmer, 2010, p. 478). Such formal or summative assessment would complement the formative assessment undertaken in the lesson and give important feedback to the teacher for forward planning.

7.3. Peer Assessment in the Diploma

As mentioned previously, students need both practice and assessment in the target language in order to become proficient at speaking. The IB promotes three practices related to PA in its Standards and Practices document. The fifth practice of the IB states: "Teaching and learning supports students to become actively responsible for their own learning" (International Baccalaureate Organization, 2014, p. 25), whilst the eighth practice states: "the school provides opportunities to participate in and reflect on the assessment of their work" (International Baccalaureate Organization, 2014, p. 26). PA is very much related to these two practices by involving students actively in their own learning process (Ozogul and Sullivan, 2007), which, according to Falchikov (2005), is beneficial for developing self-regulating skills as well as lifelong learning skills (Malan and Stegman, 2018). Judging their peers' work (Falchikov, 2005), understanding the assessment process and doing this in a collaborative way will be most likely required in their professional careers (Malan and Stegman, 2018). Boud and Falchikov (2006, p. 403) state that, "graduates in the workforce will not in general be taking examinations or writing academic essays. They will be puzzling over what counts as good work and how they will be able to discern whether they are producing it". Hence, they might need to evaluate work quality in their future workplace. Peer assessment has been evidenced as a tool to improve language learners' self-confidence, accountability, sense of ownership, social abilities as well as their cooperative and collaborative work (Esfandiari & Tavassoli, 2019).

Moreover, PA allows teachers to assess their students more regularly and consequently reduces teachers' workload (Topping, 2003), thus allowing them to focus on other things, such as lesson planning or creative learning. Finally, the fourth practice states: "the school provides students with feedback to inform and improve their learning" (International Baccalaureate Organization, 2014, p. 26). The feedback students provide each other in PA is formative, so it usually fosters the learning process (Hattie, 2009) and it is instant individualised rich feedback (Topping, 2009). Students' oral skills were found to have improved with this practice in the research by Chen and Warren, (2005). Fauzan (2016) asserts that teachers have the possibility to conduct more interactive classes, where students have the confidence and motivation to talk.

7.4. Constraints and opportunities in assessing oral skills

Assessing oral skills is not an easy task because it is usually interactive. No talk is exactly the same as another, and it is necessary to develop the criteria related to every task and not in isolation. Luoma (2004) writes that: "Assessing speaking is challenging because there are so many factors that influence our impression of how well someone can speak a language, and because we expect tests scores to be accurate, just and appropriate for our purpose" (p. 1). As a result, as explained below, teachers face various challenges when assessing oral skills.

7.4.1. Assessing oral skills is time-consuming

Teachers, first, have to define their learning objectives and structure them in a rating form according to Luoma, (2004) choosing a task and specifying the assessment criteria. Then, they have to do the assessment task with a large number of students in a short time listening to them as they talk and asking them the right questions to determine their level (Ahmad et al., 2019). Conducting oral skills assessment has always been time-consuming and is often the reason why teachers choose not to do so. Identifying in advance the students who are going to be assessed and the specific aspect(s) that needs to be tested will help teachers gather all the data required for assessment. If the activity has been recorded, performances will then need to be rated, listening to the recordings word by word. Finally, teachers are often asked to write detailed reports (Luoma, 2004), from which students can obtain the feedback on what they need to change, improve and self-regulate their learning (Nicol and Macfarlane-Dick, 2006). This process is time-consuming.

Luoma (2004) asserts that time can be saved when testing, if teachers have two students talking with each other in the exam. In this way, teachers share their responsibility with students and save time. This is one of the reasons why at present interactive tasks are replacing interviews, according to Ducasse and Brown (2009). Their work on assessment tasks involving peer-to peer interaction reveals that "peer-to peer assessment is typically also more time and cost efficient as candidates are tested together, and raters assess two or more candidates simultaneously" (p. 424). Hence, in PA the peer assessor's feedback can be as reliable as the

teacher's (Topping, 2009). Topping (2009) also says that the assessment of oral skills is not as time-consuming when practising PA. Asking the students to design the assessment questions following examples can also save the teacher time as well as using a technological environment in which students can submit their feedback in a timely manner, thus preventing the recipients from losing focus during the assessment process, which sometimes can be very long (Liu et al., 2018; Luoma, 2004). Teachers can show students examples of questions formulated by them and can ask students to create their own. They might not formulate grammatically correct questions but they are usually understood by their peers.

Mobile phone usage for EFL learning, both inside and outside of the classroom, has become a common practice (Selcuk et al., 2019). Many students when producing language prefer recording themselves with a phone or use any other technological device. Technology usage has been evidenced as fostering language learning motivation, but this needs to be undertaken with a purpose in mind (Vonkova et al., 2021).

Mobile applications can promote greater participation and provide the students with practice on providing feedback to more than one student (Ko, 2015). Ko explains how 56 students used technology in their oral assignment. They used WikiTalki, an Android-based mobile application, to create a radio talk show in pairs. Through this activity, the students progressed their oral skills by absorbing key expressions, listening to others, providing feedback to their peers, receiving and reflecting on their own feedback, and re-recording an improved version of their own show on WikiTalki to reflect the feedback provided. In conclusion, PA is potentially useful for helping teachers undertake a valid and reliable assessment, while also reducing their workload.

7.4.2. Validity and reliability

Spoken interaction can be challenging to conceptualise and assess reliably. This could be because it is dynamic and context dependent. It may also be due to the many ways in which it can be used. For example, talking can occur in a one-to-one conversation, group discussion or as a monologue. Talking can also be deliberate and scheduled or unprompted (Carter and McCarthy, 2017; Luoma, 2004).

Language proficiency assessments have increasingly included interactive activities involving oral interaction. Despite it being recognised that peer-to-peer assessment lies in allowing for the assessment of a varied and broad range of skills, there has still been only limited research into the appraisal of oral skills and presentations. The evaluators and their rating criteria are key. This, therefore, calls for an examination into how those assessing oral skills interpret the spoken interactive tasks. Research into this can help to develop the rating of oral skills and guarantee that validity claims move beyond the 'content' level to the 'construct' level (Ducasse and Brown, 2009). The validity of an assessment relies on the criteria and the evaluation procedure, just as much as the nature of the task in hand. This is because, "in any assessment involving judgement it is the criteria by which the performance is judged which define the construct" (Brown, 2005, p. 26).

Validity must be regarded as the first aspect to consider when developing a test. As Luoma (2004) suggests, tests need to be meaningful for students and teachers, with time being needed to specify the aims and make them valid. According to some researchers, PA makes conversations even more realistic than those between teacher and student (Luoma, 2004). The design of an interview that really assesses spoken interaction (Brown, 2004), with criteria that are as objective as possible (Ahmad et al., 2019) is required. This is not easy owing to the fact that the judgment of a person is involved and as we have seen above, time is an issue. Patri (2002) highlights that assessing speaking is particularly difficult for second language (L2) learners, compared to the other skills, due to the fact that it is not easy to assess appropriacy, fluency and clarity; criteria that are often used in oral assessment.

Teachers normally have a curriculum to teach with a section dedicated to assessment. If they are not given any specific criteria or are unable to develop their own, they can use those suggested by experts, such as the Council of Europe (2001), which proposes that for the assessment to be reliable four to five criteria are enough (Ahmad et al., 2019) or the IB Internal Assessment oral criteria (See Appendix 6), which uses three criteria: language, message and interactive skills (IBO, 2018). Greenstein (2010) and Li and Dervin (2018) believe that professional development is needed to help teachers create valid criteria and to ensure formative questions are valid and reliable (Brookhart et al., 2009). Moreover, students must be trained in PA to be able to understand the criteria to assess their peers critically (De Grez et al., 2012; Li and Dervin, 2018; Liu and Li, 2014; Loo, 2016; Song et al., 2017) and to provide

reliable assessments (Kappe, 2008). Cullen et al., (2008) write that students and teacher might interpret the criteria differently. However, there is no agreement on this subject. Hafner and Hafner (2003) are of the view that, if different students give feedback to the same student, this will make the assessment more reliable. With or without training, teachers who are not persuaded of the importance of PA and its benefits may only see obstacles.

It is important that teachers understand and believe that PA can improve learning, and they must also be taught how to implement it. A study in Spain carried out by Panadero and Brown (2017) on the main advantages and disadvantages of PA showed that teachers were concerned by issues, such as low reliability and trust, when students score their peers. However, despite this, teachers found good pedagogical reasons for using PA in their classrooms.

The validity and reliability of the assessment of students in some contexts when practising PA is questioned by some researchers (Chang and Lee, 2005; Joo, 2016). They assign the issue of lack of reliability to the fact that students are sometimes not proficient enough to provide quality feedback, because they have not been suitably trained (De Grez et al., 2012). However, if students commit to peer assessment and take it seriously, they do not need to have a high proficiency level to give quality feedback. As Joo (2016) has pointed out, students must be given the opportunity to give open-ended comments, providing a considerable amount of information. This can help all students involved improve their learning. The comments can be oral (Patri, 2002) or written (Babaii et al., 2016). According to Black et al., (2004), "students may accept criticisms of their work from one another that they would not take seriously if the remarks were offered by a teacher" (p. 4). Other researchers even claim that assessments by teachers can be substituted by PA (AlFallay, 2004; Patri, 2002).

Finally, a teacher's assessment might be involuntarily biased (Bennet, 2011; Elwood, 2006) and students might be underscored or overscored, because they are better acquainted with norm-reference tests than criterion-related ones. Moreover, lack of anonymity when peer assessing might also be involuntarily biased (Langan et al., 2005). They observed that "gender influences were detected as males tended to grade other male speakers slightly more highly than female speakers" (p. 21). Brown and Knight (1994) also analysed biased peer assessments and noted that students tended to over-mark their friends.

7.4.3. Anxiety: Reluctance, lack of motivation and self-confidence

Spoken interaction has often been described as the most "anxiety provoking aspect in a second language situation" (Cheng et al., 1999, p. 420). Research into this anxiety exposes a link between students' fear of receiving a negative assessment and concern of how their performance is perceived by peers (Kitano, 2001). Whilst anxiety when speaking a foreign language can have a detrimental effect on students' oral performance, teachers often associate a reluctance to participate in oral activities with lack of motivation or inability on the part of the student, rather than their speaking-related anxiety (Gregersen, 2003; Tsiplakides and Keramida, 2009). Students do not talk in the languages classroom, if they lack self-confidence or if they are not motivated enough to talk, according to Fauzan (2016). Research has demonstrated that this reluctance comes from a lack of confidence in their oral skills' abilities, rather than from lack of interest or understanding of the value of learning English. Students worry about their peers' evaluation and how they compare to others in the classroom (Ansari, 2015). De Saint Leger and Storch's (2009) work has drawn attention to the fact that when students feel they are less proficient than their peers they lack the confidence to talk. This is also true when they are competitive and do not want to reveal their flaws to their peers (Dörnyei, 2001).

There are some factors that can influence the students' willingness to communicate, including how well they know their peers, how many students are participating in a certain task, the class size, the formality of the environment and the topic under discussion (de Saint Leger and Storch, 2009). All this can provoke anxiety and low participation in the classroom, hence limited learning. Some students' anxiety is reduced when in small groups or in pairs and this is because collaboration, rather than competition, allows anxious students to feel more confident and relaxed as well as less threatened by their peers (Ansai, 2015; Dörnyei, 2001). Dörnyei also observes that tests and assessment must be completely 'transparent', and students must be involved in the negotiation of the final mark (p. 4). This environment can allow them to feel more relaxed, although others might view the situation as artificial. De Saint Leger and Storch add that through self-assessment, students' will to communicate increases and this helps them build self-confidence (de Saint Leger and Storch, 2009).

PA might help students increase their motivation and confidence in speaking (Fauzan, 2016) and their attitude towards the feedback given by their peers, which is usually positive (de Grez et al., 2012). Sometimes the judgement that students give their peers has a major impact on their willingness to produce a work of a higher quality (Searby and Ewers, 1997). This might be for fear of being left behind in front of their peers. However, some students feel they are not efficient enough in assessing their peers (Cheng and Warren, 2005) and this can affect the quality of the feedback provided, making the assessment less valid. Others might not feel comfortable (Cheng and Warren, 2005) assessing their peers, due to competitiveness, lack of trust or lack of understanding of the set task (To and Panadero, 2019).

To sum up, an open and welcoming classroom environment will allow students to overcome some of the challenges they face when speaking (Leong and Ahmadi, 2017). This research explores the impact of PA on reducing some students' difficulties, such as anxiety.

8. CONCLUSIONS

Many recent studies have focused on the importance of encouraging students to engage in lifelong learning at school and to become independent learners throughout their lives (Black et al., 2003; Clark, 2012a; Nguyen and Walker, 2016; Sanmartí, 2020). A great many authors have observed the relationship between peer assessment and lifelong learning (Boud and Dawson, 2021; Nicol et al., 2014; Panadero et al., 2018) as well as peer assessment and autonomy (Heitink et al., 2016; Jones, 2014).

The new professional world requires people trained in skills, such as being able to self-evaluate and evaluate others (Ajjawai et al., 2020; Boud and Soler, 2016). This means a complete change in the way assessment is performed in class, as current research indicates (Broadfoot and Black, 2004). They have observed that "the worldwide tendency for more young people to stay on longer in formal education that now increasingly includes higher education, coupled with a growing discourse of 'lifelong learning', has helped to shift attention towards how best to support students' learning, rather than to judge it" (p. 19).

Assessment is closely related to learning (Suskie, 2018), and in order to promote lifelong learning students need to learn how to self-regulate their own learning by taking an active role in the process (Clark, 2012b; Claxton, 1995). Traditional assessments in which students are excluded do not encourage this active role. However, it seems that formative assessment practices, and especially particular types of feedback, can help students to self-regulate, making them participants of their own learning (Brown and Harris, 2013; Hattie, 2012; Panadero et al., 2017).

Formative assessment is increasingly common in education; however, it still needs to be seen more in the languages classroom, especially in the EFL one (Araujo, 2020). Despite the previous Spanish educational policy (LOMCE, 2013, p. 10) including formative practices (Sanmartí, 2020), most teachers conduct all assessments without involving the students. It is still difficult to guarantee that the assessment practised at schools furthers the learning process rather than inhibiting it (Wiliam, 2017). It is quicker and easier to correct mistakes without giving students feedback, especially if they get good grades in their exams through this practice (DeLuca et al., 2018).

These doubts and reluctance to try new methods of assessment could be overcome with some professional development, according to some studies (Chang et al., 2012; Greenstein, 2010; Popham, 2003; Wilson-Thompson, 2005). Some explicit teaching about formative assessment and creating the right atmosphere of trust in the classroom can increase the students' willingness to communicate. Both teachers and students need to understand the benefits of these practices (DeLuca et al., 2018). Hence, teachers require being taught how to start making small adjustments to change the culture of their classrooms and create environments more susceptible to formative assessment. They also need to teach their students to give feedback to their peers and how to assess themselves (Fauzan, 2016).

This study is focused on oral skills, the most overlooked skills in Spain due to the lack of an oral exam in the SUEE, as noted by García Laborda and Litzler Jerman (2015). These authors also observe that teachers do not dedicate as much time as they should to oral skills, and that they barely assess them. Both students and teachers are more concerned with doing well in the final exam than improving their English. As a result, students will not become more proficient at speaking until there is a cultural change in the classrooms.

The IB, unlike the Spanish Curriculum, includes an oral interview in the Internal Assessment. The motivation of having to perform well in the Internal Assessment makes oral formative practices in the classroom easier, especially PA. A great number of researchers have indicated that PA might enhance motivation and self-esteem (Cauley and Macmillan, 2010), especially when students are equipped with self-regulated learning strategies that might improve attainment (Brophy, 2004) and encourage lifelong learning. Understanding the assessment process stimulates students to work collaboratively, and this is something they will need to apply in their professional lives (Malan and Stegman, 2018; Stiggins, 2002). Applying metacognition to the activities performed in class, such as planning the task, monitoring and evaluating it generates feedback internally, which will help in the creation of self-regulating learning strategies (Boud and Falchikov, 2005; Buttler and Winne, 1995).

Students need to practise the oral skills and to be assessed in order to improve their oral skills and perform well in the internal assessment. Despite its constraints, PA practices can help students understand the assessment process (Malan and Stegman, 2018) and improve it (Chen and Warren, 2005). PA will also ensure students build the motivation required to speak (Fauzan, 2016), assess their peers and assess themselves. This thesis is focused on PA in the Diploma.

The theoretical framework that emerges from this review comprises educational assessment from the point of view of formative and self-regulation learning theories, especially peer-assessment of the oral skills in the languages classroom. The data obtained at SEK-Alborán International School was filtered through this focus, when trying to establish to what extent peer-assessment practices improve students' performance. From the perspective of independent learning, this study involved exploring the extent to which peer-assessment encouraged students to build their own learning. Some issues or challenges were observed. Relating the literature to this study, the framework, thus, looks at how PA assessment practices impact on students' performance and whether these help them to become more independent learners.

CHAPTER 3: METHODOLOGY

1. INTRODUCTION

This chapter focuses on the research methodology used in the study. Intended to be inclusive for an international readership, first, the context of the research, English as a Foreign Language (EFL) in the Almería area and at SEK-Alborán School is explored, with consideration of some issues and problems of assessment that are pertinent in this context. The concept of situatedness holds, as the classroom is the defined context, where students practise their English and from which all the data has been obtained. There is an introduction to the group of participants selected for this study and the challenges found. The reasons why this small sample group was chosen are also explained.

Second, a theoretical account of the qualitative approach and the epistemological belief selected for the research of this subject is provided along with the justification for choosing this technique. Also addressed is the importance of noticing everything that happens in the classroom and analysing it, consciously trying to remove all pre-judgements.

Third, a timeline of the research for addressing each of the research questions is provided and the methodology used is described in depth. The process of data collection is divided into four stages, which are explained in detail. Then, there is a focus on data analysis, with the coding process of the interviews undertaken by the students themselves with each other being described and the procedural difficulties involved in analysing the collected data are discussed. Subsequently, the concepts of reliability and validity are explored in relation to this study. The penultimate section discusses the importance of taking an ethical approach to this research study and describes matters of positionality and reflexivity regarding the role of the teacher as a researcher. Finally, the main points of the methodological approach are outlined in the conclusion.

2. THE CONTEXT OF RESEARCH

This study took place over a five-month period in the Academic year 2020-21 at the only international school in Almería, a city in the South of Spain. This is a private, co-educational

school, which offers the Spanish National Curriculum as well as the IB Primary Years Programme (PYP), Middle Years Programme (MYP) and Diploma Programme (DP). This study relates to one DP English class of thirteen students aged 16-17 and one teacher, the author of this study, and it involves the internal assessment of their progress in English as a Foreign Language (EFL).

SEK-Alborán School started with the Diploma programme in 2006, but in Andalusia it was not recognised until 2020. The 2021 academic year's students in year 13 (and the focus of this study) were the first ones who studied only for the Diploma programme in this city, and some of them had to do some SUEE exams to obtain the marks they needed to secure a place at university. This research will prove insightful in exploring the challenges these students faced.

The current Spanish legislation considers explicitly the acquisition of a foreign language as a priority: "el dominio de una segunda lengua o, incluso, una tercera lengua extranjeras se ha convertido en una prioridad en la educación como consecuencia del proceso de globalización en que vivimos" (LOMCE, 2013, p. 10). This includes the need for students to develop their communicative competence and increasing the number of overall hours that they devote to learning a foreign language.

However, according to Cózar et al., (2020), many Andalusian high schools have not achieved these requirements. These authors carried out a study in Andalusia surveying students of the 4th level of Educación Secundaria Obligatoria (ESO) and they compared it to a study that the Ministry of Education (MEC) published in 2004 through the National Institute of Evaluation and Quality of the Education System (INECSE in Spanish) [20], with data from 2001 (MEC, 2004). They have drawn attention to the fact that, despite students of the 4th level of ESO in Andalusia now being more motivated and interested in learning the English language than twenty years ago, there is still little difference regarding their communication skills. They believe this lack of proficiency in the English language could be related to the individual teacher and their own teaching method (Cózar et al., 2020). This places Andalusia amongst the communities with the lowest command of EFL (Education First English Proficiency Index

[20] Instituto Nacional de Evaluación y Calidad del Sistema Educativo.

[EFEPI], 2017). Many Andalusian pupils finish their Secondary Education able to communicate in English to some degree, but not fluently.

> Los andaluces están entre los españoles con un menor nivel de inglés, ya que el porcentaje de la población que declara tener un nivel alto o muy alto de este idioma es del 14 por ciento, ocho puntos menos que la media, situada en el 22 por ciento según recoge el IV informe monitor de la rama editorial de la Universidad de Cambridge'. (Cambridge University Press, 2015)

Following the publication of the CEFR (2001), the Department of Education of Andalucía initiated a Plan to Promote Plurilingualism (Plan de Fomento del Plurilingüismo) to enable Andalusian students to improve their linguistic competence in a foreign language in a communicative context (Gelle, 2013). The plan involved a challenging educative project that would allow Andalusians to gain a wider knowledge of foreign languages. This plan was approved by the government through the Agreement of 22nd March 2005[21]. The Government began to develop a programme called Bilingual Schools in order to raise the students' awareness of the benefits of using English for communication. In bilingual schools, students study in Primary and Secondary Education Geography, History, Sciences and Maths in English, ostensibly using a CLIL approach to learning. Accordingly, students have more hours learning in English, with conversation assistants helping them improve their oral skills.

However, according to a study performed in Almería by Gelle (2013) in which teachers were asked to give their opinion about the bilingual programmes in their schools, most said that they did not feel comfortable teaching in the target language. Gelle (2013) cites a teacher saying:

[21] Consejería de Educación. "Acuerdo de 22 de marzo de 2005, del Consejo de Gobierno, por el que se aprueba el Plan de Fomento del Plurilingüismo en Andalucía", BOJA núm. 65 de 5 de abril de 2005: 8-39.

> La administración en un principio permitió, como en el caso mío, al principio no tenía ninguna titulación, pero ya después pude hacer el B2 y dispongo de él, pero ya no estoy en la sección bilingüe, yo empecé con ella, pero después lo dejé. Una vez que la Administración andaluza reguló por orden el funcionamiento de los Centros Bilingües, exigió que el profesorado de asignaturas no lingüísticas tuviera que tener al menos el B2 o el título de Escuela Oficial de Idiomas o un título académico de Magisterio o de Licenciatura'. (Gelle, 2013, p. 56)

The obvious need was for teachers to improve their English to feel more comfortable and enabled to teach their subject in a more communicative way. The higher the level of English of teachers, the better they can express concepts in different ways, help students to relate them with each other and give them opportunities to use the English language in a meaningful way (Arnaiz, 2017). She also states that teachers need to be familiarised with the different pedagogical methods about how to acquire a foreign language, in terms of the structures and specific characteristics of the English language, as well as with the culture where the English language is spoken (Arnaiz, 2017). Almería has taken steps to respond to this need. In Almería, in 2019, there were already 112 public primary and secondary schools operating bilingually in English. This was one of the proposals of the Plan Estratégico de Desarrollo de las Lenguas en Andalucía, Horizonte 2020 (Plan PEDLA). The objectives of this plan are to develop the skills of the students' communicative competence, to improve teachers' methodology, increasing the number of teachers who have a C1 level of English and meeting the European guidelines by ensuring that up to 50% of students achieve a level of B1 in English by the age of 15 (Serrano, 2017).

There are 18 private schools in Almería and this thesis has been undertaken in the only international one in the city. All private schools, bilingual or monolingual, place considerable importance on ESL and usually prepare students for external exams, including the prestigious Cambridge exams, and often, through extracurricular activities as well as in the curriculum. Only two private schools and one public high school teach the DP. The first one to start teaching the IB was the International School (this study's school), followed by a private school and

finally, since the current Department of Education of Andalusia decided to establish a network of Public High Schools with the IB, a public high school in the city started to teach the DP in the academic year 2020-21. This decision has resulted in one public high school in each Andalusian city teaching the DP.

As we have seen in this section, EFL has had a great impact in Spain and in the Spanish Education System. There is no doubt about a need for EFL competence for exams and general communication purposes. The entire school community in Spain is aware of the importance of being able to communicate in English and they have tried to achieve it in their regions, with varying degrees of success. Different laws and programmes have been implemented to enable Spanish students to meet the requirements of the European Economic Community (EEC), however, they have not been entirely successful in their implementation at local and school level on account of contradictory messages, a lack of training in new pedagogical and assessment practices and because of teachers' own lack of competence and confidence in English. As aforementioned, a key issue is the lack of an oral exam in SUEE and the lack of homogeneous guidance and standards. This means that oral skills are not prioritised by teachers. These are major obstacles to the implementation of new practices and it is hardly surprising that teachers have sought a safe haven in their own idiosyncratic and familiar interpretations of traditional practices. This has an impact on their students, who are potentially prevented from achieving a high level of oral competence in English.

Despite bilingual schools and CLIL programmes having increased the number of hours students study in English, the teachers who participate in these programmes sometimes do not feel comfortable speaking English in the classroom (Gelle, 2013). There is a fundamental need for additional language training to ensure they are more proficient in teaching their subject in English. Teachers need a high level in the target language (CEFR, C1 or C2), the ability to create a relaxed atmosphere as well as using the target language in academic contexts and encouraging students to talk in it (Pavón et al., 2019). Moreover, the policy guidelines regarding assessment need to be clear and homogeneous at least within each region. In addition, there needs to be more professional development for teachers to improve their level of English and to learning about different assessment practices is required. In conclusion, teachers need more guidance on different assessment practices so that they can continue achieving good

results in exams, further develop their students' oral proficiency, and make progress in their own English communication skills. At present, the good intentions of assessment practice changes have been unrealised, because these three factors are in competition and not in harmony with each other.

3. PARTICIPANTS

In this subsection, how the participants of this study were accessed and the sample was constructed is explained. Mason (2002) explains the purpose of sampling and selection as: Sampling and selection are principles and procedures used to identify, choose, and gain access to relevant data sources from which the researchers will generate data using their chosen methods. (Mason, 2002, p. 120)

Deciding on sample size and its nature in qualitative research is a subject of debate. Agar (1986, p. 12) argues that it is not always helpful to ask "How do you measure that?" or "How large is your sample?" in social science research as this kind of enquiry deals with understanding the world from first-hand experience (ibid): "Hypotheses, measurements, samples, and instruments are the wrong guidelines. Instead, you need to learn about a world you understand by encountering it first-hand and making some sense out of it". (p. 12)

As explained before in the rejection of quantitative methods in this case, the following question by Silverman (2005) nonetheless was useful in the initial stages of this research: "Is the size and method of recruitment of your sample appropriate to both your topic and your model?" (p. 83). Since this research was undertaken at the school where the researcher worked, of which she has good understanding, decisions on the sample were taken conforming to what she felt would give reliable insights into the research topic. For example, for this study, purposive sampling rather than random sampling, was the preferred method to give the answers to the processes the researcher was interested in (ibid.). Denzin and Lincoln (2011) looked critically at purposive sampling methods, stating that: "Many qualitative researchers employ... purposive, and not random, sampling methods. They seek out groups, settings and individual where…the processes being studied are most likely to occur" (p. 245).

The researcher tried to construct a sample that was theoretically meaningful in order to build in certain characteristics or criteria, which would help to develop the argument (Mason, 2002). This research involved studying the impact of peer-assessment on IB students' achievement. The sample for this research was a group of Spanish IB Diploma students in the International School SEK-Alborán studying English. This research was focused on students in the twelfth grade (sixteen-years-old), who had been studying English for thirteen years.

The researcher decided to take the sample from only one Diploma class, because it was the only DP 1 Diploma class that she taught. This was her second year teaching this sample group and she knew they would be willing to participate in the research. The chosen class was formed by thirteen students, eleven male and two female, all being white and middle class.They had been in the same class for several years and were comfortable with each other, sharing information freely, and willing to ask for help and to share strategies (Tokuno, 2008). All the participants took part in this study on a voluntary basis, with parental consent - the form signed by the participants' parents is attached in Appendix 7. There was, however, one participant who did not agree to be recorded, so notes were taken of what he said. The participants will be referred to as: L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, L11, L12, L13, as follows.

Participants	Gender	Age	Languages known	The Common European Framework of Reference for Languages (CEFR)
Learner 1	Male	16	German/ English	B2
Learner 2	Male	16	German/ English	B2
Learner 3	Male	16	French/ English	C1
Learner 4	Male	16	French/ English	B2
Learner 5	Male	16	German/ English	B2
Learner 6	Female	16	German/ English	B2
Learner 7	Male	16	German/ English	C1
Learner 8	Male	16	French/ English	B2
Learner 9	Male	16	French/ English	B2
Learner 10	Male	16	German/ English	C1
Learner 11	Female	16	French/ English	B2

Learner 12	Male	16	French/ English	B2
Learner 13	Male	16	French/ English	B1

Table 3: Research participants

Once this thesis is submitted, all participants will be informed of the project's completion and thanked for their participation. They will also be offered a summary of the thesis and time to meet and discuss the findings.

4. THE STUDENTS' VOICES

The importance of making students part of their learning goes back to the early 20th century with the work of John Dewey. The term student's voice began to be used during the 1990s. Students began to be asked in primary and secondary classrooms about their learning experiences, with the aim of making reforms in schools (Fielding, 2001; Rudduck and Flutter, 2004). These practices have significantly spread and the voices of students have now become an integral part of education (Bourke and Loveridge, 2018). Researchers involve students in their studies by asking them about their educational experiences, views and perceptions. Jones's (2017) study, for example, citing a teacher's classroom research with MFL secondary school pupils and target language use, found that teachers' thought that they used more target language than they were really using. They were not aware of the students' real feelings:

> I found a considerable amount of target language used in the case of teachers but not enough on the part of the students. The teachers thought they themselves used more target language than they did. The pupils thought that they needed and would have liked more target language than they did. (Jones, 2017, p.10)

It is very important to listen to students and integrate what they say that is fair and reasonable to improve, even transforming teaching and learning. As Fielding states: "transformation requires a rupture of the ordinary and this demands as much of teachers as it does of students"

(Fielding, 2004, p. 296). The student voice movement has become very popular recently, mainly in Anglo-Saxon countries, however, it is barely heard in Spain (Susinos-Rada et al., 2019).

Fielding (2004) writes about the model of students as co-researchers, where their voice is key. In this model, the teachers' experience is very important to carry out research, however, teachers cannot succeed if students do not engage in research, becoming enquirers and making meaning themselves. Students are not natural researchers and need to be included in a dialogue about research and how it may impact on them. In this research, students discussed their feedback with their classmates and with the teacher, learning new practices together. The researcher learned from the feedback provided by her students, because she gave them independence to speak with 'their own voice'. The students as co-researchers learnt things about independence in learning, about their individual and collective agency (Fielding, 2004, p. 307) as well as their responsibility. The lattermost point chimes very well with a principle of Assessment for Learning (AFL). That is, AFL emphasises the development of student responsibility for their own learning (Black and Wiliam, 2009).

5. RESEARCH APPROACH

Education is seen nowadays as a way of life, where students must be taught skills as well as given the knowledge to adapt to the fast-moving society. Teachers' responsibility is to help students to develop a life-long ability to learn. A qualitative approach was deemed the most suitable for this research. Qualitative methodology within a naturalistic approach has been used for researching the role of peer assessment in an IB languages classroom. A naturalistic approach implies that this research took place in a real life setting, with minimal manipulation from the researcher. Lincoln and Guba (1985) refer to naturalistic enquiry as a 'discovery-oriented' (p. 298) approach that minimises investigator manipulation of the study settings. In research, naturalistic investigation involves using naturally occurring data, i.e., not material primarily produced for academic purposes, but rather, that originating from everyday aspects of social life, thus enabling the studying of 'life as it happens'(Potter and Shaw, 2018, p. 182). Methods applied with this perspective engender inductive and data-driven analyses. Hence this

form of inquiry, rather than drawing upon prior conceptualisations, is aimed at eliciting what in the observed data is relevant, usually by considering previous findings in the literature. In this research, authentic data in a naturalistic classroom setting was collected. The researcher left the students to perform their own interviews and even devise their own questions when they interviewed each other. All the process was performed in a natural environment, where the students felt relaxed, and they did not feel they were part of a research project.

A quantitative approach was rejected, because this research involved focusing on uncovering meaningful insights and there was no comparison between different research groups. The term 'qualitative' implies that the research process was "not experimentally examined or measured in terms of quantity, amount, intensity, or frequency" (Denzin and Lincoln, 2011, p. 8). Qualitative researchers study topics in their natural settings, "attempting to make sense of or interpret phenomena in terms of the meanings people bring to them" (ibid. p. 3). For this research, the students' recordings were studied, with words and sentences being analysed and interpreted by the researcher. The researcher does not talk about quantity, intensity or frequency, she always talks about the students and interprets what they say. Her intention was to look at the feedback students gave to their classmates in a peer assessment task to see whether this task helped in encouraging students to speak in the classroom and giving them more independence. As Denzin and Lincoln (2011) point out:

> Qualitative research is a situated activity that locates the observer in the world. It consists of a set of interpretive practices that make the world visible. These practices transform the world. They turn the world into a series of representations, including field notes, interviews, conversations, photographs, recordings and memos to the self. At this level, qualitative research involves an interpretive, naturalistic approach to its subject matter. (p. 3).

The situatedness of the activity is an important point. By situatedness what it is meant is that learning cannot be separated from the context in which it takes place, as Lave and Wenger (1991) point out. Moreover, learning should not be considered as being the transmission of abstract and decontextualised knowledge between individuals, but rather, a social process

where knowledge is co-constructed. Students construct knowledge when they assess their peers, and as shown in this research, they learn from each other. Learning a foreign language is quite challenging when students are in a country where it is not spoken amongst the community on a day to day basis, as Efe et al., (2011) hold. However, if students are given a number of authentic opportunities in classroom to talk, they can construct their own knowledge, as shown in this research.

Qualitative research, as Mann (2003) states, "can offer insights into complex multiple, even evolving truths (…) and qualitative approaches enable the researcher to take an open-ended, exploratory approach where little is predefined or taken for granted" (p. 66). New ways of moving forwards and how assessment practices relate to agreed policy for the benefit of the institution (ibid.) were hoped to be discovered.

In teaching situations, much can be learnt from observation and being aware, an act that is inherent to teaching, as Mason (2002) observes: "Every act of teaching depends on noticing: noticing what children are doing, how they respond, evaluating what is being said or done against expectations and criteria, and considering what might be said or done next" (ibid. p. 7). 'Noticing' provided rich insights into classroom practices. However, Mason (2002) also notes that issues have been raised with the discipline of 'Noticing', including with regard to what constitutes data, research findings and validity. Mason suggests that there are three questions to address while 'Noticing', these being: "How do you decide what to notice, how do you select which incidents to study and how can you remove all judgement and interpretation from brief but vivid accounts?".

As Silverman points out (2013), the choice of a particular methodological approach is dependent on the topic and hypothesis. In this study on assessment practices in IB classes in private schools a qualitative approach was required in order to achieve the objectives of the research. The aim was to gain insight into how students engaged in peer assessment in an English lesson, and the effect that this produces on talking in the classroom. Blommaert (2005) describes a qualitative study as one in which there is interaction between teachers and students; through informal conversations, providing opinions and trying new assessment practices. Using the principle of deductivism (data are collected about various concepts and their interconnections are investigated), the hypotheses of relationships between peer-assessment

and improvement in students' performance can be tested and defined. Moreover, following the principle of inductivism, perceptions of how the students felt practising peer-assessment and how they could improve their performance were gathered by the data and observations (Bryman, 2016).

6. RESEARCH PLAN

The aim of this research was to study how a group of 1st year Diploma IB (International Baccalaureate) students engage in peer assessment in an EFL (English as a foreign language) classroom in an Independent school in Spain and the effect on classroom talk. The objective was to raise Spanish teachers'awareness of how best to use peer assessment formatively during the teaching of oral oral skills in upper secondary MFL classess.

This research was performed in three stages:

1. The students recorded themselves in their 4' oral presentations about the extract of the book they have had to read for the IB internal assessment and received written open comments from their peers.

2. The students worked in pairs. Firstly, they gave the 4' presentation, while their partners (acting as teachers) listened to them and took notes. Then, they asked them questions about the presentation for 5 minutes. Finally, they assessed their partners' presentations and answers to the questions based on a proforma. The feedback given by them was recorded with a phone.

3. The teacher interviewed the students. While the interview (presentation + questions) was taking place, two students acted as moderators assessing the student and taking notes, using the real IB criteria.

The data were collected through feedback that students gave on their classmates' presentations and follow up discussions. The process of the assessment task is detailed below, and the obstacles faced are outlined. The students practised the three different parts the

Language B Higher Level (HL)[22] oral assessment consists of: 1: presentation, 2: follow- up discussion and 3: general discussion, from the table below, which helped them to prepare their Internal Assessment for English B in the Diploma, which they were to take the following year.

The following table summarises the three parts into which the Internal Assessment (IA)[23] of English B (Name of the subject) Higher Level (HL) subject in the International Baccalaureate is divided.

Supervised (by any other teacher of the school) preparation time	The student is shown two extracts of the book ' (Appendix 8) of up to approximately 300 words each: one from each of the two literary works studied during the two years course. The student chooses one of the extracts and prepares a presentation focused on the content of the extract. During this time, the student is allowed to make brief working notes.	**20 minutes**
Part 1: Presentation	Students present the extract on their own. (They had be taught how to do it before). The student may place the extract in relation to the literary work, but must spend the majority of the presentation discussing the events, ideas and messages in the extract itself.	**3–4 minutes**
Part 2: Follow-up discussion	The teacher engages with the student on the content of the extract that the student has presented, expanding on observations that the student has provided in the presentation.	**4–5 minutes**
Part 3: General discussion	The teacher and student have a general discussion using one or more of the five themes of the syllabus as a starting point. The themes are experiences, human ingenuity, identities, sharing the planet and social organization.	**5–6 minutes**

Table 4: Oral Internal Assessment (Language B guide, 2020, p. 50**)**

[22] https://www.ibo.org/programmes/diploma-programme/curriculum/language-acquisition/language-b-sl-and-hl/

[23] An IA is an individual assessment that the teacher of the subject assesses following some criteria. The IA is moderated by IB examiners.

https://blogs.ibo.org/blog/2017/11/08/ib-terms-explained/

The process of this research had four stages developed by the researcher that are described below.

Stage 1 (focused on part 1= presentation)	28th November 2020	55 minutes	Class discussion and research sensitisation about the task to be done by the students.
	29th November 2020	45 minutes	Students started recording their presentations themselves with Flipgrid[24] in the class.
	3-4th December 2020	1 hour and 50 minutes	Students listened to each other's recordings individually in the platform (Flipgrid, now called Flip) and then wrote some feedback (open comments) on Flipgrid, which they all could read.
Stage 2 (focused on parts 1 and 2 explained above). Presentation and follow up discussion)	15th January 2021	50 minutes (last period before leaving for home)	Students were informed how to do part 2 (explained above). Some samples were shown to them. We all made comments related to the IB criteria (Appendix 6) and free comments.
	16th January 2021		They worked in pairs preparing their presentations and having follow up discussions. They were given

[24] An online video tool from Microsoft that allows for discussion.

https://info.flip.com

			a peer-assessment proforma to have a guide of the kind of feedback they had to give their partners. How to do the peer assessment was explained to them.
	21st January 2021	1 hour and 50 minutes	Students started doing the peer-assessment activity. First, one student asks the other one some questions (Appendix 10) as they try to have a conversation (to achieve the highest mark according to the IB criteria). Then, they change roles and do the same. They both finally are recorded giving their partners feedback about how they have answered the questions, interaction, etc (according to the IB criteria).
	22nd and 23rd January 2021	50 minutes and 50 minutes	They continued doing the activity.
Stage 3 (focused on part 3): follow up questions: 5-6	28-29th January		In pairs, students asked each other questions about the themes prescribed by the subject syllabus (Experiences, sharing the planet, identities, social organisation and human ingenuity) and they gave feedback to each other.
	4th and 5th February		They continued with the activity.
Stage 4 complete oral task: part 1: presentation, part 2	28th-29th February		The teacher carried out the IA task (parts 1, 2 and 3) with every student individually,

follow up questions and part 3: general discussion about the themes.			while two students acted as moderators taking notes and assessing the student interviewed. On this occasion, the Diploma criteria for the IA were used (Appendix 6). The feedback given by the two students acting as moderators was recorded and transcribed.
	4th and 5th March		We continued with the activity.

Table 5: Study stages developed by the researcher

The researcher collected data following four stages, so that the students had the opportunity to progressively understand the process of peer assessment. The students were also able to use the feedback provided by their peers to improve their own presentations in the subsequent stage. The stages of the activity are described in detail below.

Stage 1: Class discussion, research sensitisation and first peer assessment activity for part 1 (presentation). In class intervention.

At the first stage, the main aim was to sensitise the pupils with a class discussion about the peer assessment activity that was to be done. The research process was explained to the students. They were told that we were going to practise the activity that they would have to do for their IA. They would be practising every part of the IA using the IB criteria and taking real extracts from the literary work studied in class as part of the language B HL course. They had practised formative assessment before; however, they had never performed any peer assessment for an activity in which they gave and were given some feedback by their peers in preparation for a final summative assessment, which according to Stobart (2008) is a key aim of formative assessment.

In the IB guidelines, it is clearly stated that "The IB encourages teachers to practice teacher-led formative and summative assessment as well as peer assessment in order to develop their language skills" (IBO, 2018, p. 58). The students received the learning criteria in the form of a proforma (see Appendix 8) and with these these criteria being explained to them and they were also informed about the purpose of the oral activity. The teacher's intentions were to perform an activity to study the link between peer-assessment and improvement of oral skills. The main objectives were to improve both their oral skills in English and their knowledge about the English B HL Internal Assessment that they would have to undertake the following Academic year. The students were keen to participate in the activity and made an effort to speak, because they knew that they would soon be assessed on their oral skills for the IB (Fauzan, 2016). Moreover, they would prepare this standardised oral test doing peer-assessment, which is recognised in the Spanish Educational policies to help students develop their lifelong learning skills (Broadfoot and Black, 2004).

Whilst the activity for the English B HL oral Internal Assessment had already been explained to the students at the beginning of the academic year, this was repeated for absolute clarity. A clarification of Table 5 above was used. This was taken from the International Baccalaureate Organization Language B guide (IBO, 2018, p. 50).

Each student was asked to choose an extract of about 300 words that they considered relevant in the book they had to read in class. In the real summative Internal Assessment students are given by the teacher two different extracts of the books they have read, but they would not have studied these in detail in the class. However, this time students would be free to choose the extract, thus allowing them to participate in the whole process and to activate them as owners of their own learning, as urged by Black and Wiliam (2009). Once they had selected a text, they had to share it with the teacher and, then, start preparing the presentation individually. Once prepared, they recorded themselves on Flipgrid.

Some students struggled with the technology and found it easier to record themselves with their mobiles. They were allowed to have the extract with them when presenting and no more than ten bullet points of the ideas they were going to develop in their presentations. However, they were not allowed to read what they had prepared. Those who did not have time to finish the activity in class had to do it for homework.

On the third day of the process, the students began to listen to their classmates' recordings on Flipgrid. Some of them did not upload them on the program and, instead, sent them to the teacher. Surprisingly, even the two students who were usually more reluctant to speak in class were able to successfully record their presentations on Flipgrid and provide written feedback (open comments) to their classmates. They were not comfortable with recording oral feedback for their classmates, because as Hamzah and Ting (2009) indicate, some students' feeling that their level of English is not as good as their classmates' might make them reluctant to give such feedback. It was decided not to insist, as written feedback would also be valid. The aim was to obtain feedback that might move the learners forward (Black and Wiliam, 2009) and to help respond to the research questions. The data obtained from the feedback were coded and categorised according to the approach described by Strauss and Corbin (1990) (Appendix 11), which is explained in the process of data coding described in Chapter 4.

Since learning is most effective when it is active and social, as Black and Wiliam (2009) point out, this study was aimed ascertaining the extent to which pupils were effectively involved in their own learning process, through assessment practices. Teaching and learning is shifting "from planning based on teacher intention to planning directed towards pupil outcomes", Stoll and Fink (1999, p. 124) suggest. When students understand their teacher's expectations, they have a better chance of learning (ibid.). Students' involvement in the learning process is also key to this, it being essential for motivation and it sits at the heart of AFL. Boekaerts (1995) asserts that by developing pupils' participation in their own learning and encouraging their independence, learning self-regulation can be more effectively achieved. These were also the researcher's convictions.

Only eight students (L3, L4, L5, L8, L10, L11, L12 and L13) received feedback from their classmates and all except L5, L7 and L8 gave it to their peers. Some were reluctant to give or receive feedback, because they lacked confidence, believing that their oral skills were worse than most of the students in the class. Others were hesitant, probably because they were not ready for peer learning and assessment, as To and Panadero (2019) believe. There was also some competitiveness amongst them, and some weaker students did not want to show their level of English in front of their peers (Cestone et al., 2008). The table below indicates the students who received and gave feedback.

	Received feedback	Gave feedback
L1	No	Yes
L2	No	Yes
L3	Yes	Yes
L4	Yes	Yes
L5	Yes	No
L6	No	Yes
L7	No	No
L8	Yes	No
L9	No	Yes
L10	Yes	Yes
L11	Yes	Yes
L12	Yes	Yes
L13	Yes	Yes

Table 6: Students giving and receiving feedback

Stage 2: Peer evaluations with proforma; parts 1 (presentation) and 2 (follow up questions about the presentation) of the interviews. In class intervention.

There was more than a month's gap between stages 1 and 2, so the researcher had to remind the students of the task at hand as some had lost focus. It was not possible to continue with the activity until the 15th of January 2020. First, there was a bank holiday in Spain for 'Día de la Constitución", then the students had their term exams just before Christmas and finally, because of the Christmas holidays, which in Spain last until the 7th of January. Moreover, it was not possible to continue with the activity due to some school activities taking place and disrupting normal lessons.

At this stage, it was explained to the students that we were to do a practice of their IB internal assessment in the next few weeks, this time, parts one and two. We, as a group, listened to a

few real examples of the internal assessment taken from My IB[25] (an application for teachers and schools in the IBO) and we all made comments both related to the IB criteria and free comments. It was quite difficult to understand the audio recordings since we had not read the books from which the extracts had been taken. There were comments such as "I would like to read that book", "I think he is Spanish", "She is not Spanish" "His English is very bad" and some funny faces when we listened to a student who stammered a little bit. Some students did not pay much attention to the content, just to their English. They thought they could guess their marks just by listening to them for a few seconds. They were surprised that accent is not so important and even if you have a problem of stammering, for example, you can achieve a good mark. It was not done on purpose, but it was later observed that it was a good example since there was in class a student who stammered a bit and was shy to speak for that reason.

The teacher explained to them the most important things to bear in mind about the internal assessment and what we were to do. They started to guess the mark randomly. They still did not know exactly what they had to do in the second part. Others did not do well in part one and I needed to explain it to them again. This was done during the last period of the day, just before leaving.

The following day they were given time to prepare the presentations, this time in pairs. Taking into consideration the fact that students' objectivity of their peers' work might depend on the classmate they assessed, as indicated by Foley (2013), the teacher tried to group them in pairs with the same level of English. However, this was not easy since there were a couple of them who felt a bit inferior to the rest and they were very shy. L12 was very good at written English, however, his oral English was not so good, and he stammered a little bit. L2 was clearly the one with the lowest level of English in the class. So finally, they were allowed to choose their partner. The teacher separated L6 and L11 of the class because they became distracted very easily. One of them would be working with the most 'difficult' student in the classroom, who never wanted to do anything in pairs or in groups. As suggested by Webb and

[25] https://internationalbaccalaureate.force.com/ibportal/IBPortalLogin?lang=en_US

Jones (2009), some students prefer to work on their own and receive feedback from their teachers. The teacher was not sure how this would work.

They started preparing their presentations and the possible questions, examples of some of which can be seen in Appendix 9. It was thought that asking their peers their own questions would give them more autonomy and would make them more responsible in the activity, as Jones (2014) recommends. The students were only recorded when they were providing their peers' feedback. They were given the peer-assessment proforma again so that they had a guide of what they needed to listen to and finally, to give their partners some feedback based on the IB criteria. The teacher's intervention included asking the students questions which they had not asked their peers, such as: did they give their opinion? What did you think of the language they used? Was there any interaction? Did they provide a brief summary of the book? The teacher also asked questions like: What advice would you give them to improve next time? Do you now know exactly what you have to do? Using these questions, the intention was to highlight what exactly they had to do in the interview to score highly. This detailed description gives some insight into how, with student research, it is necessary to provide scaffolding at appropriate stages (van de Pol et al., 2010).

L4 and L1 were a bit reluctant to do this activity. L4 thought that his English was already perfect, with his American informal English, whilst the other was not very proficient and sometimes did not talk for fear of making mistakes in front of his classmates and usually turned back to speaking Spanish. He was always aware of any mistake that his classmates made. At the end L1 said, "*These IB people are really strict*". He meant that it was very easy to make mistakes and very difficult to score high. Next time, it was decided, they would not work together.

L2 and L6 were really reluctant to do the activity, especially L2, who always refused to work in pairs or in groups. L6 did not normally take activities very seriously, but both did this activity, because all their classmates were doing it. However, they did not seem to see the point of the activity and did not want to speak and make mistakes in front of their classmates.

L11 and L5 took this activity seriously and gave excellent feedback. L3 worked hard to give feedback to L9, because he could see the effort L9 was putting into the task. L11 volunteered

to do the task with L5. At this stage, all the pairs gave and received feedback, except L12 and L13. This was because of time pressure; L13 was not able to provide feedback to L12.

Stage 3: Part 3 of the interview: general questions about the themes studied in class.

The questions relating to part two were removed as they only referred to the extract and the students were not practising this at this stage. The students could not focus on all the questions of the proforma at the same time. Hence, some more specific questions were also removed from stages three and four. The researcher decided to include only those that appeared in the Language B guide with the bullet points, which summarised clearly all the important ideas the students needed to remember (IBO, 2018). The students were free to choose questions to ask their partners on the five themes studied in class, according to the Language B guide (Identities, Experiences, Human Ingenuity, Social Organisation and Sharing the Planet) (IBO, 2018). As Boekaerts (1995) points out, students must participate in their own learning.

L2 was more at ease when his answers were not being recorded and therefore, the teacher decided to note down his responses during the interviews instead of recording them. The format of the interviews was semi-structured, but all students were asked questions relating to the five themes studied in class, according to the English B guide. In order for it to reflect as closely as possible the real IB interview, the activity was semi-structured, requiring quick decision-making by the students on the order of the questions. Moreover, for this thesis, the researcher was particularly interested in the students' feedback to each other on how they might improve their oral skills. While they were given a proforma with some questions to give feedback to their classmates, they were not required to follow the order of the questions, and more open comments were also considered valid.

As the concepts of teaching and learning change, a shift is required from "planning based on teacher intention to planning directed towards pupil outcomes" (Stoll and Fink, 1999, p. 124). Therefore, throughout the activity, the researcher was observing to what degree students were actively involved in the learning process, particularly as learning is most effective when it is active and social, as indicated by Coyle (2007).

The recordings were always made by the teacher (researcher) and she did not have time to record L12 giving feedback to L13 nor L5 and L6 giving feedback to each other. The following day, the students were asked to provide oral feedback, which the researcher then noted down. L6 was reluctant to give more feedback and she even told the teacher to listen to the recording herself and she summarised her feedback in one sentence. While L4 enjoyed giving feedback, L2 simply did not like the task and did not care very much about doing so. He did not admit doing anything wrong and did not accept any constructive criticism. L7 was a bit reluctant at the beginning, but as he saw that most of his classmates were enjoying the exercise and seemed to find it useful, he started to enjoy himself too. L9 and L13 took the activity seriously from the start. L10 was excellent at giving feedback and L1 finally agreed to do the activity, although he was hesitant about speaking in front of his classmates. He did not want to accept that most of them spoke English better than he did. L11 volunteered to participate in two interviews since there was an odd number of students. The only student who did not give any feedback was L8, because he did the interview with L11, who had already done the activity with another student.

Stage 4: The researcher undertook the activity with a student (parts 1 and 3), which was recorded while two students were listening and taking notes. At the end they needed to give the student some feedback and a mark.

At this stage the aim was to do some collaborative learning and social interaction, which could help students to self-regulate their own learning, as Coyle (2007) maintains. Two students need to agree on a mark out of 30, according to the IB criteria. This practice, according to Fauzan (2016), is believed to enhance the students' confidence and motivation, making them participants of the learning process (Topping, 2009). Both the peer assessed, and the assessor can benefit from the activity, they suggest. Most students felt comfortable giving their classmates a mark. This is contrary to the opinion of researchers, like Cheng and Warren (2005), who consider that students usually do not feel comfortable giving feedback to their classmates due to their lack of confidence in their personal peer-assessment skills.

Seven students received feedback and a mark from two students. Two received feedback and a mark from three students. Three received feedback and a mark from one student and one did not receive any feedback (L7). L2 was consistently unwilling to participate in the activity

and did not accept the feedback given to him as he was not able to accept the fact that his English was of a lower level than that of his classmates.

Seven students	Received feedback and a mark from two students
Two students	Received feedback and a mark from three students
Three students	Received feedback and a mark from one student
One student	Did not received any feedback

Table 7: Students receiving feedback and a mark

The design of the interview questions allowed for the students to relax, with those to start with relating to the themes they had previously told the teacher they would be comfortable talking about. Then, more challenging questions were asked, if the teacher felt that the students would be able to answer them. The students seemed to be comfortable and trust the teacher, who always gave them time to think, as advocated by Black and Jones (2006) and Hodgen and Marshall (2005). The intention was to create an atmosphere of trust so that students were interested in each other's opinions (Liu and Zhang, 2018). Before forming a conclusion and extracting the main themes, the students' replies were analysed according to assessment and feedback practices. This process is explained later in Chapter 4.

The researcher was aware that her positionality, whereby she had been their teacher for two years, could impact on the results of the research as "Who you are (or are becoming) determines to a large extent what and how you research" (Saldaña, 2011, p. 22). Therefore, she emphasised the importance of "morally responsible research behaviour" (Kvale and Brinkmann, 2009, p. 74). Knowing her position as the main instrument for gaining knowledge through the interviews and transcripts, her "moral integrity, sensitivity and commitment to moral issues and action" (Kvale and Brinkmann, 2009, p. 74) were of great importance.

Reflexivity played an important role in this research, which Robson (2002, p. 551) describes as "the process of researchers reflecting upon their actions and values during research

and the effects they may have". For the purposes of this study, reflexivity can be defined "as explicit, self-aware reflection and analysis toward increasing richness and integrity of understanding" (Finlay, 2012, p. 317). When the researcher was in the field, after each interview, she focused on what she could have done better, including re-thinking the process of the interview in order to recognise any omissions or mistakes and to improve her research techniques. Throughout the interview process, the researcher attempted to identify possible insincerity in her participants' responses, but in nearly all cases, she felt that her participants were sincere.

7. INSTRUMENTS FOR DATA GATHERING

The researcher used various tools in this study, the main one being interviews. The presentations were recorded on Flipgrid in the first stage, whilst the interviews were recorded with a phone. Two different proformas (rubrics) were used by the students with the relevant aspects of the criteria, which helped them to give feedback to their peers.

7.1. Interviews

Depending on the level of formality required, there are three interview types: unstructured, semi-structured and structured. Unstructured interviews usually start with open questioning regarding the subject of interest, with follow up questions being dependent on the participant's responses (Holloway and Galvin, 2016). However, the term 'unstructured' is rather a misnomer as no interview is completely without any structure. Hence, despite such an interview being non-directive and flexible, an interview guide is utilized by the researcher, but one comprising themes and examples of questions, rather than specific questions. Analysing the data from such interviews is challenging and takes much time, as it involves collating the responses from various participants and identifying links between them, not an easy task. Researchers have to attentively listen and record any new or interesting data provided by the participant (Doody and Noonan, 2013).

For the current research, semi-structured interviewing was selected as pre-set topics and questions allow for basic information being collected, whilst also creating opportunities for interviewees to provide more detailed responses. Moreover, they allow interviewers to probe what is intended by and also exploring issues of particular interest in more depth (Berg 2009). Further, the researcher can vary the wording and order of the questions (Power et al., 2010), according to the direction of the interview. One key challenge for the interviewer is when to ask prompt questions and what these should entail. Finally, structured interviews are basically administered questionnaires, containing set questions and possibility for follow-up questions to responses to elaborate any futher on this.

Interviews are a helpful format for the exchange of opinions and views on specific topics (Kvale, 1996), also being well-suited to a small-scale educational study, where knowledge is produced through social interaction. The researcher was aware, nonetheless, of the problems with interview as a technique. It was difficult to make some students feel comfortable. Whilst the researcher had already established rapport with the students in class, it was necessary to explain them about the research, confidentiality, and the use of the audio recorder. As the class was the location where the interviews were carried out, there were some interruptions and some noise. The researcher decided that interviewing was a suitable tool in this particular study since the activity was exactly what students would have to do in their real exam. Besides, they were undertaken in a natural way, enabling vivid and accurate accounts and allowing the researcher to explore the students' understandings of questions, experiences, social situations, etc. The qualitative interview according to Powney and Watts:

> Avoids reducing the complex responses or behaviour of an individual to a single number in a maze of statistical calculations. The focus becomes the detail and quality of an individual or small group's experience rather than the number of people who responded in a particular way (1987, p. 22).

A mobile phone was used to record the interviews, thus allowing the teacher to focus entirely on the interviewing process, giving all the attention to the students and maintaining eye contact with them. It was felt that this approach created a more natural social environment.

Prior to the interviews, the students were asked whether they were comfortable with this particular method of record keeping. It was decided to use notetaking during the interviews to put them at their ease and to make them less anxious. Whilst notetaking can be distracting, according to Blaxter et al., (2001), the researcher's personal experience suggested that it was a preferred form of recording information from the interviews. Despite recordings taking a long time to transcribe and analyse, it was felt that live recording their responses was an appropriate and accurate method to use. Moreover, this is the form used in the real IA English B oral interview. The collected data, with all the names deleted, were kept safely locked up in the researcher's study at home.

This task was approached from the perspective of a detached observer, so that the researcher would code objectively, rather than being influenced by what she wanted to find. She made every effort to distance herself from her perceived beliefs. She adopted a more neutral stance in order to answer the research questions with as much validity and accuracy as possible. This was achieved through reflection, critically reflecting on her observations and field notes and references to pertinent literature to keep a balanced view.

7.2. Flipgrid

Research (Burston, 2015; Kukukska-Hulme, Lee and Norris, 2017) agrees on the benefits of using technology to access to a wide range of learning opportunities. Social networking platforms such as Facebook or WhatsApp have become very popular over the last few years (Statista, 2019) and they have been widely used to foster second language acquisition for the following reasons: (1) interaction leads to negotiations regarding meaning (Bueno-Alastuey, 2013); (2) peers and teachers provide feedback, thus promoting collaborative learning (Sauro, 2009); (3) cognitive load and affective filter are reduced, which leads to students feeling more comfortable when producing language (Sykes, 2009); and (4) information is retained for subsequent analysis and assessment, with the interaction remaining within the application, thus providing the opportunity for future evaluation (Andujar and Cruz-Martínez, 2017; Tang and Hew, 2017).

To foster peer assessment of English writing, Facebook has been used (Parejo, 2016; Shih, 2011; Winet, 2016). That is, it has emerged that it can be a positive tool for developing EFL writing and fostering cooperative learning as well as increasing interest and motivation. However, skills such as speaking and reading have yet to be thoroughly assessed in this application. Andujar and Salaberri (2019), when comparing the application being used in a mobile context or in a computer context found there were differences. Specifically, whilst the mobile environment generated a greater degree of interest in students than the computer one, the latter produced higher cognitive engagement, with the students being more focused during the interaction than in the mobile context scenario.

Flipgrid's website was created by a professor at the University of Minnesota in 2014 and at present, it is used by millions of educators in a great many countries all over the world, as Craig (2020) observes. Barlett (2018) also explains that Flipgrid works like a forum for discussions, where the teacher prompts a task and then, the students have to post their responses using a video instead of typing them. Students respond to each other's posts, which increases communication among them. In this research, Flipgrid was used in the first stage. Students recorded their first presentations and gave each other feedback. Some students used their computers and others their phone to do the recordings, usually depending on what they preferred or if they needed to find a a quieter place outside of the classroom. This is an excellent way to motivate students to talk (Craig, 2020) and to learn from each other.

As technology has become an inseparable part of our daily lives, teachers have been integrating Flipgrid into their classrooms to assess students' oral skills (McLain, 2018), to improve students' presentation skills (McClure and McAndrews, 2016), to increase student engagement (Bartlett, 2018) and to practice English communication skills (Petersen et al., 2020). As Edward and Lane (2021) observed, some learners might be hesitant to use Flipgrid, but it has the potential to provide an effective platform for interaction and communication in a digital environment.

7.3. Rubrics (proformas)

Classroom assessment gauges student progress and provides feedback to learners, whilst also allowing for teachers to evaluate the efficacy of their instruction. Teachers need to be able to design and undertake "varied and valid assessments…to support student learning" (Short et al., 2018, p. 60).

Open-ended tasks, like that investigated for this research, are focused on course communication objectives, which require assessment tools able to measure the quality of the language produced whilst they are being carried out. Clearly ranked descriptions of language performance will enhance the consistency of assessment and accordingly there has been increased usage of rubrics for assessing performance quality during open-ended tasks, like speaking.

In this study, the term "rubric" (proforma) refers to a specific assessment tool that ascribes descriptions of different levels of performance quality. Goldberg (2014) defined a rubric as "a scoring guide that outlines features of work at different levels of performance" (p. 1), there being two types: holistic and analytic.The former can be practical, with only an overall score being needed (e.g., a quick placement test), whislt the latter are more effective for assessing the learning process (Brookhart, 2018). For these, multiple traits or categories were listed separately, according to a rater, with the aim of ranking depending on descriptions of levels of performance. As the teacher can assign separate ratings for each category, the student acquires specific feedback on what they are able to do and what needs attention (Brown, 2018). Moreover, with an analytic rubric, the results provide information regarding learner's strengths and weaknesses to the teacher. Hence, when assessing open ended extended language performance, they are an appropriate choice, as they can deliver consistent assessment and feedback to the student.

Providing rubrics to inform leaners about assessment criteria and level of quality regarding a task assignment, is a commonly utilised approach for carrying out self-assessment (e.g., Panadero and Jonsson, 2020). Using assessment rubrics has become very popular as they stress the aspects of the task that are related to the objectives, however, as some research has observed

(Bailey, 2009; Sendziuk, 2010), the feedback given when using a rubric, especially the holistic rubrics, is sometimes very generic. Incorporating peer-assessment (Davies, 2002) or making students participants in the process makes the task focus on learning through feedback, rather than ticking the aspects achieved in the rubrics to measure performance.

In the current research, first, students were provided with a detailed rubric with the IB criteria for an oral activity (Appendix 6). However, the teacher came to realise that it was too specific, which made it difficult to implement and time-consuming (Vercellotti and McCormick, 2021). To address this, she produced a shorter one, without the more specific aspects of the criteria and the less relevant ones. Vercelloti and McCormick (2021) contend that the rubric should be of manageable size. According to Brookhart (2018, p. 5), only " descriptive language helps students envision where they are in their learning and where they should go next" and it is essential to "provide a clear description of the learning goal" (p. 5). In addition, the focus should be predominantly on what the learner has demonstrated, rather than what is missing. Analytic rubrics can assess language production in open-ended tasks, for which the learner is given the freedom regarding the vocabulary and grammar constructions to utilise. In this study, the teacher always encouraged the students to give their own comments and to use their own words when providing their feedback.

8. RELIABILITY AND VALIDITY

Validity and reliability of the gathered data are key in effective research, as Cohen et al., (2002) point out. They note that: "in qualitative data collection, the intensive personal involvement and in-depth responses of individuals secure sufficient level of validity and reliability" (p. 106). These two concepts were key not just to the design stage of the research, but also, to the data collection, analysis and interpretation. These were addressed in different ways: thirteen students gave comments to their classmates in their own classroom and were recorded and transcribed by the teacher. The activity, which the students found meaningful, lasted five months. Whilst the students were given independence so that they could participate in their own learning, the teacher was always there to support, sometimes intervening in order to contribute to the validity of the activity. The students at SEK-Alborán are used to being

recorded and were familiar with peer-assessment, hence, the whole process was natural, and they saw the activity as part of their learning process, being a meaningful activity for the vast majority of them from the very first day.

Validity is key to effective research (Cohen et al., 2002) as it ensures that the methods used accurately measure and relate to the research phenomena as well as what they purport to measure (Cohen et al., 2002). Kirk and Miller (1986) describe validity as "a question of whether the researcher sees what he or she thinks he or she sees" (p. 1). In qualitative research, where opinions and attitudes can be subjective and biased, it is important to minimise invalidity (Cohen et al., 2002). The researcher agrees with Hammersley (1992) observing that in a qualitative study the researcher might be confident in the results, however there is no certainty. In this study, the researcher judged that applying the coding method on the collected data from the interviews would produce valid results.

With regard to reliability in qualitative research, this refers to the link between recorded data and real-life events in the researched situation (Bogdan and Biklen, 1998). In this research, the interviews were not influenced by any external factors, but rather, they were spontaneous. The students were requested to follow the proforma as a guide, but they did not have to touch on every point and were allowed to give open comments to make the exercise more authentic, valid and enjoyable.

In this research, data reliability was prioritised by making detailed written notes of the research topic, and consistently categorising and coding the relevant data (Appendix 11). The data was broken down in parts that were given labels. And as Bryman (2016) indicates below:

> The analyst then searches for recurrences of these sequences of coded text within and across cases and also for links between different codes. Thus, there is a lot going on in this process: the data are being managed, in that the transcripts are being made more accessible than if the researcher just kept listening and relistening to the recordings. (p. 11)

The approach was also consistent because there was only one researcher undertaking not just the collection of data, but also, the recordings of interviews and transcriptions (Silverman,

2005). The researcher was trying to link the process of making sense of the data with the research questions provided. She faithfully transcribed what she heard, including student linguistic errors and major hesitations. She doubled checked these aspects to ensure maximum reliability. This has been described by Bucholtz (2000) as 'denaturalized' or 'full verbatim' transcription, which makes no adaptations and leaves everything in.

9. ETHICAL CONSIDERATIONS

In any social science research, there will be many ethical considerations. The main guidelines while conducting this research was never to violate human rights and to respect human dignity. Canvan's definition (1977) helps to explain the importance of ethics in social science research, principles that the researcher stood by throughout the current study:

> …a matter of principled sentitivity to the rights of others. Being ethical limits the choices we can make in the pursuit of truth. Ethics say that while truth is good, respect for human, dignity is better, even if, in the extreme case, the respect of human hature leaves one ignorant of human nature. (p. 810)

Firstly, research in education requires seeking consent from those involved, including with regard to the use of data. When researching with children, parents' consent is necessary, alongside the pupils' own agreement. As well as consent forms, verbal explanations were also provided to those involved on the reasons behind the study, and how the students' identities would be protected. During this study, the confidentiality of students and their identity had to be considered (Cohen et al., 2002). To ensure anonymity, the students' names have been omitted from the thesis.

Another variable needing to be considered when researching in one's own institution was allocating specific time for research purposes, which was separate to 'duty time'. In this case, all the collection of data was done in the classroom as this was a peer assessment activity planned for this year in preparation for the students' internal assessment. The rest of the

research process, including transcribing and coding, was performed at home. As it was felt that there was a constant need to step back and reflect, a research diary provided a useful aid, a sample page of which is included in Appendix 11. This research diary was helpful when researching in one's own institution. SEK-Alborán School always encourages teachers to undertake their own research, and so seeking permission from the institution did not present any issue. It was beneficial carrying out the study in the institution where the researcher worked, because she was able to collect data as part of her normal routine. In addition, the researcher was familiar with the school day and activities; the school calendar helped her to plan her own research timetable. With regard to her role as a teacher, it was not believed that this had any great impact on the students' responses and they were able to acknowledge both roles as a teacher and researcher. This enabled them to respond to the activity with their usual honesty and confidence.

10. CONCLUSIONS

In conclusion, the choice of methodology in this study has been described and justified, reflecting on why alternatives were rejected and asserting that the methods used were the most apt and fit for the purposes of this research. A fully detailed account of the process as it happened has been provided, including the tensions and difficulties and how they were managed. The coding process has been explained and samples included in the Appendix 10 for reader scrutiny. It has been explained that ethical issues relating to working with young people have been foremost in the researcher's mind throughout the study. In the next chapter, the interview data of the participants are analysed and discussed.

CHAPTER 4: DATA ANALYSIS

1. INTRODUCTION

In this chapter, the interview data of the thirteen students who participated in this study is analysed, focusing on the use of peer-assessment practices in the classroom and how these practices allowed the students to take an active role in assessment and build their own learning, as recommended by Topping (2009). The students, who aspired to different levels of achievement, performed an oral task in pairs for which they had to give each other feedback. They were paired with different partners in every stage without a specific criterion, mainly according to students' availability and willingness to participate, except for one case in which the researcher decided that two students reluctant to do the activity in the second stage should not work together again. Hence, some students participated more than others, providing feedback to a greater number of their peers.

With regard to allowing students to choose their own group members, Storch (2013, p. 163) describes the advantages: "Students choose work with peers with whom they are familiar, and this means that they may be more comfortable and willing to challenge each other's suggestions and offer repairs". To and Panadero (2019) note that pupils enjoy the independence and autonomy experienced during the process of selecting peer assessors, who were often either high-performing students or their friends.

Analysing the data was a critical part of this qualitative research, allowing the researcher to make sense of the data collected. Each interview was coded (Appendix 11) following a process of open coding, during which "data are broken down into discrete parts, closely examined and compared for similarities and differences" (Corbin and Strauss, 1998, p. 102). After reading each interview transcript, first, the insightful ideas raised were identified, particularly those either in agreement or in contrast to the researcher's previous research. Then, using Corbin and Strauss's (1998, p. 102) argument that "events, happenings, objects, and actions/interactions that are found to be conceptually similar in nature or related in meaning are grouped under more abstract concepts termed categories", the ideas were organised into themes, making notes in the margins. Grouping similar ideas under abstract categories was useful to the research and had "analytic power", with the "potential to explain and predict" (Corbin and Strauss, 1998, p. 102). Whilst there were many ideas and issues that were not directly related to this study,

throughout the coding and analysis the research questions were at the forefront of the researcher's mind. An example of the way in which the interviews were coded is attached in Appendix 11.

2. THE PROCESS OF DATA CODING

Qualitative data analyses (QDA) software is increasingly being used in qualitative research (André, 2020) and it is becoming more sophisticated. Kuckartz and Rädicker (2019) state that these computer programmes are efficient for analysis, visualisation and organisation of a large amount of data. However, André (2020) contends that they are only a support for the researcher as they do not do any analysis automatically and it is the researchers who have to evaluate the work intellectually. QDA programmes are not very flexible and might restrict or suppress creativity. In all qualitative research there is an artistic approach (Knoblauch, 2013; Strauss and Corbin, 1998). Using creativity in qualitative research is a way to gain knowledge and to present scientific research results (Gergen and Jones, 2008). QDA programmes have a limited scope that does not let researchers take various paths to analyse ideas. As a result, researchers cannot interpret analytically the interview extracts as they cannot be looked at in context. The researcher of this study rejected QDA since she wanted to discuss categories related to each other and not the categories identified by the program. As there were not a large amount of data, the researcher could analyse, visualise and organise it herself, interpreting the meaning of the interviews in depth.

During the process of data coding and analysis four emerging categories (EC) were identified, which were repeatedly reported by the students when giving feedback to their peers. The following list of emerging categories is not exclusive, but they are the most relevant ones with regard to the research questions and objectives:

1. EC 1: Students' emotional responses practising peer-assessment.
2. EC 2: Benefits for the peer assessed and for the assessors (peer leadership).
3. EC 3: Constructing their own learning.
4. EC 4: Teacher's perceptions.

Different colours were used for the different categories as follows:

EC 1: : Students' emotional responses practising peer-assessment: PINK

PA (Positive attitude)
LC (Lack of confidence)
R (Reluctance)
C (Competitiveness)
A (Affection)

EC 2: Benefits for the peer assessed and for the assessors: YELLOW

CF (Critical feedback)
P (Praise)
SF (Specific feedback)
BA (Benefits for the assessor)

EC 3: Constructing their own learning. BLUE

I (Interactions)
RL (Reflection of their own learning; students' acceptance of their own mistakes)
IC (Internalisation of the criteria)
T (Talking)
TM (Tense mistakes)
SA (Self assessment)
EQ (Effective questions)

EC 4: Teacher's perceptions. PURPLE

TP (Teacher perceptions)

The qualitative data for the main study was collected from semi-structured interviews, firstly, of the students with their classmates and then, of the students with their teacher. The

data analysis is presented as richly as possible with illustrative quotations from different interviews.

The data was coded in order to explore the meanings and subtexts with accuracy. A sample of the transcripts with codes is included in Appendix 11. In the following sections, specific extracts from the interviews have been selected to illustrate the key themes most effectively. Whilst in some cases there is an overlap between emerging categories, for clarity they are presented separately. For example, in the following sentence two different emerging categories overlap, namely 2 and 3, as the student gives specific feedback praising his peer and shows some internalisation of the criteria.

L6: *I think this audio is a really good example of what we were supposed to do (BA) You clearly explain the extract, you make a "magnificent" summary, and you explain why that extract is important to the understanding of the book. To me it has been a clear and excellent presentation of the book* (SF) (P)(IC).

3. EC 1: STUDENTS' EMOTIONAL RESPONSES PRACTISING PEER-ASSESSMENT

The responses of the students towards the peer-assessment task they had to perform in the class to prepare for their internal assessment were varied. Most reacted positively, however, some of them showed a lack of confidence in themselves and in their peers, whilst others exhibited a spirit of competitiveness. Some of their responses are quoted below.

3.1. Positive emotional responses

With regard to students' emotional responses, word frequency (Baayen, 2001) relating to positive attitudes (PA), while practising peer-assessment, was analysed. However, only L5 showed some excitement through words, while he was being assessed, saying: *Toma!!! (Yes!!!)* when he was given a good mark by the assessor. It seemed that he trusted his partner,

considered him to be sufficiently proficient and was excited to receive a good mark from him. This evidence is in contrast to the findings of To and Panadero (2019) – that engagement in peer assessment could be reduced due to distrust of the student's evaluative ability.

In this research, positive attitudes were expressed by five students when assessing their classmates. As Selcuk et al., (2019) observed, praise can improve students' mood and serve as motivation, encouraging them to work harder and perform better. In stage 1, L4 and L9 praised L13 when providing feedback to him:

L4: *Thank you for your recording.*
L9: *I've enjoyed so much your presentation.*

In stage 2 (Peer evaluations with proforma; parts 1 (presentation) and 2 (follow up questions about the presentation) of the interviews. In class intervention), L6 shows a positive attitude while giving feedback to L2. L6 was initially reluctant to take part in the activity. Her mother is American and so with her good level of English, she was able to perform to a high standard in English and expected by her peers to to do so. Whilst she spoke fluently, she made a lot of basic grammar mistakes, however, she was comfortable with L2 since he was not one of the best students in the class. L2 liked working individually and was not very keen to do this activity in pairs since he felt that he had a lower level of English than the rest. They were probably the two most difficult students in the classroom. According to Sargeant et al., (2019), shyness or the fear of appearing inadequate can limit the quality of the feedback. L6 uses some positive vocabulary to talk about what she liked and what L2 had done well:

L6: *He used very well vocabulary, he talked about the book, the extract and he also related identities with the characters of the book. And that's something that I really liked about that.*

Murillo-Zamorano and Montantero-Fernández (2018) stress the importance of the student being able to ask for clarifications and answer their assessor, during the feedback process. As noted by Kurt and Atay (2007), it is also important to provide constructive, positive feedback, which is never mocking or unkind and hence, might make students feel ridiculed. In stage 4,

L4 and L10 were required to give feedback to L5, while he was doing his interview with the teacher. There were some laughs from L4, which gave some humour to the conversation.

L4*: Then you invented some words such as bioligli.*
L5: *That doesn't exist?*
Teacher and L4: *Biologically*
L4: *jajaja*
L5*: I was close*
L4: *He invented "bioligli".*

There are more students who also laugh while giving feedback. For example, L12 laughs when giving feedback to L9. The word 'synonyms' being mispronounced causes laughter between the teacher and the students in the example below:

L12: *He talked about the extract the whole time. He knew what he was talking about. He needs more vocabulary, more synonyms* (mispronounced)
Teacher: *synonyms* (well pronounced)
L12: *ah synonyms* (well pronounced), *I thought it was synonyms* (mispronounced)

In reciprocal peer assessment, where students play both the role of assessor and student, providing emotional support is important to the feedback process (Boud, 1995). L4 giving feedback to L6 realises about the importance of making independent contributions using critical thinking. L4 also showed an interest in L6's points.

L4: *She made some points that she thought for herself which were really interesting.*
L4: *And that's it, so in general I think it was good, but I think there are things that should be improved and that's why I said this recording, so yea that's it.*

L4 understands that the feedback he is giving his classmate is meant to make him improve and he tries to motivate him, saying:

L4: *But there are some aspects which you could improve in order to make it as good as I know you can do it.*

However, their clear involvement and interest in providing feedback shows their positive attitude as well, for example:

L10: *You also use a lot of good grammar stuff, like adopting or rejecting new ideas or, I mean, everything was very clear, you used complex sentences and a lot of phrasal verbs, which was quite good. One thing I didn't like was that you used too much the eee.*

L11: *yes?*

L10: *And you say it quite a lot and also the word like*

L11: *Yes?*

L 10: *Yep, so you have to be very careful with that and the main problem with your first part is that you didn't give your opinion, just kept speaking about the extract and the book, which was very good, because it was very good, but you need to try to save just a few seconds to just give your opinion and you also did a very good job connecting the, you speak your speech with the topic, and saying the important parts and all of that, that was very good and, you actually compensate it that part of the opinion saying your opinion when she asks you the questions and going back to the topic, I mean, that we have seen in class you talk very well about racism and classism and selfishness and all of that and you connected it very well with the extract that you were talking about but one thing that you should have done, but this is just if you wanted to and if, that you could have talked how sexism took a very important role in that extract, the moment when you said that Trout Walker wasn't actually caring about what Kate was looking for her own, that's a thing you could have said, that men in the past didn't actually think what women thought just thought for their own and yea, the questions were very good because you actually improvise a lot at the beginning and you introduced all your answers with a well, or something like that, which gives you time to think and connected some topics that came in the questions with other topics like education, which was very good.*

Some studies of oral presentations have noted that positive attitudes to peer assessment have been a valuable source of learning (De Grez et al., 2012; Murillo-Zamorano et al., 2014). The examples of emotional responses above suggest that some students enjoyed giving and receiving feedback from their classmates. They were very much aware of what kind of feedback could be helpful and motivating for their classmates and found it easy to articulate. Such specific feedback is only given by someone who is really enjoying the activity and who

sees how helpful this is. Cheng et al., (2014) note that students with positive attitudes often approach peer assessment with more enthusiastic participation than those with less positive attitudes. Their enthusiastic attitude was measured by the researcher in terms of the quantity of comments from peers. The students in this study were providing feedback in their target language. In the case of L10, this helped his own language learning and made him a more autonomous learner. He seems to have interiorised the criteria.

3.2. Lack of confidence

As Fauzan (2016) observes, students will not talk, if they do not have any self -confidence. Students need to be confident in their abilities. If not, as Panadero (2016) states, they may struggle to deliver or use peer feedback. In this study, three students started to exhibit a low level of comfort when assessing their peers in stages 3 and 4, not completely trusting their abilities, for example:

L3: *I don't remember grammatical errors, but I am not a good corrector, so I probably skipped one.*

L3 thinks he is not good enough at English to be able to tell the grammatical mistakes that his partner made. As he is a student with one of the highest levels in the class, his lack of confidence might be more related to the activity itself, rather than his language skills. It is important to mention that this is a new student at school, and he is not familiar with any formative practices. He still thinks that providing feedback is just the teacher's job. Zou et al., (2018) observe that some students lack confidence in the language capability of their peers and themselves, thus feeling that assessment should be conducted solely by the teacher.

L4, on the other hand shows some lack of confidence, because he seems not to know the criteria:

L4: *And you use a lot of informal words, and I don't know if that's bad or good I just wrote it.*

L10 realises that he has given a lot of feedback to his partner, giving him suggestions to improve. Zou et al., (2018) found that the students who were concerned about damaging peer-to-peer relationships were more likely to support and help each other during the peer assessment. L10 is one these students and he probably doesn't want to sound too harsh, so he concludes by saying:

L10: *Just if you wanted to.*

Two peer-assessed students (L2 and L6) showed some reluctance when performing the activity. They both did not seem to enjoy the task, only doing it because all the others were. L2 did not want to be recorded; he felt that his oral skills were worse than those of his classmates and did not want to speak in front of them. L6 felt she should be among the best in the class, but as this was not the case she did not want to be assessed by her classmates. Their peers were also careful with their comments, trying not to say hurtful things, as can be observed in the following extracts:

L12 about L2: *I think what he was actually saying made sense, but I think the problems maybe was the tone he was using. He was so relaxed. He should focus more on his pronunciation.*

L4 about L6: *She wasn't in the mood of interacting.*

Both L2 and L6 gained some confidence in the stage 4 as they understood the purpose of the activity better and they had seen all their classmates doing the task.

3.3. Competitiveness

In stage 4, L4 and L5 have a conversation where there is a sense of competitiveness. In this case both have the same level of English and L5 does not particularly like to be given negative feedback. He tries to excuse his bad pronunciation saying that it does not interfere with

communication, which demonstrates that he has learnt the criteria well. However, he feels relieved when the activity finishes, as can be observed at the end of the following extract:

L4: *He invented "bioligli".*
L5: *But teacher that doesn't interfere with communication.*
Teacher: *It doesn't interfere, go on.*
L4: *And you didn't just as me, just like me you didn't use idiomatic expressions.*
L5: *That's false.*
L4: *When, when did you use some idiomatic expressions?*
L5: *I said, 'till the cause come home', I said that, come on.*
L11: *What's that? What does it mean?*
L5: *A long time, really long time.*
L4: *He didn't say it*
Teacher: *I think so, I will listen to that again anyway. I was looking for the questions, I didn't want to.*
L4: *In general, congratulations.*
L5: *I am free.*

The data indicates another case of competitiveness. However, this time L8, who has a lower level of English than L11, does not feel very comfortable being given feedback from her. Cestone and Levine (2008) observed this reluctance of lower-level students to speak in front of their peers. L8 did do the activity, but he did not enjoy it very much and made the following comment:

L8: *It is a proper name; I can say the way I want.*

This example shows L8's rejection of L11's feedback. L8 does not take negative comments very well as he usually thinks he does better than the really does. He is not very happy with the feedback given by L11 and he wants to show that he is better than everyone in the class thinks.

4. EC 2: BENEFITS FOR THE PEER ASSESSED AND THE ASSESSOR

There is extensive research finding that learners benefit from peer assessment, in both the assessor and assessee roles (Patchan et al., 2011). According to Jhangiani (2016), peer assessment in one particular assessment has a positive impact on performance in subsequent assessments. In this study, students were, for the most part, eager actively to participate, which is in contrast to the findings of Zou et al., (2018), who noted in their study of 234 Chinese undergraduates providing reviews for their peers the students' unwillingness to participate in activities involving peer assessment. This may be due to cultural reasons relating to the view of many Chinese students, who prefer the teacher in the assessor role. As expressed by Salaberri (1995), "The act of sharing information in the classroom and helping others to understand creates a feeling of success and reduces inhibition" (p. 5). She also noted that the more opportunities students are provided with to use the target language the broader their interactive experience and the more independent they become.

A crucial part of peer assessment is feedback, which, according to Tunstall and Gipps (1996), can be classified in two groups: evaluative and descriptive. In this study, three different types of feedback can be found: criticism, praise, and specific feedback. While the first two types are evaluative and focus mainly on conveying approval and disapproval and occasionally explanations to their peers as to why their answer is right or wrong, the third type is more descriptive and tells the students what they can do to improve. It is important to mention that this differentiation is not clear cut, and they sometimes overlap. These three different types of feedback are described below as they were found in the data: giving critical feedback, giving praise and giving specific feedback.

4.1. Giving critical feedback

The most common thing for students to do when giving feedback is to tell their peers what they have done, or not done. Only a few provided examples of what they could have done instead. Nine out of the thirteen students involved in this study made some critical comments

expressing what their peers did not do or did wrong. For example, in stage 1, L4, L10 and L9 criticise L13 as follows:

L4: *Sometimes you didn't mention really important events, which led to the extract which you were analysing.*
L10: *There are a few mistakes in grammar and vocabulary you should correct, like in the formation of the sentence 'I have chose" where past perfect was not correctly used.*
L9: *But there are some grammar mistakes, such as 'I have chose'.*
L3 and L2 criticise L4 thus:

L3: *He has used an informal register in words such as "gonna".*
L2: *The register is a little bit informal.*

L2, L9, L3 and L10 agree on the fact that L8's presentation was too long:

L9: *One of the main problems is that it lasts too long.*
L3: *But its length has bored me. The proportion between the extract and the whole summary is a bit 'decompensate'.*
L10: *Maybe it is too long and boring. I suggest you to use another structure, pronunciation and grammar, because you sound a bit like an Indian and you mix the verb tenses at the same time that you use vocabulary, which you confuse with Spanish words.*

None of this feedback includes what the correct way of using this tense is or which structure should have been used. However, the fact that two or three different peers gave the same feedback makes it more valid since, as MacMahon (2010) observes, the greater the number of people involved in the feedback process the more valid it is. It is important to mention that, at this stage, the students had not been given the proforma with the IB criteria yet, so they focused on what they thought was important. Van Zundert et al., (2010) note that many students have a better understanding of the difficulties faced by their fellow students in the classroom, and they are able, therefore, to articulate feedback in a more simple and concise way, an observation that emerged in this study.

In Stage 2 students were given the IB peer assessment proforma (See Appendix 12) and this led to more specific feedback from them. They focused on what they considered more important or on what drew their attention the most. Some examples are the following:

L10 to L8: *You use a lot of 'I mean'. You could have used a bit more vocabulary and grammar, like more interesting grammatical expressions.*

L2 to L6: *I think she didn't use a huge number of idiomatic expressions.*

L9 to L7: *Sometimes he doesn't relate it with the past and it's 192 pages so it's a lot of background behind that extract.*

L9 to L3: *He described the characters and the feelings, but not his opinion. Maybe sometimes could change the intonation.*

After considering that there were too many questions for the students to answer in the previous stage, some more specific questions were removed from the proforma for stages 3 and 4 (See Appendix 12). Moreover, for stage 1 the students had to focus mainly on engaging with the literary extract, so there were a lot of questions about the extract, which were removed as the students had to pay attention to many other aspects relating to the IB criteria. Hence, students found it easier to give feedback to their peers. However, it is interesting to point out that some students kept answering the questions removed from the first proforma, despite not being in the new proforma anymore. The students continued referring to the questions for criterion B related to the students' opinion on the characters, the event and the ideas and emerging categories of the extract. This might have been because the teacher had insisted on the importance of giving their opinion. The questions removed from all the criteria are indicated below:

1. Related to how well the student engaged with the literary extract (Criterion B1) all the following questions below were removed and became just one, namely: How well did the students engage with the literary extract?

- How did the students present their opinion on the characters?
- How did the students present their opinion on the event of the extract?
- How did the students present their opinion on the ideas and themes of the extract?

- How did the students make any connection with the themes studied in class? Identities, experiences, etc.
- How did the students centre the discussion on the extract and the work?
- Is the presentation consistently relevant to the literary extract and convincing?
- Does the candidate make effective use of the extract to develop and support observations and opinions?

2. Related to the language (Criterion A) the following questions were removed:

    - How easily can one work out what meaning is intended, or how much of an 'effort of translation' on has to make?
    - How rich or complex or subtle is the meaning conveyed- are there nuances or implications of irony?
    - How nuanced is the vocabulary?
    - Do pronunciation and intonation enhance communication?
    - Do they use idiomatic expressions?
    - Do pronunciation and intonation enhance communication?

3. Related to the message (Criterion B2) the following questions were removed:

    - How much does your partner contribute beyond a simple relevant response?
    - Are the responses broad in scope and depth?

4. Related to the Interactive skills (Criterion C) only part of a question was removed:
    - How much does your partner contribute to the flow of the conversation?

At stage 3, it is interesting to note how L7 and L11 both focus on interaction, which is one of the most important points in the IB criteria, when they give feedback to each other:

L7 to L11: *There is almost no interaction in the response given to the question, so that needs to be improved.*

L11 to L7: *It was only I asked him a question, he answered it and we moved on to the next one.*

At stage 4, L8 is again told by a different peer that '*he uses too many "I mean"'.* Most students continued to focus only on what attracts their attention the most. However, their feedback was always interesting and useful. L12 and L13 help L2 realise that he has to take the activity seriously. They are not very happy with his attitude, and they tell him so, as can be observed in the third extract below:

L10 tells L8: *You use a lot the word no, I mean it's always no no no. One thing I didn't like was that you used too much the 'eee'.*

L7 about L10: *There are some connections with the themes studied in class that may not be that clear.*

L12 about L2: *You should take it more seriously, because through all the interview Teacher: what happened throughout the interview?*

L13: *I think he has to take it more seriously.*

L2, as previously mentioned in this thesis, was probably the most reluctant student to participate in peer assessment. His classmates could clearly see that he was not taking the activity seriously and they considered that it was the most appropriate feedback to give him.

L9 made cautious and almost meaningless comments about L4, who is a student proficient in English. After every comment he said: *it's fine.* He may have been reluctant to give honest critical feedback in case it undermined his relationship with L4 (To and Panadero, 2019):

L9 about L4: *But there are sometimes he makes; he repeats so many times 'like' and also stays waiting instead of talking in some moments, but it doesn't interfere with communication, so it's fine and I think that the part he did the worse but it is fine is the last part because of the questions, because the questions were a bit difficult but he tried to answer the questions fast and giving some solutions like he starts but he didn't know how to finish, but in general is fine.*

In general, students acting as the assessors become more self-confident when judging the quality of work, according to McMahon (2010). L4 is more and more confident giving feedback to his friends and his feedback becomes more and more specific:

L4 about L12: *I thought you might need to be more like varied in your vocabulary because you very often said the same kind of words. You use complex grammatical structures, but sometimes I just noticed that you were low talking and sometimes maybe boring for the examiner, because they want you to go faster, because maybe you would express more ideas in the same amount of time.*

Critical feedback then, as it can be seen, can be relevant and focused when the students have a clear idea of the criteria. When the criteria are unclear, general and non-specific comments tend to be given, but in this study with one exception, feedback was readily given and received.

4.2. Praise

The data in this study contains forty-four praise comments. Only L8 did not make any praise comments and he did not make any critical ones either. Some examples of praise comments are the following:

L4: *It was actually pretty good.*
L11: *You did a great summary of the book itself and the extract you were talking about.*
L10: *You did a good job.*
L13: *Great job! Your presentation is almost perfect.*
L6: *You clearly explain the extract, you make a magnificent summary, and you explain why that extract is important to the understanding of the book. To me it has been a clear and excellent presentation of the book.*
L2: *Congratulations. I appreciate your great pronunciation. I also consider that the message is clear.*
L1: *Well summarised, good job!*
L3: *It was such a great recording!*

L5: *She developed her ideas really well and gave some contributions of herself.*
L7: *The questions are fully understood!*
L9: *It has been a good presentation. The summary of the extract was quite good.*
L12: *I think his ideas were very clear.*

Research suggests that students tend to prefer praise, which serves as encouragement and helps them to appreciate their own achievements (Ferguson, 2013; Walker, 2009; Weaver, 2006). Praise comments at stages 1 and 2 in this study were usually general and bland and did not provide useful feedback to the students. In fact, as Hattie and Timperley (2007) observe, they offered little to learning and could have even undermined the students' self-regulation. However, Selkuk et al., (2019) found that those who receive praise find it motivational, leading to improved mood and encouraging them to perform to an even higher standard. In this study, praise was often before or after criticism or specific feedback comments, even in stage one. An example of this would be the paragraph below:

L11 about L5: *The structure followed corresponds to the one we were supposed to do, and the message is clear enough. Your recording is fine, but for next time I will recommend you to spend more time explaining and introducing your extract, because it feels like we are missing some part of the information. Talking now about pronunciation, it's really good, but it would be better if you try for next time not to force too much your American accent even I know it's difficult. Well done!! Keep going!!*

This is in contrast to the findings of MacMahon (2010), that good students tend to receive praise and that feedback tends to ignore weaknesses. L13 started his feedback to L12 saying: *Your recording is awesome. Everything is clear and perfectly organised. Your pronunciation is very clear. If I have to point you some possible improvements, I would say that you have to speak louder.*

He tells his peer in a polite way that he has to speak louder, which is something that the teacher has also observed and has told the student about this many times. Taylor (2017) observes that feedback should focus on the positive more than the negative. In this data, these general praise comments are mixed with criticism and more specific comments:

L13 feedback for L1: *He didn't contribute to the flow of the conversation, but he answered the questions very well, so it is fine. He gets stuck from time to time, but it is normal, and it was for a short period of time, so great job.*

Dweck (2002) notes that praise could only be effective, if it involves "an interchange with the student rather than simply an evaluation" (p. 54). Praise linked to grades can be counter-productive to learning. Saddler and Andrade (2004) highlight the strengths of peer feedback – praise leads to motivation and additional information on weaknesses provides suggestions for improvement. Only at stage 4 in this study were students told to give their peers feedback as well as a mark according to the IB criteria. Conversations among the students in which the teacher sometimes participated demonstrated this interchange mentioned by Dweck. L10 giving feedback to L11 is an example of these conversations:

L10: Y*ou also use a lot of good grammar stuff, like adopting or rejecting new ideas or, I mean, everything was very clear, you used complex sentences and a lot of phrasal verbs, which was quite good. One thing I didn't like was that you used too much the eee.*
L11: *Yes?*
L10: *And you say it quite a lot and also the world like*
L11: *Yes?*
L10: *Yep, so you have to be very careful with that and the main problem with your first part is that you didn't give your opinion, just kept speaking about the extract and the book, which was very good, because it was very good, but you need to try to save just a few seconds to just give your opinion and you also did a very good job connecting the. You speak your speech with the topic, and saying the important parts and all of that, that was very good and you actually compensate it that part of the opinion saying your opinion when she asks you the questions and going back to the topic, I mean, that we have seen in class you talk very well about racism and classism and selfishness and all of that and you connected it very well with the extract that you were talking about ….*

L10 showed a good understanding of the feedback assessment practice and was able to articulate clearly views on future improvement of their peer's work. The semi-structured nature of these interviews allowed the students to interact and gave them the flexibility to ask questions and seek clarification.

The students used praise in a natural and spontaneous way, using simple unequivocal positive language in an encouraging tone. Some of the feedback was quite sophisticated, delivered in a feedback 'sandwich', for example, and the constructive, positive nature of the comments was noticeable.

4.3. Specific feedback

Formative feedback, in essence, provides advice on how to improve. However, to be truly formative, it should also cover how to go about this improvement. Wiliam (2011, p. 3) observes that "information does not become 'feedback' unless it is provided with a system that can use that information to affect future performance". He also notes that a strong formative assessment includes explanation and specific activities for the student to take on in order to improve. The feedback that some students gave their peers in this study fits under this description; it will be called specific feedback. All students gave some specific feedback to their peers, especially at stage 4, when they seemed to have understood the most important points in the IB criteria. However, it must be pointed out that L4, L9, L10 and L11 were the students who provided very rich feedback with more specific comments, as well as criticism and/or praise comments. It is interesting to see how L11 recognises the fact that L7 rambles on too much when answering the questions in the interview:

L11 to L7: *Maybe focus and not extrapolate the questions to other areas, to be straight, to answer the questions straight.*

L10 also gives very specific feedback to L8 telling him what he could have done better, with very specific examples, like how marginalised were these two people … and how they were different from the other kids:

L10 to L8: *You spoke a lot about the characters and how they evolved and you express your opinion about the characters, but you could maybe have expressed your opinion more about the event that happened during the extract and also the main ideas, for example how marginalised were these two people or the society they belong in and*

how were they different from the other kids and also about education and how that changes people, how the fact that Standley changes, eh, like teaches the alphabet and all those staff to Zero changes Zero and also Standley. That's it more or less, well you could have engaged a bit more the, the people you are talking to by connecting this extract to some personal experiences or some personal ideas, that's all.

L9 to L13: *When you are improvising and talking it is difficult to make complex grammatical structures, so I think that is a point he needs to improve. Also, like many times, you have to intonate or to make some changes to make the person is listening you like make it engage. I think that he has given some examples of other people, but also, his own example, for example, about climate change. He has to improvise and make the conversation more fluent.*

L4 for L5: *In the last part in the questions you were pretty good in mostly all of them, but in the one if we should be worried about technology it was a little confusing what you said, but then at the end you make yourself clear on your point and then some things you should improve is not saying my word: remarkable.*

L1 for L8: *There was no interaction, and this makes it a bit boring. The girl should have asked him more questions in between. He related everything to globalisation. His tone is monotonous. He understood the questions. The flow is not good, because he sometimes needs some time to think before he starts talking. It seems he doesn't know what to answer and he fills in the gaps and becomes repetitive.*

Many of the points made by the pupils above are sharp and redolent of the type of comments teachers make. After years of experiencing teacher feedback, these students have absorbed and are able to use the feedback discourse with ease. According to Gaynor (2020), students find the process of providing feedback more helpful than receiving it. The students in this research found that giving feedback improved their understanding of the task at hand. There are not many examples in the data that reveal students' perceptions on giving feedback and they were not asked about this. One example could be when L4 admitted not having used idiomatic expressions, reflecting that he should have done so, or when L12 learnt how to pronounce the

word ‘synonyms’ after being corrected by his peer. There is a great amount of research about knowledge construction in Higher Education (HE), especially in technology-based learning environments (Hadwing and Oshige, 2011; van Aalst, 2009). Some of the data in this research shows some co-construction of learning in terms of English language learning in a secondary school. The student below is trying to say that he has not thought about the use of idiomatic expressions and now he has learnt, thanks to his peer, that he must use them.

L4 tells L5: *And you didn’t just as me, just like me you didn’t use idiomatic expressions.*
L12: *Ah, synonyms (well pronounced). I thought it was synonyms (mispronounced).*

It can be clearly observed from the above data how the students were increasingly giving more specific feedback to their peers and, thus, improving their interview skills. Selcuk et al., (2019) define this concept ‘peer facilitation’. According to Boud et al., (1999) and as the data of this study demonstrates, peer learning is a “two-way reciprocal activity” (p. 3) that students learn from each other by sharing their knowledge and their experiences.

5. EC 3: CONSTRUCTING THEIR OWN LEARNING

Various studies have assessed the efficacy of peer assessment in allowing students to become more independent by using self-management skills to enable them to participate actively in their own learning (Jones, 2014; Topping, 2009). As noted by Fauzan (2016), formative assessment allows students to evaluate themselves and their peers, including them in the assessment process. This helps to progress their language skills and allows them to become more autonomous with regard to their own learning (Archer et al., 2007; Ion et al., 2019; Webb and Jones, 2009). In this study, a few indicators of this were identified:

1. Interactions.
2. Students’ acceptance of their own mistakes.
3. Internalisation of criteria.
4. Asking effective questions.

5.1. Interactions

As Coyle (2007) observed, students can learn to self-regulate by interacting with their peers and talking about their own learning. In this research, the students were encouraged by the teacher to talk and interact with each other, especially from stage 2, when they were already familiar with the criteria. The teacher's intention with her intervention was to highlight what had to be done in the interview to score highly. The students recognised the importance of interacting with their partner to get the highest score in the IB, which was also something the teacher reminded them of. It is observed that scaffolding is necessary at appropriate stages when researching students (van de Pol et al., 2010). Only four students (L1, L5, L6 and L8) did not give their peers feedback related to the interaction criterion.

At stage 2, only three provided this type of feedback and it became clear which students took the activity more seriously. The answers to the same questions, asked by the teacher, were different, for example, the question: *And was there any interaction, engagement in the conversation, or was it just questions and answers?* L2's feedback about L6 was vague as had been the norm with this student throughout the research: *L2: I think that the main part was a relation of questions and answers, but I think that there was a little bit of dialogue about the content of the extract.* L8, despite not giving L10 feedback to the interaction criterion interacts with L10 and his teacher providing his peer with some feedback:

L8 feedback for L10: *So, it was almost perfect, but you could have done some things better like for example, something that you did really well is reflecting about everything on the extract, but you could have explained more the ex tract, I mean, what happened before, what leaded to those circum stances. I mean, talking about the form of your speech, that's the only thing. And now grammatically, maybe some more complex sentences, but I mean, you used some of them, but I mean you can always use some more idiomatic expressions. That's something that, I mean, that's the biggest point and maybe some more specifically vocabulary about the extract about the book, but apart from that it´s quite perfect.*

Teacher's questions for L8:

Teacher: *Did you understand everything? And how did he answer the questions?*

L8: *Really good, like he did during his speech. He reflected on everything.*

Teacher: *And what about the theme?*

L8: *He talked about racism and things related to those camps, I mean, he reflected on every thing.*

Teacher: *And did he give his opinion?*

L8: *Yeah, he gives, I mean, during the questions and before the questions he gave his opinion*

L8 feedback for L10: It´s true that I may have done some questions to help you develop more your opinion, but I mean, you made your opinion with me not saying nothing so it's fine.

Teacher: *Ok, thank you.*

However, L11 and L3 were able to interact with their peers and they seemed to have learnt what was required in order to achieve a high score. In his influential study, Boekaerts (1995) showed that when students asked their peers their own questions they engaged more, as they felt responsible for their own learning. One example of this can be seen below.

L11 providing feedback about L5:

Teacher asks about interaction again: *And in the end when you started to ask him questions, was there any kind of interaction?*

L11: *Yes, but I don't remember what I asked.*

L5: *You asked about the importance about this extract at the end of the story.*

L11: *Ah, yea, I remember I asked him which like, if he would…*

L5: *(Interrupting) If I would change the end of the story.*

L11: *And like, by that, we established like a little conversation.*

L3 now distinguishes between an interview, where there are only questions and answers and an interview in which there is a natural dialogue or conversation. As the CEFR (2001) notes, talking is when the student listens and speaks with one or more interlocutors, thus creating

meaning and having a conversation. L3 had a conversation with the teacher about his interview with L9 in which both participated.

Teacher: *Were there any personal interpretations about the extract?*
L3: *Yes, he asked about how does the extract can be transformed into a normal situation. He asked me how could I solve that situation.*
L9: *Because he said that his extract was how he escaped from some places and I asked him, like, how did he solve that situation, if he was there.*

At stage 3, L4 also seems to have understood the importance of talking during the interview and he comments about L2:

L4: *About interactive skills was pretty good because we were talking to each other, we agreed and we disagreed in some aspects then, we had a great interaction between us. I think he understands the questions because he didn't have to think his question, he just went into it, states his point and tries to convince me about what he said being right.*

L9 and L13 also mentioned interaction in their feedback, but their feedback was vague. The students either did not distinguish where there was interaction in the interview or they did not know how to explain this interaction. L9 considered fluency important matter and for L13, a conversation is when both speaker and listener express ideas. Both are in some way right.

L9: *I think that we have interacted about and that's a good point, because like I said before it still makes the interviewer still want to listen and he has understood all the questions and answered with a fluent English and fast, so that's a good point.*
Teacher: Was it a real conversation between you two?
L13: *The first one was more a conversation in which both expressed ideas.*

At stage 4, five students mentioned interaction in their feedback to their peers (L7, L9, L10, L12, L13). However, their language was vague. Below, is an example from a segment of conversation between students in which, although they do not explicitly refer to interaction, they are interacting with each other, by giving and receiving feedback. From this example, one can see how students learn from each other, always conversing in the target language. As noted

by Zyoud (2016), students need to practise oral skills frequently at all levels of learning to develop fluency.

L5: *I said till the cows come home. I said that come on*
L11: *What's that? What does it mean?*
L5: *a long time, really long time*

L9 refers to fluency as something very important: *"He makes different intonations to make the conversation fluent and easy to understand and the pronunciation is also excellent, so it is quite easy to understand"* and L10 says*:"was quite fluid".*They have understood that a conversation must be fluid, and this is what attracts their attention the most. However, L13 and L4 focus more on what is said than how it is said – both points are in the criteria.

L13 says about L1: *His responses were not just responding to the question, but he has amplified with some real situations and with his opinion.*
L4 feedback about L12: *I think you interact well, because you were both asking questions, indirectly you were given questions to the air and they you were asking him more and more, and those were good questions for him to answer and to talk about them and to establish a conversation between both of you and I think that was really well, yea, that's it.*

Interaction is important in maintaining a conversation and it is one of the main points in the IB criteria. Peer assessment can be used to improve pupils' self-assessment skills, because it requires them to apply their knowledge of the criteria to critique each other's work, while justifying their judgements through conversation and dialogue (Nicol et al., 2014). It is interesting to see how the students in this research chose different aspects to provide feedback on, while interacting with their peers or the teacher: fluency, both interlocutors asking questions, making independent contributions including their opinions, expressing ideas, agreeing and disagreeing. Some of the feedback provided was positive and constructive and helped both the assessor and the assesee to reflect on their own learning and accept their own mistakes.

5.2. Critical reflection on their own learning

Ten out of the thirteen students in this research accepted their own mistakes and showed no fear of speaking. As stated by To and Panadero (2019), students who receive frequent marking and review practice are likely to use their experience of peer assessment to develop their own self-evaluation skills. In their study, they note that "through participating in peer assessment activities, the students could enrich understanding of quality, refine subjective judgement and deepen self-reflection" (To and Panadero, 2019, p. 931)

L1, L4, L5, L6, L10 and L11 in stages 2, 3 and 4 make comments that show their acceptance of their own mistakes either giving or receiving feedback. L5 admits to not having used any idiomatic expressions required to score highly in the IB, while L1 and L10 think they should have asked more questions:

L5: *I think I didn't use any expressions.*

L1: *I should have asked more questions so to contribute to the flow and have him to interact with my ideas.*

L10: *Maybe I should have included more questions in the middle, like trying to connect one question with another so that he could also give more thorough opinion on each topic and explain his experiences.*

L6 and L4 also reflect on their own learning and comment on L2:

L6: *I think he also committed the same mistakes as me, because we didn't use like a structure. The gap year told us to go from the book, a summary of the book, a summary of the extract, characters and opinions, and I couldn't stop to go from more to less. And I think neither of us did that, so I think that is a thing we should improve.*

L4: *There were no personal examples, there were interpretations of course, but it wasn't with personal examples, for example, I didn't do that too, and that's it.*

L11 competently provides feedback to peers, including L5:

L11 about L5: *Yes, yes that like to extend a bit more, because he didn't reach the four minutes. Like for example me, I extended a lot. He had to stop me because I was…. Yes,*

because doing it I should like time what I want to say and practice before because I over-talk.

Teacher: and you?

L5: *I think I need to spend longer in the summary, but the connection of the ideas and answering the questions was quite good.*

L11 about L7: *I gave him maybe five seconds to realise, because the questions were a bit strange, maybe I don't know for example the art one like I gave him 5 seconds to think the answer.*

Both L5 and L11 reflect on their interview, with a focus on time. They know they cannot exceed four minutes in their presentations; one has been too short and the other too long. It is true that L11 usually talks a lot, but the important thing is that she admits it in front of her partner. L5 also admits that his presentation was a bit short, but he wants to make clear that he did well answering the questions. Hopefully, this self-assessment will lead them to perform better next time, as is Brown and Harris's belief (2013), and also mine.

Only six out of thirteen students made self-assessment comments. This may have been because the task was about peer assessment and not self-assessment. However, these few examples illustrate how participating in peer assessment leads to self-assessment, which helps the students reflect on their own learning. The first student to provide feedback usually chooses the main theme of the conversation, as so happened with L11. She was aware that she had taken too long in the presentation, so when it came to providing feedback to L5, the first thing on her mind was time. These students feel comfortable sharing their 'weaknesses' with their peers, which makes it easier for them to be accepting, learning to shape these weaknesses into 'strengths'.

6. EC 4: TEACHER'S PERCEPTIONS

The teacher in this study mostly agreed with the students' feedback provided and received from their peers, in other words, there was large measure of agreement. There were only a few times when the students were not sufficiently involved, resulting in less useful feedback. The teacher intervened to try and reverse the situation, but it did not always work.

6.1. Agreement with students' views

Most students in this research mirrored the teacher from the first day. However, some tried to act independently and confidently providing their peers with more comments of their own. As Joo (2016) observes, open-ended comments offer an affordance for an extended amount of information. This is also the researcher's view in the light of the findings from the data.

Seven out of thirteen students started behaving in a teacherly way at stages 1 or 2. L1, L2, L5, L6, L7 and L8 behaved like their teacher throughout the research, barely providing comments of their own. They followed the IB criteria focusing on the areas the teacher usually prioritises. For the most part, these students did not include any oral task that the teacher had planned for them in class, like debates, or they simply did not participate in the conversations about the IB themes, where they were required to express their opinion. Two examples can be seen below that show how the students tried to imitate their teacher by focusing on relevant points in the IB criteria, such as the use of relatives as complex grammatical structures, the use of idiomatic expressions, the importance of having personal interpretations, the development of the questions and the fact that grammar mistakes are not so important, if they do not interfere with communication:

L1: *About the language he used an appropriate and varied vocabulary that has relationship with all the topic he uses correctly. Then, he used relatives and complex grammatical structures with some mistakes that did not interfere with communication. And he has a good pronunciation and intonation that also helps to the flow of the interview and his responses were relevant to my questions and he showed a lot of development. And then, my partner doesn't engage too much with the interpretations. He doesn't have personal interpretations.*

L1's language is similar to that of the IB criteria, as he uses 'appropriate and varied vocabulary', 'complex grammatical structures', 'mistakes which do not interfere with communication', 'development in the answers' and 'personal interpretations'. This is the vocabulary the teacher normally uses to provide them with feedback.

L8: *So, it was almost perfect, but you could have done some things better like, for example, something that you did really well is reflecting about everything on the extract, but you could have explained more the extract, I mean, what happened before, what leaded to those circumstances. I mean, talking about the form of your speech, that's the only thing. And now grammatically, maybe some more complex sentences, but I mean, you used some of them, but I mean (SF), you can always use some more idiomatic expressions. That's something that, I mean, that's the biggest point and maybe some more specifically vocabulary about the extract about the book, but apart from that it´s quite perfect.*

L8 knows the importance of complex sentences, idiomatic expressions and sophisticated vocabulary to score highly in the IB, so he focuses on those aspects. It is interesting how he starts saying that it is almost perfect and finishes saying that it's quite perfect, something that the teacher sometimes does as well.

L1, L2, L6 and L8 were some of the weakest participants of this research and the way they provided feedback sometimes caused tension in the interview. L2 undertook it at stage 4, but he did not wish to be recorded, with the situation being tense and uncomfortable for his peers and especially for the teacher. The feedback from his peers was limited, which chimed with the rather disrespectful attitude towards the teacher:

L12: *I think you should take it more seriously, because through all the interview, I think what he was saying actually made sense, but I think the problems maybe was the tone he was using. He was so relaxed. He should focus more on his pronunciation.*

L1 did not agree to be recorded until stage 4, whilst L5 and L7, despite being two of the best students, were also reluctant to perform the task. They did not often participate in oral tasks, because they believed their English to be good enough already. Some of the feedback

they provided for their peers, as seen below, was vague but correct. This reflects their lack of commitment to creating good feedback:

L5: *Very good, I liked a lot her summary before like, the context before her extract and then, she connected it well. Then, she gave her opinion on Zero, Standley and the group of kids that were with them.*

L5's feedback is very simple. He does not tell his peer what she could do to improve.

L7: *So, vocabulary, the range of vocabulary was very wide, so vocabulary doesn't need to be improved also complex grammatical structures were used and errors do not interfere with communication I think at any point and pronunciation and intonation they also don't interfere with communication and the responses are indeed relevant and they show development to questions made. I think they do include personal interpretations.*

L7's feedback is also very general, without any examples. Again, this student does not tell his peer what he might to do to improve.

L9, L10, L11 and L12 took the task seriously from the first day and most of the feedback they provided was of their own devising. As Cheng et al., (2014) note, often those students with positive attitudes participate more enthusiastically in the peer assessment activity. These students also made some comments similar to those of their teacher, especially at stages 1 and 2. They used connectors like the teacher and explained what their peers needed to do to improve. Moreover, like their teacher, they provided examples of the specific vocabulary that had been used:

L9: *First of all, it has been a good presentation, however, one of the main problems is that it lasts too long. Nevertheless, the summary of the extract was quite good, but there are some things missing in the summary of the book, maybe you need to improve with the time administration and some word pronunciation.*

L9 uses comments of his own devising, like the fact that it lasts too long. He also uses connectors like 'nevertheless' and tells his peer that he/she needs to administrate his/her time better.

L12: *He didn't explain much about the characters. He should have explained more his real opinion, rather than what it is already said in the extract. He should explain what happened before and after. His opinion on the event of the extract is very relevant, it is the clue to find the treasure. He didn't express his opinion on ideas and themes of the extract. About the language: he sometimes struggles to express what he wanted to say. He was looking for the word 'carry out' and he said 'do' instead. He could have used more specific vocabulary. He used: treasure, lipstick, camp, digging holes. It was easy to understand, but there were still mistakes, such as the s of the 3rd person.*

L12 provides feedback that tells his peer what to do to improve as well as what he did well. The teacher agreed with all the feedback provided. It is interesting how L12 realises that his peer was looking for the word 'carry out'.

L3, L4, L9, L10, L11 and L12 initially imitated their teacher when performing the task. The more they practised, the more confident they became. Some of the feedback given during the last stages was even more meaningful to the students than that provided by the teacher (Black et al., 2004). The teacher was surprised and pleased to see her students behaving so confidently and independently, acknowledging that students' comments were often actually more helpful to their peers than her own. This is in accord with AlFallay (2004) and Patri (2002) finding that teacher assessment can sometimes be replaced by PA.

The teacher could not have told L3 and L4 what she thought about their presentations in a better way than they did. L3 used the word '*messed around'* to describe how one student's presentation was not very clear or structured and that '*its length has bored me',* in other words, according to him, it was too long. L4 clearly told his peer to vocalise; something the teacher had already told him to do a few times.

L3: *He must improve his effectiveness, in other words, the number of words he needs to say something, because he has messed around a lot during the whole presentation.*

L3: *It is such a great recording, but its length has bored me. The details are very concise and precise so in that section your mark is the maximum. Nevertheless, the proportion between the extract and the whole summary is a bit 'descompensate'.*

L4: *Open your mouth when you talk.*

L4: *You located really detail the extract and you describe the character, the main character Standley and you site the extract in some occasions then you really related the extract with the next events happening in the book and then, you had really good English but sometimes you make grammatical mistakes. I think we really understood you. You had a great pronunciation and then in the last part in the questions you were pretty good in mostly all of them, but in the one if we should be worried about technology it was a little confusing what you said, but then at the end you make yourself clear on your point and then some things you should improve is not saying my word: remarkable and you use a lot of informal words and I don't know if that's bad or good I just wrote it.*

L4 provides specific feedback using his own vocabulary and specifying, with examples, how his peer did. L4 even refers to one question about technology.

Most students provided feedback in a natural way with their own 'affective' vocabulary: *it wasn't a big thing, quite good, all of that, really good, so it's fine, but in general is fine, maybe, he did a good job, that's it, I think it is complete, a lot of grammar stuff, it was perfect, almost perfect, and things like that, I've enjoyed so much your presentation, keep going, incredible result, it's too good, super clear pronunciation, such a great recording, more or less okay, that is a point that is so important*. As indicated by Salaberri (1995), "The language that the classmates produce helps to act as a bridge between the teacher and the individual knowledge of each learner". The students in this study understood each other well and did not want their peers to be discouraged by negative feedback. So they introduced, from time to time, their own encouraging words. For example:

L10: *I like that you spoke, you said a lot of phrasal verbs and you spoke about ethics and that's tough but, eah, everything was quite good. The only thing is that try to speak more naturally instead of always with the same tone, and with the same intonation and all of that.*

Students' own vocabulary as well as the fact that they sometimes compare their performance with their peers makes the peer-assessed student feel more comfortable, and this is something the teacher cannot do. For example:

L11: *Yes, yes that like to extend a bit more, because he didn't reach the four minutes. Like for example me, I extended a lot.*

All the students were familiar with the assessment language, as the teacher had practised it with them during other formative tasks about written work. L13's language, for example, was exactly as it appears in the criteria, however he added the last part with his own words. This last part was more helpful for the peer-assessed student:

L13: *I am going to talk about the internal assessment of L1 and I will give him in criterion A an 8/12, because the command of the language is effective and mostly accurate. In criterion B1, I will give him a 4/6, because the presentation is mostly relevant to the literary extract and in criterion C, I will give him a 4/6 again, because the comprehension and interaction are mostly sustained. To summarise it, he has done a great work, but in the criterion A he uses appropriate vocabulary, but he doesn't make so many complex sentences, but he tried, and he expressed the principal message correctly. He used a great intonation, maybe he talked about the story of the book too much, but it is quite good. His interactions with the teacher are very good and he answered appropriately the questions. He didn't contribute to the flow of the conversation, but he answered the questions very well, so it is fine. He gets stuck from time to time but it is normal, and it was for a short period of time, so great job.*

6.2. Divergence from students' views

At stage 1, most students focused on pronunciation when giving feedback to their peers, although the teacher had never particularly insisted on this aspect. They made comments such as:

I appreciate your great pronunciation
Your pronunciation is well enough
Super clear pronunciation
You have almost no mistakes in grammar, vocabulary or pronunciation
You need to improve with the time's administration and some word's pronunciation
I suggest you to use another structure, pronunciation and grammar, because you sound like an Indian and you mix the verb tenses
Talking now about pronunciation it is really good, but it would be better if you try for next time not to force too much your American accent even if I know it is difficult

The comments above were written by the students on OneNote and most of them were for L4, who has a strong American accent. Even though they were told that accent is not so important in the IB, they still wrote these comments. Being able to see their peers' comments before writing their own and not having been given the proforma yet, could be the reason there were so many comments about pronunciation at this stage, as they copied each other for the sake of security.

At stage 2, L1 and L4 performed the task, and gave each other very positive feedback. The teacher needed to probe them further in order for them to reflect on what they had not done so well. L2, who was reluctant to participate in the exercise, often provided students with vague feedback. The teacher's scaffolding sometimes helped the students to stay focused and encouraged them to reflect more on their peer's performance, thus providing better feedback. Teachers need to guide their students on the practice of providing feedback and self-assessment, as noted by Fauzan, (2016). For example:

L2: *Also consider that she spoke very deep about the content of the extract and also, she explained very well the psychologic filters of the characters.*
Teacher*: The psychology of the characters?*
L2: *Yes. And she also related the extract with the whole book. However, I think that L6 didn't express completely the whole content or the whole summary of the book. I consider that* L6 *was very good at answering questions, because she could relate the content of the questions with the extract and the whole book. And also, because she could use a good recovery to express the answers of the questions.*

Teacher: *So, in short, what advice would you give her for the next time?*

L2: *I would tell her that she shouldn't speak just of the extract, she should also relate the extract to the whole book.*

Teacher: *And was there any interaction, engagement in the conversation, or was it just questions and answers?*

L2: *I think that the main part was a relation of questions and answers, but I think that there was a little bit of dialogue about the content of the extract.*

The students occasionally did not provide any negative feedback or points for improvement. The teacher was then required to step in and encourage them to reflect.

L7: *So, vocabulary, the range of vocabulary was very wide, so vocabulary doesn't need to be improved also complex grammatical structures were used and errors do not interfere with communication I think at any point and pronunciation and intonation, they also don't interfere with communication and the responses are, indeed, relevant and they do show development to the questions made. I think they do include personal interpretations*

Teacher: *Could you remember any personal interpretation?*

L7: *No. There is almost no interaction in the response given to the question, so that needs to be improved and the questions are fully understood.*

Teacher: *What would be your best piece of advice to do better next time? I think that the only thing that needs development and improvement is the part of interaction after the question is made, so that's the only thing that needs some improvement, the rest of it is perfect.*

DeLuca et al., (2018) note that students value feedback from their teacher more highly than comments from their peers, and that there is a notable contrast in grades when assessing their peers. They suggest that students need to be taught about peer-assessment. Chiming with what Brown and Knight (1994) observed, the students in this study overmarked their peers, but only in the first stage, where they were still not very familiar with the criteria. However, at stage 4, four students slightly undermarked their peers, whilst three were accurate and only one overmarked his peer. It is interesting to note how L6, who was marked by two different peers, received a mark with only two points of difference. In general, the teacher was both pleased and surprised with the accuracy of the marks given by the students at this last stage. Moreover,

and contrary to some research (Langan et al., 2005), the teacher did not detect any biased assessment as a result of gender influences or lack of anonymity.

The data indicates that most students initially mirrored their teacher when providing feedback to their peers, mainly sticking to the set criteria. They sometimes needed some support from the teacher. However, when they give more independent feedback using their own encouraging language, this emerged as being more successful and meaningful for their peers.

7. CONCLUSIONS

The data has revealed an extensive use of affective language and encouragement, which led to students performing to a higher standard. They were concerned about damaging peer-peer relationships and always tried to motivate each other with their own positive language. This emotional support indicated that the students saw the activity as something helpful and enjoyed the process more so than when doing it with the teacher, whose language was not so affective.

The data has also shown that the teacher sometimes needed to scaffold her pupils' learning in peer assessment. Her interventions usually made students reflect more and focus on the important points in the IB criteria. However, pupils' answers sometimes demonstrated a great degree of independence. They undertook the task almost completely in the target language, English, with very few instances of needing to use Spanish. It seemed almost paradoxical that pupils had better understanding of what was the most effective feedback than the teacher who was accountable for their progress. Practice and support helped develop confidence and produce very effective feedback, which included a great amount of independent contributions. The teacher always encouraged critical constructive feedback and through practice the students learnt how to provide it, with both the criteria wording as well as their own words.

Finally, it is important to note that some students, finally, understood the importance of making independent contributions using critical thinking in reciprocal peer assessment. Whilst

the teacher had to remind some students to make independent contributions, others needed no such reminding and attempted to make them by themselves. Some were more successful than others, but they seemed to understand what was required in order to be successful in the interview.

CHAPTER 5: RESULTS

1. INTRODUCTION

Chapter five presents and discusses the research findings. Following critical discussion, this chapter provides responses to the objectives and the research questions.

The growing interest in EFL around the globe has called for learning opportunities that engage in learning English and motivate them to continue doing so (Vonkova et al., 2021). Motivation towards learning EFL is an ever-quickening demand in a world where English is fast becoming the lingua franca for effective communication in global trading, policy, and diplomacy is crucial (Woodrow, 2017). Students are particularly interested in learning English oral skills, so as to be able to communicate in these domains, however, the desire to learn EFL does not mean learner motivation is a given (Lamb, 2017). Students often lack confidence in their oral skills' abilities and in particular, worry about being compared or criticised by their peers (Ansari, 2015).

If learners anticipate gaining a bigger repertoire of material and symbolic resources that will lead to an increase in the value of their cultural capital and social power, then they will invest in the target language (Darvin and Norton, 2017). Instead of presuming that silent or disengaged learners are 'unmotivated', teachers should seek to find out "to what extent are students and teachers invested in the language and literacy practices of a given classroom and community?" (Darvin and Norton, 2016, p. 20). If the classroom practices are racist, sexist, or homophobic, then the learner may well show little engagement in language lessons, and thus, exhibit little progress, even if they are highly motivated (Darvin and Norton, 2021). Hence the importance of creating the right culture in the classroom where everyone respects each other's beliefs such that students feel confident to talk.

Classroom culture has been described variously, in several or different ways, as a feeling of comfort (Jones, 2015), a safe academic climate (Holley and Steiner, 2005) and a classroom "way of life" (Kafele, 2016). In general, the teacher aimed to provide a positive classroom culture, with Selcuk and Jones (2022) talking about creating a 'trustful affinity space' (p. 1). Most participants in the current research felt comfortable when communicating with their classmates and sometimes used informal language when assessing them. They gave and

received encouraging as well as critical feedback. The classroom atmosphere in this study was one of trust, friendship, humour and mutual peer respect, and this was paramount for this study, as it has been shown in the reported data.

2. REVIEW OF THE PURPOSE OF THE STUDY

The researcher of this study focused on oral peer-assessment practices in a secondary school in Spain, investigating challenges and benefits in terms of improving the language learning experience, which has been demonstrated is of outmost importance for teenagers' future lives. Throughout this study, research has been undertaken to address the following general objective: to raise teachers' awareness of how best to use peer assessment formatively during the teaching of oral skills in Diploma Programme language acquisition.

From this general objective described above, the following specific objectives and research questions (RQ) arise:

1. To explore how a group of 1st year Diploma IB (International Baccalaureate) students engage in peer assessment in an English as a Foreign Language (EFL) classroom in an independent school in Spain and the impact it has on classroom talk.

2. To study the students' emotional responses towards peer assessment of oral skills.

3. To examine the teacher's perceptions regarding the indicators of improved student performance.

4. To investigate the benefits of peer assessment for the assessed and assessor.

5. To look into how the practice of peer assessment can help students become more Independent learners, with the ability to build their own learning.

RQ1: To what extent do peer-assessment practices in an EFL classroom in an Independent School in Spain improve classroom talk?

RQ2: What are the students' emotional responses to practising peer assessment in oral skills?

RQ3: What are the teachers' perceptions about indicators of improvements in students' performance?

RQ4: How do the peer assessed or the assessor benefit from peer assessment?

RQ5: To what extent does peer-assessment help students to become more independent learners, able to build their own learning?

This chapter brings together the key findings derived from the data, providing responses to the objectives and the research questions.

2.1. To explore how a group of 1st year Diploma IB (International Baccalaureate) students engage in peer assessment in an English as a Foreign Language (EFL) classroom in an independent school in Spain and the impact it has on classroom talk

In chapter 2, the difference between speaking and talking was explored. A key objective of this study was to ascertain whether and if so, how, peer assessment could improve students' oral skills in a class of thirteen secondary school students. The students were following the IB curriculum in an independent school. The data from this small-scale study indicates that students improve their oral skills when they participate in an activity in which they assess their peers and are assessed by their peers, and when they participate seriously. With regard to this objective the researcher found five important points that have emerged from the data, as follows:

1. Peer assessment has been proven to be an effective means of encouraging discussion amongst peers, thereby improving performance (Johnston, 2015). The classroom is a space for the building of shared knowledge mediated by teachers and their interaction with students (Salaberri, 1995), moreover, together, all can co-construct knowledge (Tognini et al., 2010), as this study has shown.

According to Crichton (2009), students learn collaboratively when the teacher knows how to instigate them to talk. He drew this conclusion after observing a group of students in interactions in the target language with their teacher. This was successful because the teacher utilised strategies that encouraged the students to talk and to support each other in interaction. What is crucial, is that the teacher advances the right questions that avoid seeking to control their learning. In sum, the teacher needs to engage in pedagogy that promotes opportunities for students to engage collaboratively.

The evidence contained in the data in this qualitative analysis shows that despite during the first stage some students were a bit reluctant to give feedback to their peers, they finally took part when they saw that their peers used affective language in their comments and were trying to help by scaffolding the learning with each other. Most students improved in every stage, according to the feedback given by their peers. However, there were two students who showed very little improvement according to the feedback provided by their classmates. One was L2, who was reluctant to participate in the activity from the beginning, whilst the other was L8, who also was not very keen to join in. The researcher was not surprised by their attitude since these students had a similar attitude to other tasks performed in the classroom.

However, the data shows how L8 started to participate in Stage 2, interacting with his peer and generating a discussion with his teacher about his peer's performance, while his peer was also present. L8 tried to express what he thought about his peer's presentation, referring to the criteria: grammar, vocabulary, and content. The teacher sometimes interjected to ask him some questions that required his reflection on all the important aspects. It is interesting to note how he realised that he could have asked his peer some other questions that would have helped him develop more his opinion, however, his peer expressed his opinion without being asked about it, which indicates that he knew what he had to do. This comment: "*it is true that I may have done some questions to help you develop more your opinion, but I mean, you made your opinion*

with me not saying nothing so it's fine" shows how L8 was already reflecting on his own performance regarding the task and on his own learning.

2. Peer assessment gives students a sense of empowerment in the classroom with a voice and the power of influencing learning. This sense of empowerment helps to generate talking between students. Students little by little see that they are learning while enjoying being active learners playing the teacher's role (Black and Wiliam, 1998). According to Rotsaert et al., 2018) students are more likely to develop expertise for making sound evaluative judgements on peers' work, when opportunities for peer assessment practice opportunities are provided along with rubrics and appropriate questions for guidance. In this regard, their content analysis of the peer feedback messages elicited that students clearly informed their peers about what was wrong and why as well as putting forward suggestions as to how performance could be improved:

> Peer assessment practice in combination with clearly defined assessment scaffolds constitute a valuable classroom assessment practice since students experience a tangible educational value of peer assessment through the perceived and actual growth of their peer feedback skills. (Rotsaert et al., 2018, p. 268)

Clear instructions of what to do to improve were constantly seen in the data obtained from this thesis. Even during Stage 1, all students told their peers what, in their opinion, they could do to improve. The teacher had always told her students the importance of telling their peers what they needed to do next time to close the gap. Some examples of the vocabulary used were:

If I have to point you some possible improvements, I would say that you have to…
It would be better if you try for next time not to…..
You should / you should have……..
He could/ could have
Be careful with……..
You must
Maybe you need to improve

Instead of…….you could have
I just advise him to try to……..
If I had to give him a piece of advice……
I really encourage you to……..

Students' sense of language identity grows as they inhabit their feedback roles in engaging in another language and as competent users of English, thus they enjoy speaking in English. They forge their language identity through peer assessment which then develops into something much more fundamental. Learners renegotiate their language identity as they learn and use another language for their own purposes. This identity emerges from the language interaction in a particular social context (Bucholtz and Hall, 2005).

This study identifies manifestations of learner agency related to communicative participation, which is a quality of any foreign language teaching and learning approach aiming to support communicative competence development. In this respect, it is worth noting that several students talked about real situations conversations, for example:

L13 says about L1: *His responses were not just responding to the question, but he has amplified with some real situations and with his opinion.*

L4 feedback about L12: *I think you interact well, because you were both asking questions, indirectly you were given questions to the air and they you were asking him more and more, and those were good questions for him to answer and to talk about them and to establish a conversation between both of you and I think that was really well, yea, that's it.*

According to Brookhart (2017), whilst teachers may deliver better sustainable feedback than students, it is clear that the latter can provide their peers with explanations that are in accord with their level of understanding, something the teacher may find difficult to do. In fact, studies have found the effectiveness of peer feedback to be greater than that of the teacher for this reason (Ashton and Davies, 2015; Eksi, 2012; Ruegg, 2015). The data in this study revealed that the use of affective language by peers encourages students to try to improve, as noted by Carless and Boud (2018). Moreover, some comments made by their own peers were sometimes considered more useful than the teacher's. The progress that some students made was recorded

by the researcher, who wrote specific comments that the students provided for their peers in every stage.

Students are most motivated by successful experiences. In this study, some students participated more than others. It was noticed by Wu and Miller (2020) that occasionally due to lack of confidence in themselves and their peers, students lacked motivation to do peer assessment. Trust among students, and between students and their teacher, is, therefore, essential. The lack of motivation displayed by certain students was due to their own unwillingness to expose their lower level of English in front of peers. This was the case for L2 in this study, who participated but he did not involve himself like his classmates. However, the feedback given by his classmates helped him realise he needed to improve.

In order to improve student motivation, teachers could use achievable mastery learning goals and, through examples, emphasise the importance of feedback in reaching these goals (Carless and Boud, 2018). It was observed that praise could motivate students to perform better (Selcuk et al., 2019). Even students like L2 and L6, who were initially reluctant to talk in front of others, finally practised their English in pairs and, according to their peers' feedback, there was some improvement in their oral activity. The students need to form habits and develop fluency, thus requiring frequent practice at all stages and levels of learning (Zyoud, 2016). As the students in this study provided feedback in their target language, this helped them with their own language learning. In addition, they were exposed to listening to their classmates all the time, which also helped them to improve (Conti and Smith, 2019).

The study has demonstrated how students fully engaged with the oral activity and socialised with their classmates. Whilst the teacher was present, her role was that of observing, supporting and scaffolding, thus fostering development of students' skills. Rather than focusing on what she was putting into the teaching and learning process, she concentrated on what her students were obtaining from the activity. This led to increased student engagement and making the classroom processes more in line with student requirements (Wiliam, 2011). Students' talk increased and improved in every state as they came to feel more relaxed in the classroom, in particular, because they became increasingly aware that their classmates also made mistakes:

L6: *I think he also committed the same mistakes as me. I think neither of us did that, so I think*

that's a thing she should do

They also learned from their classmates what to do to improve. As the task progressed, they felt more confident with themselves that they were doing what they were supposed to do and started to enjoy talking to their classmates as well as giving them feedback.

3. Students in this study made grammar mistakes in their conversations giving feedback to their peers, however, these did not usually interfere with the communication. The study has shown that students do not need to be corrected by the teacher unless it affects communication (Emerson, 2004). They sometimes used Spanish words like "Toma!", which means "Yes!" to show excitement and as examples of translanguaging whereby students showed they were focused on trying to make meaning to enhance experiences, and to develop their own identity (Creese and Blackedge, 2015). As Wei (2018) points out, translanguaging is "a process of knowledge construction that goes beyond languages" (p. 15). Classes are still dominated by monolingual pedagogies despite the recognition in recent years of the importance of L1 in learning new languages as the data of this study has shown. It is generally asserted by the community of practice of language teachers that the target language should dominate but students should not be penalised for using their own language sometimes or for using 'Spanglish'.

Students corrected each other's grammar mistakes mainly during Stage 1, when they still had not been given the rubric with the criteria and they had to listen to the presentation on Flipgrid, for example:

L11: *you did a great summary of the book itself and the extract you were talking about. (P) However, it is true that you should pay attention to grammar mistakes as you talk, such as verb tenses (e.g., I have chose, instead of I have chosen)*

However, in the next stages students talked about grammar mistakes but they do not specify these:

L3: *I don't remember grammatical errors, but I am not a good corrector, so probably I skipped one.*

Other references to the register or language were made by some students when giving feedback, for example, the use of "camps", instead of "fields" (L8). However, there are not many words like this to be found in the fieldwork data although the angle of register is one to further pursue, since, in fact, the cited examples of register e.g. using "gonna" or the use of 'bro' is correct and authentic. These could have been acquired through more leisure oriented-sources, such songs or films or games outside the class (Lyrigkou, 2021).

L4: *You use a lot of informal words, and I don't know if that's bad or good, I just wrote it.*
L3: *He has used an informal register in words such as "gonna"*
L9: You nailed it bro

Also, the use of some informal discourse markers, such as "*I mean",* was considered by some students to be something incorrect in the school context, as they were assessing their peers based on the IB criteria.

L11: *You use too many I mean, like every time and many likes aaaand*

Providing feedback, the students in this thesis employed a high frequency and wide range of informal discourse markers, such as *I mean, yeah, like*. The teacher did not correct this use as it is considered an integral part of spoken discourse. They can lead to the establishment of a social rapport between speaker and hearer as well as being associated with successful spoken communication (Blakemore, 2002, Buysee, 2011). However, they are not deemed to have a place within an exam orientated EFL setting. However, the potential of discourse markers needs to be acknowledged, for students use these in creating successful talk, as the data in this thesis has shown:

L8: *Yeah, he gives, I mean, during the questions and before the questions he gave his opinion It´s true that I may have done some questions to help you develop more your opinion, but I mean, you made your opinion with me not saying nothing so it´s fine.*
L11*: Yes like, in that aspect he did it really well. Like, he talked about was, like the extract there was like two characters I think, three, and he talked about like every single one, the importance in the extract in the whole story and yes, I think that.*

2.2. To study the students' emotional responses towards peer assessment of oral skills

MacIntyre and Gregersen (2012) assert that positive emotions during language learning are not simply about feelings: learners experiencing the former are more cognizant of their classroom environment, becoming more conscious of language input and this results in their taking on board more of the foreign language. The findings of this study confirm the importance of practising peer assessment in secondary schools to improve foreign language oral skills. As discussed in the previous chapter, there was a significant emphasis in the data about relationships and the importance of using affective language when giving feedback. Some students provided specific guidance to their peers (Tasker and Herrenkohl, 2016). Of some surprise to this researcher, some of those comments were deemed more helpful than the teacher's comments. It was observed that students did not want to disappoint their friends (Zou et al., 2018) and therefore, gave them meaningful comments using very accessible language (Nicol and Macfarlane-Dick, 2006). This was possible because the teacher had worked to create a positive classroom culture. According to Dewaele, Gkonou and Mercer (2018) a good language teacher requires an acute awareness of this and the following qualities:

> Essentially, a good language teacher needs to be in a position to manage the emotional tenor of the classroom. This means not only should they be able to harness the emotions of their learners, but they should also be able to regulate their own emotions to ensure they are in the right frame of mind to create positive rapport with learners, generate enjoyment and manage any anxieties. (p.126)

The data in this study confirm that when teachers are able to regulate their own emotions and have an awareness of the emotional impact of TL efforts and use, they are more likely to manage the emotional tenor of the classroom, establish strong connections and teach with confidence, optimism and passion (Dewaele et al., 2018). The teacher in this study had to manage sometimes her emotions and show her confidence and passion about what she was doing in order to instigate her students to continue with the task.

In Chapter 4, extended accounts of the students' emotional responses were provided. The researcher was not sure about what the students' reactions to the task would be, because even though she had already done some peer assessment with the students, it had never been to assess oral skills. The atmosphere in the classroom was supportive, except for three students who reflected a low level of comfort when assessing their peers. Two of them lacked confidence in their abilities, and this prevented them from providing as much feedback as their peers, as Fauzan, (2016) also found in her research. One was new and had never done any kind of peer assessment and, hence their inadequate responses were unsurprising. If the students trust the teacher, they usually do what they are asked to (Carless, 2013) and thus, building it up is a crucial part of developing peer assessment, as noted by Erkan et al., (2021). Clearly, these two pupils needed more time and support.

Most students in this study reacted in a positive way and did not distrust their peers' evaluative ability. This study has shown that the affective language used by the students helped to create an atmosphere in which there even were some touches of humour and they seemed to enjoy the role of acting as a "teacher". They never provided comments that could hurt their peers' feelings as they did not want to damage their relationships with their classmates. There was only one case of defensiveness when a student received some critical feedback. The researcher had expected more of these cases due to the competitive nature of some of the students. This suggests the importance of having faith and confidence in others, and providing a supportive atmosphere (Carless, 2013) to enhance the students' uptake of feedback. This is in the hands of the teacher, who must be respected and trusted by her students to create the desired environment.

2.3. To examine the teacher's perceptions regarding the indicators of improved student performance

One of the most common AFL strategies widely employed in the classroom has been peer assessment (Yan and Boud, 2022). Feedback from peers has been identified as being a highly effective practice owing to the finding that this feedback improves learning and performance (Ion et al., 2019; Panadero et al., 2019). It can transform the role of students by encouraging

them to target, generate and interpret feedback when interacting with each other (Ion et al., 2019). Students focus on the contributions of their peers so as not to make the same mistakes again, thus meaning the teaching process is focused on them.The activities for this research led to the students delivering feedback to their peers as well as reflecting on their own input. By giving feedback and conveying their ideas, this means they became active learners and could transmit to their peers all their knowledge about effective speaking. Evidence has suggested that students do not value teacher feedback as much as that of their peers and during the process the teacher's role is changed to that of facilitator, rather than having active engagement (Fajar, 2019), and this was also observed in the current study.

Shen et al., (2020) elicited that peer assessment facilitates learner autonomy and enhances confidence in the ability to study, thereby resulting in an elevated agency and making advances in speaking proficiently. This research has provided evidence of the positive impact of PA on student learning. The findings have shown how the students' speaking performance improved progressively across the four different stages of the intervention. The students' feedback to their peers showed how they learned from them, internalised the designated criteria as well as adjusted their goals and strategies regarding speaking performance and assessment behaviour. Through their participation in assessment and taking up the responsibility to assist each other, they were actively engaging with their peers' learning (Yin et al., 2022). The result was student to student comments expressed in a dialogue of equals in a democratic and maningful way. Peer assessment promotes student collaboration (Topping, 2009), thus, as in the case of this research, enhancing the target language ability (Joo, 2016).

As it can be observed in this study performing an oral activity in pairs and even in threes or fours can save time. At stage four there were two students listening to the conversation between the teacher and another student, with them both giving some feedback. Moreover, when two or more candidates are assessed simultaneously by each other (Ducasse and Brown, 2009) time is also saved, thus, PA can save the teacher's time. The use of a technological tool like flipgrid might prevent listeners from losing focus (Liu et al., 2018) and the use of mobile applications also allows students to give feedback to more than one classmate. In conclusion, the assessment of oral skills must not be relegated as they are practical enjoyable skills. This study has shown that students can enjoy assessing their classmates, for example, when L4 and L9 were providing feedback to L13:

L4: *Thank you for your recording*
L9: *I've enjoyed so much your presentation*

In response to RQ 3, the researcher observed that half of the students became more confident as they continued to practise providing feedback. This confidence was visible through the quantity of comments provided and increased use of the criteria vocabulary. Using more fluent language, they also reinforced their feedback with examples and made sure to answer the teacher's questions. As aforementioned, it was interesting to note that in the last stages some of the students' comments were more meaningful to the students than the teacher's comments. Students used straightforward language as well as IB criteria language in a very harmonious way. They had not been taught how to do this by the teacher. Hence, this study has shown that secondary school students, with a little support and scaffolding, can quickly learn to peer assess in a positive, structured way.

Comparing their own performance with peers was evidently very helpful for the students as they felt more comfortable, which was demonstrated by the fact that they sometimes made very long feedback presentations and gave a great amount of feedback to their peers. In the first stages, the students gave more vague feedback and copied each other for the sake of security. They also mirrored their teacher's comments. However, the interventions of the teacher sometimes helped the students to reflect more, as was noted by Fauzan (2016), which allowed them to provide some independent feedback, thus resulting in more meaningful comments for their peers.

Contrary to some research (DeLuca et al., 2018), which noted that students usually tend to overmark their peers, the researcher of this study found that this happened only in the first stage. Moreover, she was surprised to see that the students were quite accurate by the final stage. According to Zhu and Carless (2018), teacher's input during the peer assessment process can have a positive impact. The teacher's scaffolding, it can be deduced, probably had some impact on this moderation of feedback.

2.4. To investigate the benefits of peer assessment for the assessed and the assessor

In this study, the researcher distinguished three different types of feedback: criticism, praise and specific feedback. Praise helps to create a respectful environment that is necessary in peer assessment and " when teachers create a comfortable and accepting environment that fosters student development, students' academic results improve" (Selcuk et al., 2019, p. 6). As this research reported, it was found to be combined with criticism and more specific feedback in a way that seemed to encourage students to perform better. As Brookhart (2017) asserts, both the receiving and the giving parties benefit when students give appropriate feedback. However, the literature fails to present a comprehensive picture as to how these two sides function differently, whilst serving interrelated and complementary objectives to facilitate student learning: "Few studies have clearly differenciated between the effects of assessing peers versus the effects of being assessed by peers" (Topping, 2010, p. 339). In this study the teacher perceived that although both phases contributed to their learning, more learning benefits were obtained from giving feedback rather than receiving feedback as they need to be proactive studying content and marking criteria and making judgements.

In response to this research question, it is considered that both the peer assessed and assessor benefitted from the task. Peer assessment can promote self-reflection and learning gains are not restricted to the assessed, instead the assessors can learn through assessing as it involves deep cognitive activities (e.g., evaluation, questioning) that can help them consolidate their knowledge and deepen their understanding of the subject matter (Panadero et al., 2018).

Students sometimes know better than the teacher the difficulties their peers might face and therefore, provide simpler and more comprehensible feedback (Van Zundert et al., 2010), so that the peer assessed benefits from it. When giving feedback during peer assessment, students can observe their peers' production and become aware of the elements of good practice. According to Marks, Meek and Blakemore (2016), by evaluating the work of their peers, students are exposed to solutions, strategies and perspectives that they would not see it otherwise. This mutual benefit was also observed in this study.

The assessors were given a peer assessment proforma, however, they used it selectively and focused on what they considered more important. When students act as assessors, they become very self-confident and provide more specific feedback. The researcher did not ask them what they thought had benefited them more, but the fact that they seemed to have absorbed their teachers' feedback discourse showed how helpful the process of providing feedback was. According to Gaynor (2020), it is more beneficial for the assessor than for the assessed. The participants of this study interacted with each other, and this contributed to the researcher's understanding that both assessor and assessed could construct knowledge regarding English language learning, as discussed in the previous chapter. However, the data clearly demonstrates that some students were more involved in the exercise than others and that, despite a willingness to participate, the students do not all collaborate equally and thus, the benefits are more nuanced in terms of reciprocal learning.

According to Ion et al., (2019) providing feedback helps students improve their learning. Moreover, students want to play an active role in their own learning and see their involvement in the design of teaching and learning processesss as being crucial in this regard. In this thesis, the teacher interacted with her students to promote positive connections with the set task, addressed doubts and clarified the received comments from the students. As a result, she obtained valuable insights into the effectiveness of her teaching.

2.5. To look into how the practice of peer assessment can help students become more independent learners, with the ability to build their own learning

The students in this study conversed with and listened to each other in the target language. The feedback they provided allowed both the assessor and assessed to understand their errors and reflect on their progress. This could be observed when the students recognised what they had done wrong and articulated what changes to their performance would have led to a higher score. It was also observed that the students compared their own mistakes with those of their peers. By comparing their own work with a few good examples, which come from the students' own performances, students can improve their own performance (Carless, 2020). When

students use task goals, examples of good work and the set criteria, they more effectively absorb the requirements of a task (Nicol 2009; Price and O'Donovan, 2006) and assess their own work more critically (Nicol, 2010). Whilst not all the students in this study made this kind of self-assessment as they were not asked to do so, some examples emerged that illustrate how participating in peer assessment can lead to effective self-assessment. Self-assessment, it has been shown, can result in better performance (Brown and Harris, 2013).

In addition, students demonstrate a higher level of independence through their comments (Molloy et al., 2020). As Kim (2009) noted in his work, learners are more reflective of the assessment and overall learning process when they have the space to reflect on the feedback. Kim also observed discussion between the assessor and assessed allows for new insights. In this study, in order to develop a level of responsibility among the students, they were provided with an opportunity to discuss their work and ask each other questions about the feedback. The meaning of the feedback was developed through this interaction between students, rather than being relayed directly from the teacher to student. The teacher only intervened occasionally to guide the students to focus on the most important points, but she barely gave them any actual feedback, thus providing an insight into classroom practices that could develop student self-regulation. Students had a sense of empowerment in the classroom with a voice and the wherewithal to influence learning.

Technology can provide motivational opportunities for collaborative peer learning and co-construction of knowledge (Hsu, 2019; Liu & Lan, 2016) as has been seen with the use of Flipgrid. This research has demonstrated that engaging with their classmates in a natural collaboration while assessing each other gives students the motivation to cooperate with each other. They become responsible of their own learning. Most students in this research mirrored the teacher from the first day and some were so motivated that tried to act independently, confidently providing their peers with more comments of their own. They understood sometimes their peers' feedback better than that of the teacher.

However, whilst motivation is important, it is not enough. Students have to be taught explicitly the skills of PA, especially in terms of how to assess their own work and that of other students by giving constructive feedback that enables them to move forward (Wiliam and Leathy, 2016). These are skills that will help them in the real world as they will always have

to self-assess and assess the work of others. The students in this study had had the same teacher for two years and they were familiar with the assessment vocabulary they had to use.They always had a rubric in front of them, which the teacher had previously explained. PA needs practice and this is shown in this thesis, with the students giving the best feedback in the last stage after having been practising it in the previous stages guided by their teacher.

It is important to recognise that some voices do express scepticism about the usefulness of peer feedback. They warn that it cannot be seen as a simple solution to improving language skills (Lam, 2010; Xu and Carless, 2017; Yu and Lee, 2016). The researcher has been aware of this scepticism and the complexity of the process throughout the study. The literature acknowledges that true knowledge construction is still unusual (Siquin et al., 2015) given that students face issues when seeking to regulate their own learning. However, it is interesting to note that there were some cases of co-construction of knowledge in this study in terms of English language learning, for example, when students interacted and discussed how the word 'synonyms' is pronounced or when they explained to each other the meaning of the idiom 'till the cows come home'.

The researcher of this study also noticed how peer assessment encouraged reflective learning through observing others' performances and becoming aware of performance criteria (Saito, 2008). Additionally, there was a focus on strengths and weaknesses. That is, the students conceptualized the feedback process in this way and it was anticipated that this would improve critical thinking skills and potentially their own learning process and autonomy.

3. CONCLUSIONS

In conclusion, the findings of this study have demonstrated three important points.

Firstly, the teacher's role was of utmost importance in the classroom, as she was able to manage her students' emotional responses, thereby facilitating their continuing with the task when reluctance appeared. Everyone respected each other's beliefs and did not make the sort of mocking comments that could make their peers feel uncomfortable or too unsafe to talk. The

teacher was successful at engaging in a pedagogy that promoted opportunities for students to engage collaboratively, thus creating the right culture.

Secondly, it was also observed that, when given the power, students are able to influence both, their own learning and their peers' learning, not just responding to the teacher. This study has shown how they made suggestions for improvement to their peers and how they co-constructed knowledge all together (Tognini et al., 2010).

Finally, by practising peer assessment students developed not only their linguistic skills and habits, but also their own language identity and personal growth. They were aware of the intercultural dimensions of language learning, understanding "otherness" through seeing culture through other eyes through language as Byram (2020) asserts. They developed and acted as intercultural speakers, who were positioned between the foreign and their own language and culture (Soler et al., 2007). The intercultural speaker is "a person who has managed to develop his or her own third way, in between the other cultures he or she is familiar with" (House, 2007, p. 19). As the findings in this study have shown, practising peer assessment provides a conduit for students developing their critical thinking skills and their own learning process and autonomy. "Learners need to be autonomous in order to able to learn independently" (Lamb and Reinders, 2005, p. 226). A peer assessment-rich classroom has been shown in this research to promote such learning.

CHAPTER 6: CONCLUSIONS

1. INTRODUCTION

Assessment for Learning, and especially peer assessment, has been demonstrated to be a very effective learning strategy to improve students' performance in speaking a foreign language (Black and Jones, 2006; Fauzan, 2016; Panadero and Brown, 2017; Topping, 1998). Whilst a whole host of studies have focused on peer assessment in higher education (HE) (Boud and Falchikov, 2005; Boud et al., 1999; Cheng and Warren, 2005; Langan et al., 2005), research on the secondary school context has been scarce (Araujo, 2020; Baran-Lucarz, 2019).

Interaction is one of the important factors stimulating student learning. Through peer interaction it is possible to maximise the effectiveness of peer assessment (Kim, 2009). The more students actively engage with the task, goals, criteria and exemplars, the more likely they are to internalise the requirements (Nicol, 2009; Price and O'Donovan, 2006). The meaning of the feedback is not only transmitted from teacher to student, but also, through interaction and dialogue amongst the students (Nicol, 2010). However, collaboration to improve understanding and to construct new knowledge is not easy to achieve (Kuhn, 2015). One reason for this is the fact that secondary school students are not always used to peer assessment. The researcher has been interested in promoting a greater awareness of and competence in peer assessment by school students.

To remind readers, in the first part of this thesis, the background to the study was outlined, including the researcher's personal interest in AFL. In Chapter 2, the differences between assessment and evaluation were explained and educational assessment was briefly considered, as well as the context of EFL assessment in Spain, specifically formative ways of assessing oral skills in an international school. The methodological framework was set out in Chapter 3, including the researcher's choice of a qualitative approach and her data collection strategy. Over the course of the academic year 2019-20, semi-structured interviews were conducted with thirteen students in an IB Diploma classroom. Chapter 4 presented the data, whilst Chapter 5 explained how, given the quality of the feedback of the participants, practising peer assessment in secondary schools to improve foreign language oral skills is important, thereby fulfilling the thesis objectives.

In this chapter, firstly, this work is considered in the context of existing theoretical knowledge and explains its contribution to the field. Secondly, the limitations and constraints of the study are discussed, including how the researcher dealt with these challenges. Finally, possible areas for further research are proposed.

2. PEDAGOGICAL IMPLICATIONS FOR PRACTICE

This study has enabled the researcher to consider peer-assessment in the context of a specific classroom and school culture. The researcher has an improved awareness of educational assessment and its impact on ongoing classroom and future learning. She has gained a deeper understanding of how peer assessment can lead to reflective learning. Most importantly, she has a better awareness of why and how to use peer assessment in the classroom, in essence, fully integrated into the language learning process.

A problem still remains around how to bring out successfully the positive impacts of PA and how to reduce students' hesitance and resistance towards this type of assessment. One method to resolve this could be through more intensive peer assessment practices. This includes allowing students to provide extensive feedback to peers (Panadero, 2016). Both teachers and students need to develop an understanding of respect during the assessment process. This will enable students to learn and it will prepare them for future working environments. Students who share a mutual trust with each other are more likely to be engaged in peer assessment and peer support, thereby gaining more out of the process (To and Panadero, 2019). Therefore, it is important that an environment of trust is supported in the classroom. This atmosphere will encourage students to accept any critical feedback in a positive and mature manner (Xu and Carless, 2017). While researchers can identify what can work in a classroom, the teacher is responsible for enabling it in practice (Black, 2015).

As Carless and Boud (2018) noted, through the use of attainable and negotiated learning goals and examples, the students' motivation can also be improved. More professional development that meets teachers' needs should be promoted (Lee and Wiliam, 2005) to convince teachers of the value of peer assessment. Through sharing skills and tried and tested

changes, teachers can improve their pedagogical content knowledge (Harrison, 2005). It is helpful for them to see successful examples of peer assessment being carried out in the classroom, for this will lend support to the professional learning process of teacher peer observation and feedback scenarios as well as teacher-led instructional coaching. A variety of approaches can help with this, for example videos of student peer assessment practices, teachers coaching peers, learning walks (teachers walking round classrooms) and teacher peer observation.

Moreover, teachers also need to understand that this is not something completely new, but rather, a different way to redistribute their time and effort where assessment is concerned (Chen, 2015). This will require a different way of thinking about their teaching practices, where the students play a fuller role in their own learning and engage in classroom assessment. A more collaborative and comprehensive teaching, learning and assessment language ecology must be put in place in the national programme, not just in foreign languages, but also, across the curriculum, thereby providing a more coherent learning experience for students.

While many pupils do achieve a high academic level, for each to reach their full potential and skills for independent learning, it is important that they practise regular self-evaluation and open dialogue with the teacher. In order to do this, teachers might use summative assessment for peer assessment, with a focus on feedback rather than marks. This approach could improve the students' competence and self-confidence, as well as their academic results. Black et al., (2003) regarding their research with teachers to develop formative assessment and their concerns with unavoidable summative assessment, state that:

> The overall message is that summative tests should be, and should be seen to be, a positive part of the learning process. Such tests should be used to chart learning occasionally rather than to dominate the assessment picture for both teachers and students. (Black et al., 2003, p. 56)

Such an approach challenges the view that formative and summative have different purposes and need to be kept separated (Black et al., 2003). They do not. Indeed, peer assessment can

work perfectly well with summative assessments in either the preparation stage or the feedback stage. It is necessary to adopt a new conceptualistion of classroom assessment. In broader terms, formative assessment becomes a pedagogical support and a necessity; and not an add on. It must be used to support and complement summative assessment. Teachers need to think 'outside the black box' (Wiliam, date) fixed habits and change their assessment procedures by introducing a more reflexive and fluid classroom atmosphere, where questions and questioning, feedback and discussion are predominant, rather than just teacher to students top down instruction. They need to find their ways of incorporating the lessons and ideas that are set in this thesis into their own pattern of classroom work.

PA needs to become the norm rather than the rule in secondary language classrooms. Teachers normally set the rules in a particular context, however, norms arise from the students' own feelings. For PA to be successful all the students in the classroom must agree about what they expect from each other and from the teacher.

As has been seen in this study, students sometimes understand each other's feedback better than they understand their teacher's. They can give their peers more meaningful and affective comments than the teacher and students will always listen carefully to what their peers have to say, if there is an atmosphere of trust. PA gives students the chance to seek clarifications, thus allowing the assessor and assessed to have conversations that can help them to construct knowledge about their English language learning. Students like acting as teachers and they learn even more by providing feedback than receiving it. They are honest and know how to tell their peers the truth, without hurting their feelings. They can encourage each other to try to improve, as was observed by Carless and Boud (2018), and this has also been seen in this study.

By using peer assessment as the norm, students will become active participants in their own learning, thus having a clearer understanding of what exactly they have to do to improve their English and achieve higher scores. It also motivates them to always try to do better as they know that they are constantly being assessed. Students must be assessed in their oral skills on a regular basis, but it is not possible for teachers to do this on their own. They can give regular oral or even written feedback for oral tasks, however, students will receive more feedback, if they are regularly assessed by their peers. To do this, students must be provided with samples, rubrics and clear instructions.

As noted by Jones and Coffey (2006), PA can be developed from an early age with success. This study has shown us that it is possible to undertake it in secondary schools and that shifting power to give pupils more agency in their learning can provide huge benefits beyond the task in hand, including greater self-confidence, improved critical thinking skills as well as belief in self-assessment and independent learning. The findings of this research have shown how students gained increasingly in confidence in assessing their peers and themselves, with even the most reluctant students eventually becoming enthusiastic in doing so. Teachers need to show students the methods to learn on their own as they will not always be able to be their guides. By focusing on strengths and weaknesses in the feedback process students can improve their critical thinking and become more autonomous learners.

This study has developed the researcher's critical reflexive skills. The different research tools utilised have advanced her methodological expertise and comprehension of the relevant theory. She also came to understand the issues one is faced with when carrying out research in the same school as you teach. She is intent on improving the teaching of foreign language oral skills and will continue to develop further peer assessment practices in order to optimise her students' learning independence. The students of her study responded well to the research task, thanks in large part to the trusting atmosphere the researcher had successfully built over the three years teaching these students. She has always believed in the importance of creating this atmosphere and her research reinforced this argument even more. Trust should be the first thing to be built in a classroom in order to practise PA with success

There is also a need for better communication between the school and parents, particularly given the importance of parental support for students. The school is responsible for promoting and explaining its teaching practices and policies to the parents. Parents may well be suspicious of peer assessment when their experiences are rooted only in teacher marks and possibly some feedback. By explaining fully to parents, in this way, and not totally excluding summative assessment, including grades, the school can gain their support and understanding. This is not a problem at Sek-Alborán school, where the parents' community is very supportive, as they know the school is consistently working to be innovative in the education it provides.

This study has been an extraordinary personal and professional journey, the findings of which, the researcher hopes can lead to improvement in the general understanding of assessment and influence the use of peer assessment in schools.

3. LIMITATIONS OF THE STUDY

As this study was conducted in one independent school in Spain, care must be taken when comparing the findings to the situation in other independent schools or state schools. That is, whilst the findings are noteworthy, they are not necessarily widely applicable. Thus future enquiry is needed applying the methods employed in this thesis to a wider sample across a range of school contexts and perhaps, different disciplines. By so doing, strengths and weaknesses of using peer assessment in secondary schools in Spain could be further elucidated.

While this research was conducted in a short period of time, the implemented procedure was fully and attentively completed. In this study, the group of students was small and with varying levels of English. There were thirteen students (eleven boys and two girls), the researcher, and a private school involved in the study. The size of this sample would ideally be extended to a larger group. The results of the research require further investigation, certainly with a larger sample and over a longer period of time as well as with different pupil populations and types of class. However, the findings of the study have demonstrated the potential impact and effectiveness of peer assessment in English classes in the context of an independent school in Spain that may, arguably, have currency in other similar and possibly, dissimilar schools, given the considerable evidence of developing formative assessment practices in schools in Spain (Panadero and Brown, 2017).

The data gathering process was not without errors and there are certainly changes that could be made to improve the validity of the research. For example, regarding the interviews, most of them were undertaken in the classroom, with the students in the room. There were two that had to be carried out through Microsoft Teams (an online platform) since they coincided with the lockdown due to the coronavirus pandemic. The researcher noted the advantages and disadvantages of each method, but a greater uniformity would have increased validity. There

were also a couple of times when the students had no time to give feedback to their peers on the same day and were forced to continue the exercise in the following lesson, thereby potentially losing the thread of reflection and continuity.

During the analysis of the data, the researcher realised that she could have asked the students directly whether they found giving or receiving feedback more useful in improving their oral skills. In addition, whilst she did carry out some research on how to motivate students to do peer assessment, some more research would have been useful to investigate inhibiting factors and to ensure that all students were equally involved.

Comparing the findings of this research with those of peer assessment in the students' mother tongue could prove beneficial. Moreover, extending the enquiry to oral contexts in the native language would shed light on whether students exhibit similar or significantly different strengths and weaknesses in both that and the target language.

4. FUTURE RESEARCH LINES: CONTRIBUTIONS TO KNOWLEDGE IN THE FIELD OF MFL METHODOLOGY

Most literature found by the researcher of this study was about PA in HE. This study has shown that with only a bit of support, secondary school pupils can learn quickly to provide feedback to their peers in a structured way. When in a trusting environment, secondary students can seemingly give more meaningful and useful feedback than their teachers, acting with independence and confidence. While there has been research finding that PA assessment motivates students to perform at a higher level (as reinforced by this study), there has been scant enquiry with regard to motivating students to take part in PA for the first time. This study has provided some insights into how to avoid the initial reluctance that some students have towards peer-assessment, such as building an atmosphere of trust in the classroom between the students and between the students and the teacher, as well as providing detailed explanations on the benefits of such assessment. However, further research is needed.

Giving students the freedom to ask each other questions for clarification helped to create a relaxed atmosphere, in which they sometimes used humour, for example, when a word was mispronounced. In this atmosphere, they also were able to co-construct some knowledge in terms of English language learning. On these occasions, it can be seen that the teacher barely intervened and the students were in a position to provide a type of feedback to their peers, that their teacher would not have been able to give them. The language was the students' language, 'their language', the language they could better understand. Moreover, they were reassured that they could always ask for clarification, if necessary. Further research is also needed that explores student assessment talk, with the aim of using it more extensively in the classroom.

Another beneficial research avenue would be comparing the thesis findings with the students' performance in other MFL IB classrooms other than English (e.g. French and German) or in the mother tongue (Spanish). Probing oral contexts in other languages and at varying levels would determine whether students have the same or different strengths and weaknesses in these different contexts.

Finally, use of peer assessment for oral skills in a 1st Bachillerato EFL class in a public high school and comparing the findings with those of the current thesis is proposed. This would provide more in depth understanding of the needs of high school students, in general, thus stimulationg changes to teaching practices and the adoption/development of appropriate resources for improving peer assessment of oral skills. Further, such inquiry, would help in identifying the personal developmet training teachers require for successfully implementing peer assessment in their classrooms.

CHAPTER 7: REFERENCES

Aben, J. E. J., Dingyloudi, F., Strijbos, J. W., & Timmermans, A. C. (2019). Embracing errors for learning: Intrapersonal and interpersonal factors in feedback provision and processing in dyadic interactions. In M. Henderson, R. Ajjawi, D. Boud, & E. Molloy (Eds.), *The impact of feedback in higher education: Improving assessment outcomes for learners* (pp. 107–125). Chan: Palgrave Mcmillan. https://doi.org/10.1007/978-3-030-25112-3_7.

Acher, A., Arcà, M., & Sanmartí, N. (2007). Modeling as a teaching learning process for understanding materials: A case study in primary education. *Science Education*, *91*(3), 398–418. https://doi.org/10.1002/sce.20196

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Research Association.

Agar, M. (1986). *Speaking of ethnography*. Beverly Hills CA: Sage Publications.

Ahmad, M., Qasim, H. M., & Khushi, Q. (2019). Specifying criteria for the assessment of speaking skill: A library-based review. *International Journal of Educational Sciences, 24*(1–3), 23-33. https://doi.org/10.31901/24566322.2019/24.1-3.1075

Ajjawi, R., Tai, J., Huu Nghia, T. L., Boud, D., Johnson, L., & Patrick, C. J. (2020). Aligning assessment with the needs of work-integrated learning: The challenges of authentic assessment in a complex context. *Assessment & Evaluation in Higher Education*, *45*(2), 304–316. https://doi.org/10.1080/02602938.2019.1639613

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, *14*(2), 115–129. https://doi.org/10.1093/applin/14.2.115

Aleksandrzak, M. (2011). Problems and challenges in teaching and learning speaking at advanced level. *Glottodidactica. An International Journal of Applied Linguistics*, *37*, 37–48. https://doi.org/10.14746/gl.2011.37.3

Alexander, R. J. (2006). *Towards dialogic teaching: Rethinking classroom talk*. Dialogos: York.

Alfageme-González, M. B., & Nieto Cano, J. M (2017). Los docentes de la enseñanza obligatoria en España y las actividades de formación continua. *Perfiles Educativos*, *29*(158), 148-165.

AlFallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self and peer assessment. *System*, *32*(3), 407–425. https://doi.org/10.1016/j.system.2004.04.006

Allal, L. (2016). The co-regulation of student learning in an assessment for learning culture. In L. Allal & D. Laveault (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 259-273): Springer, Cham.

Alonso, R. A. (2014). Teaching speaking: An exploratory study in two academic contexts. *Porta Linguarum, 22*, 145-160. https://doi.org/10.30827/Digibug.53717

Alqassab, M & Panadero, E. (2020). Peer assessment. In Brookhart, S. et al., (Eds.) *Routledge Encyclopedia of Education.* New York: Routledge.

Altbach, P. G., Reisberg, L., & Rumbley, L. E. (2009). *Trends in global higher education: Tracking an academic revolution*. Chestnut Hill, MA: Boston College Center for International Higher Education.

Amengual-Pizarro, M. (2006). Análisis de la prueba de inglés de selectividad de la Universidad de les Illes Balears. *Ibérica*: *Revista de la Asociación Europea de Lenguas para Fines Específicos* (AELFE), *11*, 29–59.

Amengual-Pizarro, M. (2009). Does the English test in the Spanish university entrance examination influence the teaching of English? *English Studies, 90*(5), 58-598. https://doi.org/10.1080/00138380903181031

Amengual -Pizarro, M. (2010). Exploring the washback effects of a high-stakes English test on the teaching of English in Spanish upper secondary schools. *Revista Alicantina de Estudios Ingleses*, *23*, 149-170. https://doi.org/10.14198/raei.2010.23.09

Amengual-Pizarro, M. (2019). Do prospective primary school teachers suffer from foreign language anxiety (FLA) in Spain?. *Vigo International Journal of Applied Linguistics*, 16, 9-30.

Amengual-Pizarro, M. & Méndez García, M.C. (2012). Implementing the oral English task in the Spanish university admission examination: An international perspective of the language. *Revista de Educación,* 357, 105-128.

Amir, A. (2015). Examining English-only of an EFL classroom of a Swedish school: a conversation analytic perspective. In C. J. Jenks y P. Seedhouse (Eds.), *International perspectives on ELT classroom interaction* (pp. 208-218). Londres: Palgrave Macmillan.

Andrade, H. L., & Brookhart, S. M. (2020). Classroom assessment as the co-regulation of learning. *Assessment in Education: Principles, Policy & Practice*, *27*(4), 350-372.

André, E. (2020). Reflections on qualitative data analysis software – Possibilities, limitations and challenges in qualitative educational research. *Revista Electrónica en Educación y Pedagogía*, *4*(6),41-55. https://revedupe.unicesmag.edu.co/index.php/EDUPE/article/view/141/403

Andujar, A., & Salaberri-Ramiro, M.S. (2019). Exploring chat-based communication in the EFL class: computer and mobile environments. *Computer Assisted Language Learning, 34(*4), 434-461. https://doi.org/10.1080/09588221.2019.1614632

Anghel, B., Cabrales, A. and Carro, J. M. (2016). Evaluating a bilingual education program in Spain: The impact beyond foreign language learning. *Economic Inquiry*, *54*(2), 1202-1223. https://doi.org/10.1111/ecin.12305

Ansari, M. S. (2015). Speaking anxiety in ESL/EFL classrooms: A holistic approach and practical study. *International Journal of Education Investigation*, *2*(4), 38–46. http://ijeionline.com/attachments/article/41/IJEIonline_Vol.2_No.4_2015-4-04.pdf

Araujo, J. C. (2020). Trying self- and peer-assessment with intermediate Spanish students of English. *Applied Linguistics Research Journal*, *4*(2), 40-53. https://doi.org/10.14744/alrj.2020.69885

Arnaiz Castro, P. (2018). El inglés como lengua extranjera en España. ¿En qué convocatoria superaremos esta asignatura? En M. I. Amor Almedina, R. Serrano Rodríguez y E. Pérez Garcia (eds.), *La Educación bilingüe desde una visión integrada e integradora* (pp. 51-64). Madrid: Síntesis.

Ashton, S., & Davies, R. S. (2015). Using scaffolded rubrics to improve peer assessment in a MOOC writing course. *Distance Education*, *36*(3), 312-334.

Baayen, R. H. (2001). *Word frequency distributions.* Dortrech, The Netherlands: Springer. https://doi.org/10.1007/978-94-010-0844-0

Babaii, E., Taghaddomi, S., & Pasmforoosh, R. (2016). Speaking self-assessment: Mismatches between learners' and teachers' criteria. *Language Testing*, *33*(3), 411-437. https://doi.org/10.1177/0265532215590847

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

Bachman, L. F. & Purpura, J. F. (2008). 32 Language assessments: Gatekeepers or door openers?. In B. Spolsky and F. Hult (Eds.), *Handbook of Educational Linguistics* (pp. 456–468). Malden, MA: Blackwell.

Bahadorfar, M. & Omidvar, R. (2014). Technology in teaching speaking skill. *Acme International Journal of Multicisciplinary Research*, *2*(4), 9-13.

Bailey, R. A. (2009). Undergraduate students' perceptions the role and utility of written assessment feedback. *Journal of Learning Development in Higher Education*, *1*, 1–14. https://doi.org/10.47408/jldhe.v0i1.29

Baird, J. A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: Fields apart? *Assessment in Education: Principles, Policy & Practice*, *24*(3), 317–350. https://doi.org/10.1080/0969594X.2017.1319337

Baran-Łucarz, M. (2019). Formative assessment in the English as a foreign language classroom in secondary schools in Poland. Report on a mixed-method study. *Journal of Education Culture and Society*, *10*(2), 309–327. https://doi.org/10.15503/jecs20192.309.327

Barbosa, M. de A., & Beserra, L. S. (2015). Formative assessment in the foreign language classroom. *Brazilian English Language Teaching Journal*, *6*(1), 100–109. https://revistaseletronicas.pucrs.br/index.php/belt/article/view/20200

Bartlett, M. (2018). Using flipgrid to increase students' connectedness in an online class. *ELearn*, *2018* (12), article 9. https://doi.org/10.1145/3302261.3236703

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

Benzies, Y. J. C. (2016). Spanish students' reflections on the importance of written and spoken skills in their EFL classes at high school and at university. *RAEL: Revista Electrónica de Lingüística Aplicada*, *15*(1), 1–20.

Berg, B. L., & Lune, H. (2009*). Qualitative research methods for the social sciences*. Boston: Allyn and Bacon.

Black, P. (1998). *Testing, friend or foe? The theory and practice of assessment and testing.* London: Falmer Press.

Black, P. (2003). The nature and value of formative assessment for learning. *Improving Schools*, *6*(3), 7–22. https://doi.org/10.1177/136548020300600304

Black, P. (2005). Formative assessment: Views through different lenses. *The Curriculum Journal*, *16*(2), 133–135. https://doi.org/10.1080/09585170500135889

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead: Open University Press.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, *86*(1), 8–21. https://doi.org/10.1177/003172170408600105

Black, P., & Jones, J. (2006). Formative assessment and the learning and teaching of MFL: Sharing the language learning road map with the learners. *Language Learning Journal*, *34*(1), 4–9. https://doi.org/10.1080/09571730685200171

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

Black, P., & Wiliam, D. (1998b). *Inside the black box: Raising standards through classroom assessment*. London, UK: Kings's Colleg London School of Education.

Black, P., & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal*, *29*(5), 623-637. https://doi.org/10.1080/0141192032000133721

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. https://doi.org/10.1007/s11092-08-9068-5

Blakemore, D. (2002). *Relevance and linguistic meaning. The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.

Blaxter, L., Hughes, C. & Tight, M., (2001). *How to research*. Philadelphia: Open University Press.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Essex, England: Harlow.

Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. *Teachers College Record*, *70*(10), 26–50.

Bloomaert, J. (2005). *Key topics in sociolinguistics.* Cambridge: Cambridge University Press

Boekaerts, M. (1995). *Motivation in education*. London. British Psychological Society.

Bogdan, R., & Biklen, S. K. (1998). *Qualitative research for education: An introduction to theory and methods*. Boston, MA, USA: Allyn and Bacon.

Bolívar, A. (2015). The comprehensive school in Spain: A review of its development cycle and crises. *European Educational Research Journal*, *14*(3–4), 347–363. https://doi.org/10.1177/1474904115592496

Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education*, *22*(2), 151-167. https://doi.org/10.1080/713695728

Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment & Evaluation in Higher Education, 24*(4), 413–426. https://doi.org/10.1080/0260293990240405

Boud, D. & Dawson, P. (2021). What feedback literate teachers do: an empirically derived competency framework, *Assessment & Evaluation in Higher Education*, 1-14. https://doi.org/10.1080/02602938.2021.1910928

Boud, D. & Falchikov, N. (2005). Redesigning assessment for learning beyond higher education. *Research and Development in Higher Education*, *28,* 34–41.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment and Evaluation in Higher Education, 31*(4), 399-413. https://doi.org/10.1080/02602930600679050

Boud, D., Lawson, R., & Thompson, D. G. (2015). The calibration of student judgement through self-assessment: disruptive effects of assessment patterns. *Higher Education Research & Development*, *34*(1), 45-59. https://doi.org/10.1080/07294360.2014.934328

Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, *41*(3), 400–413. https://doi.org/10.1080/02602938.2015.1018133

Bourke, R., & Loveridge, J. (Eds.). (2018). *Radical collegiality through student voice: Educational experience, policy and practice*. Singapore: Springer. https://doi.org/10.1007/978-981-13-1858-0

Bradbury, A. (2014). 'Slimmed down' assessment or increased accountability? Teachers, elections and UK government assessment policy. *Oxford Review of Education*, *40*(5), 610–627. https://doi.org/10.1080/03054985.2014.963038

Broadfoot, P. (1996). Educational assessment: The myth of measurement. In P. Woods, (Ed), *Contemporary issues in teaching and learning* (pp. 203-230). Buckingham, UK: Open University Press.

Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education. *Assessment in Education: Principles, Policy & Practice*, *11*(1), 7-26. https://doi.org/10.1080/0969594042000208976

Brookhart, S. M. (2007). Expanding views about formative classroom assessment: A review of the literature. In J.H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 43–62). New York: Teachers College Press.

Brookhart, S. M. (2017). *How to give effective feedback to your students*. Association of Supervision and Curriculum Development, Alexandria, VA.

Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education 3*(1-12). https://doi.org/10.3389/feduc.2018.00022

Brookhart, S.M., Moss, C.M., & Long, B. A. (2009). Promoting student ownership of learning through high-impact formative assessment practices. *Journal of Multidisciplinary Evaluation, 6*(12), 52–67.

Brophy, J. E. (2004). *Motivating students to learn*. New York: Routledge. https://doi.org/10.4324/9781410610218

Brown, A. (2005). *Interviewer variability in oral proficiency interviews*. Frankfurt: Peter Lang.

Brown, G.T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, *11*(3), 301-318. https://doi.org/10.1080/0969594042000304609

Brown, J. D. (2018). Rubrics. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 1436 – 1440). CA: SAGE.

Brown, G. T. & Gao, L. (2015). Chinese teachers' conceptions of assessment for and of learning: Six competing and complementary purposes. *Cogent Education, 2*(1). https://doi.org/10.1080/2331186X.2014.993836

Brown, G. T., & Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources heighten conceptions of assessment as school accountability. *Journal of MultiDisciplinary Evaluation, 6*(12), 68–91.

Brown, G. T. & Harris, L. (2013). Student Self-assessment. In J. McMillan (Ed.), The *SAGE handbook of research on classroom assessment* (pp.367-393). Thousand Oaks, CA: SAGE.

Brown, G., & Yule, G. (1983). *Teaching the spoken language: An approach based on the analysis of conversational English*. Cambridge: Cambridge University Press.

Brown, S., & Knight, P. (1994). *Assessing learners in higher education*. London: Kogan Page.

Bryman, A. (2016). *Social research methods*. Oxford: Oxford University Press.

Bucholtz, M. (2000). The politics of transcription. *Journal of Pragmatics, 32*(10), 1439–1465. https://doi.org/10.1016/S0378-2166(99)00094-6

Bucholz, M., & Hall, K. (2005). Identity and interaction: A sociolinguistic approach. *Discourse Studies, 7*(4-5), 585 – 613.

Bueno-Alastuey, M. C. (2013). Interactional feedback in synchronous voice-based computer mediated communication: Effect of dyad. *System, 41*(3), 543-559.

Bueno-Alastuey, M. C., & Luque Agulló, G. (2012). Competencias en lengua extranjera exigibles en la prueba de acceso a la universidad: Una propuesta para la evaluación de los aspectos orales. *Revista de Educación, 357,* 81-104.

Burgos B. (2017). *El bilingüismo en debate*. [Bacherlor's Thesis, Universidad de Valladolid]. UVaDOC Home. https://bit.ly/2kR9an6

Burner, T. (2016). Formative assessment of writing in English as a foreign language. *Scandinavian Journal of Educational Research, 60(*6), 626-648. https://doi.org/10.1080/00313831.2015.1066430

Burston, J. (2015). Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL, 27*(1), 4–20. https://doi.org/10.1017/S0958344014000159

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245-28. https://doi.org/10.3102/00346543065003245

Buysse, L. (2011). The business of pragmatics. The case of discourse markers in the speech of students of business English and English linguistics. *ITL: Journal of Applied Linguistics, 161*, 10-30.

Byram, M. (2020). *Teaching and assessing intercultural communicative competence: revisited.* Clevedon: Multilingual matters.

Cabrera, B. (2007). Políticas educativas en clave histórica: La LOGSE de 1990 frente a la LGE de 1970. *Témpora*, *10*, 147–181.

Cambridge University Press (2015). *Cambridge monitor 5 innovación en educación.* https://www.cambridge.es/nosotros/cambridge-monitor/innovacioneducacion

Canvan, S. (1977). Review of J.D. Douglas (1976) investigative social review: individual and team field research. *The American Journal of Sociology*, *83*(3), 809–811.

Caraker, R. (2016). Spain and the context of English language education. Research Bulletin, 92, 23-35.

Carless, D. (2012). *From testing to productive student learning: Implementing formative assessment in confucian-heritage settings*. London: Routledge.

Carless, D. (2013). Trust and its role in facilitating dialogic feedback. In D. Boud and E. Molloy (Eds.), *Feedback in higher and professional education: understanding it and doing it well* (pp. 100-113.). New York, NY: Routledge

Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice*. London: Routlege.

Carless, D. (2020). From teacher transmission of information to student feedback literacy: Activating the learner role in feedback processes. *Active Learning in Higher Education*, *23*(2), 143-153. https://doi.org/10.1177/1469787420945845

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315-1325. https://doi.org/10.1080/02602938.2018.1463354

Carter, R., & McCarthy, M. (2017). Spoken grammar: Where are we and where are we going? *Applied Linguistics*, *38*(1), 1–20. https://doi.org/10.1093/applin/amu080

Cattell, R. (1944). Psychological measurement: Normative, ipsative, interactive. *Psychological Review*, *51*(5), 292–303.

Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *83*(1), 1–6. https://doi.org/10.1080/00098650903267784

Cestone, C. M., Levine, R. E, & Lane, D. R. (2008). Peer assessment and evaluation in team-based learning. *New Directions for Teaching and Learning*, *2008*(116), 69-78. https://doi.org/10.1002/tl.334

Chang, C., Cheng, C., Fun, D., & Lin, J. (2012). The development of formative assessment practice through teachers' professional learning community and its impact on student learning achievement. *Bulletin of Educational Psychology*, *43*(3), 717–734.

Chang, S., & Lee, S. (2005). Learner involvement in self-and peer-assessment of task-based oral performance. *Language Research*, *41*(3), 711–735.

Chela, B. (2008). Hacia la optimización de la enseñanza de la pronunciación de un segundo idioma. En R. Monroy and A. Sánchez-Pérez (Eds.), *25 Años de lingüística en España: hitos y retos* (pp. 285–293). Murcia: Editum, Servicio de Publicaciones de la Universidad de Murcia.

Chen, J. (2015). Teachers' conceptions of approaches to teaching: A Chinese perspective. *The Asia-Pacific Education Researcher*, *24*(2), 341–351. https://doi.org/10.1007/s40299-014-0184-3

Cheng, K.H., Hou, H.-T., & Wu, S.Y. (2014). Exploring students' emotional responses and participation in an online peer assessment activity: A case study. *Interactive Learning Environments*, *22*(3), 271–287. https://doi.org/10.1080/10494820.2011.649766

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

Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing*, *22*(1), 93–121.

Cheng, Y., Horwitz, E. K., & Schallert, D. L. (1999). Language anxiety: Differentiating writing and speaking components. *Language Learning*, *49*(3), 417-446. https://doi.org/10.1111/0023-8333.00095

Cizek, G. J., Andrade, H. L., & Bennett, R. E. (2019). Formative assessment: history, definition and progress. In H.L. Andrade, R.E. Bennet, & G.J. Cizek (Eds.), *Handbook of formative assessment in the disciplines* (pp. 3-19). New York, NY: Routlege.

Clark, I. (2012a). Formative assessment: A systematic and artistic process of instruction for supporting school and lifelong learning. *Canadian Journal of Education*, *35*(2), 24–40.

Clark, I. (2012b). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, *24*(2), 205-249. https://doi.org/10.1007/s10648-011-9191-6

Clarke, D. (1991). Reviews. *ELT Journal*, *45*(1), 79–80. https://doi.org/10.1093/elt/45.1.79

Clarke, S. (2001). *Unlocking formative assessment: Practical strategies for enhancing pupils' learning in the primary classroom*. London: Hodder & Stoughton Educational.

Claxton, G. (1995). What kind of learning does self-assessment drive? Developing a 'nose' for quality: comments on Kalinowski (1995). *Assessment in Education: Principles, Policy & Practice*, *2*(3), 339–343. https://doi.org/10.1080/0969595950020307

Coffield, F., Ecclestone, K., Hall, E., & Moseley, D. (2004). *Learning styles and pedagogy in post-16 learning: A systematic and critical review*. London: Learning and Skills Research Centre.

Cohen, L., Manion, L., & Morrison, K. (2002). *Research methods in education* (5th ed). London: RoutledgeFalmer.

Coll, C., & Martín, E. (2014). Curriculum reforms and the field of curriculum in Spain in Pinar WF (Ed.), *International handbook of curriculum research* (pp. 445–465). New York: Tailor and Francis.

Conti, G., & Smith, S. (2nd August 2019). *Breaking the sound barrier: Teaching language learners how to listen.* The Language Gym. https://gianfrancoconti.com/2019/08/02/breaking-the-sound-teaching-learners-how-to-listen-why-we-wrote-it-and-what-it-is-about/

Convenio de colaboración entre la Universidad de Almería, la Universidad de Cádiz, la Universidad de Córdoba, la Universidad de Granada, la Universidad de Huelva, la Universidad de Jaén, la Universidad de Málaga, la Universidad Pablo de Olavide y la Universidad de Sevilla, para la acreditación de lenguas extranjeras. Firmado el 2 de Julio de 2011 y modificado por Adenda de 13 junio de 2016. [Archivo PDF]. https://www.us.es/sites/default/files/internacional/Convenio_refundido.pdf

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, *13*(1), 3-2. https://doi.org/10.1007/BF00988593

Correia, R. C. (2016). Assessing speaking proficiency: A challenge for the Portuguese EFL teacher. *E-TEALS*, *7*(1), 87–107. https://doi.org/10.1515/eteals-2016-0009

Council of Europe (2001). *Common european framework of reference for languages: Learning, teaching, assessment.* Cambridge: Cambridge University Press.

Council of Europe (2018). *Common european framework of reference for languages: Learning, teaching, assessment.* Companion Volume with New Descriptors. Strasbourg: Council of Europe. https://bit.ly/2MLNMdV

Coyle, D. (2007). Strategic classrooms: Learning communities which nurture the development of learner strategies. *Language Learning Journal*,*35*(1), 65-79. https://doi.org/10.1080/09571730701315774

Cozar, M. F., Acero, J., & Perez, F. (2020). The challenge of the skill of speaking in monolingual secondary education centres of Andalusia. *Encuentro Journal, 28,* 33-48. https://doi.org/10.37536/ej.v0i28.34

Craig, M. (2020). Engaging flipgrid: Three levels of immersion. In E. Alquarashi (Ed.), *Handbook of research on fostering student engagement with instructional technology in higher education*, (pp. 185–210). Temple University, USA: IGI Global.

Creese, A., & Blackledge, A. (2015). Translanguaging and identity in educational settings. *Annual Review of Applied Linguistics*, *35*, 20-35. https://doi.org/10.1017/S0267190514000233

Cribb, A. (2011). Integrity at work: Managing routine moral stress in professional roles: Integrity at work. *Nursing Philosophy*, *12*(2), 119-127. https://doi.org/10.1111/j.1466-769X.2011.00484.x

Crichton, H. (2009). 'Value added' modern languages teaching in the classroom: An investigation into how teachers' use of classroom target language can aid pupils' communication skills. *The Language Learning Journal, 37(*1), 19-34.

https://doi.org/10.1080/09571730902717562

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*(4), 438–481. https://doi.org/10.3102/00346543058004438

Cruz, J. M. (2016). The industry of testing and mutual recognition in southern Europe. *The Grove - Working Papers on English Studies*, *23*(11-22) . https://doi.org/10.17561/grove.v23.a1

Cullen, W. R., Shuker, D. M., Langan, A. M., Penney, D., Preziosi, R. F., & Wheater, C. P. (2008). Relationships between student characteristics and self-, peer and tutor evaluations of oral presentations. *Assessment & Evaluation in Higher Education*, *33*(2), 179–190. https://doi.org/10.1080/02602930701292498

Darvin, R., & Norton, B. (2016). Investment and language learning in the 21st century. *Langage et Société, 157*(3), 19-38.

Darvin, R., & Norton, B. (2017). Language, identity, and investment in the 21st century. In T. McCarty & S. May (Eds.), *Language policy and political issues in education*. Encyclopedia of language and education (pp.227-240). Switzerland: Springer.

Darvin, R., & Norton, B. (2021). Investment and motivation in language learning: What's the difference? *Language Teaching*, 1-12. https://doi.org/10.1017/S0261444821000057

Davies, A. (2002). Using student reflective self-assessment for awarding degree classifications, *Innovations in Education and Teaching International*, *39*(4), 307-319.

Davies, A. (2010). Test fairness: A response. *Language Testing*, *27*(2), 171-176. https://doi.org/10.1177/0265532209349466

Decreto 11/2016, de 28 de junio, por el que se establece la ordenación y currículo de la Educación Secundaria Obligatoria en la Comunidad Autónoma de Andalucía. *Boletín Oficial de la Junta de Andalucía*, 122, de 28 de junio de 2016. https://www.juntadeandalucia.es/boja/2016/122/2

De Grez, L., Valcke, M., & Roozen, I. (2012). How effective are self- and peer assessment of oral presentation skills compared with teachers' assessments? *Active Learning in Higher Education, 13*(2), 129–142. https://doi.org/10.1177/1469787412441284

de Saint Léger, D., & Storch, N. (2009). Learners' perceptions and attitudes: Implications for willingness to communicate in an L2 classroom. *System, 37(*2), 269-285. https://doi.org/10.1016/j.system.2009.01.001

DeLuca, C., Chapman-Chin, A. E. A., LaPointe-McEwan, D., & Klinger, D. A. (2018). Student perspectives on assessment for learning. *The Curriculum Journal*, *29*(1), 77-94. https://doi.org/10.1080/09585176.2017.1401550

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2011). *The sage handbook of qualitative research.* USA: SAGE

Department for Education. (2015). *Modern foreign languages, GCE AS and A Level Subject Content* [PDF file]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/485569/GCE_A_AS_level_subject_content_modern_foreign_langs.pdf.

Desforges, C. (1989). *Testing and assessment.* London: Cassell.

Dewaele, J. M., Gkonou, C., & Mercer, S. (2018). Do ESL/EFL teachers' emotional intelligence, teaching experience, proficiency and gender affect their classroom practice?. In Martínez AGudo, J. (eds), *Emotions in second language teaching* (pp. 125-141). Springer, Cham. https://doi.org/10.1007/978-3-319-75438-3_8

Distrito Único Andaluz. (2022). *Acceso a la universidad*. https://www.juntadeandalucia.es/economiaconocimientoempresasyuniversidad/sguit/?q=grados&d=g_b_procedimiento_acceso.php

Dixon, S. (2022, July 28). *Facebook: Global daily active users 2022*. Statista. https://www.statista.com/statistics/346167/facebook-global-dau/

Doody, O., & Noonan, M. (2013). Preparing and conducting interviews to collect data. *Nurse Researcher*, *20*(5), 28-32.

Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge: Cambridge University Press.

Drummond, M. J. (2011). *Assessing children's learning.* London: Routledge. https://doi.org/10.4324/9780203801338

Ducasse, A. M., & Brown, A. (2009). Assessing paired orals: Raters' orientation to interaction. *Language Testing*, *26*(3), 423-443. https://doi.org/10.1177/0265532209104669.

Dung, L. Q., & Dieu Ha, N. T. (2019). Portfolio—An alternative form of assessment in EFL context. *International Journal of Scientific and Research Publications*, *9*(1), 439-445. https://doi.org/10.29322/IJSRP.9.01.2019.p8557.

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessments: The limited scientific evidence of the impact of formative assessments in education. *Practical Assessment, Research, and Evaluation*, *14*(1), 1-11.

Dweck, C. S. (2002). Messages that motivate: How praise molds students' beliefs, motivation, and performance (in surprising ways). In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 61-87). Orlando: Academic Press.

Ekşi, G. Y. (2012). Peer review versus teacher feedback in process writing: How effective. *International Journal of Applied Educational Studies*, *13*(1), 33-48.

European Commission/ EACEA/ Eurydice. (2017). *Key data on teaching languages at school in Europe.* Luxemborg: Publications office of the European Union. Http://data.europa.eu/doi/10.2797/839825

European Commission/ EACEA/ Eurydice. (2017). *The Structure of the European education systems 2017/18: Schematic diagrams.* Luxemborg: Publications office of the European Union http://dx.publications.europa.eu/10.2797/2705.

Education First. (2017). *English proficiency index* [PDF file] https://www.ef.com/__/~/media/centralefcom/epi/downloads/full-reports/v7/ef-epi-2017-english.pdf

Education First. (2018). *English Proficiency Index.* http:// www.ef.no/epi/

Edwards, C. R., & Lane, P. N. (2021). Facilitating student interaction: The role of flipgrid in blended language classrooms. *Computer Assisted Language Learning Journal, 22(2)*, 26-39.

Efe, H., Demiröz, H., & Akdemir, A. S. (2011). A situated learning practice for language teaching classes: Teaching spoken English with authentic sketches. *Sino-US English Teaching*, *8*(9), 549-555.

Elwood, J. (2006). Gender issues in testing and assessment. In P. Murphy and K. Hall (Eds), *The sage handbook of gender and education* (pp. 87–99). London: SAGE.

Erkan, I., Unal, S., & Acikgoz, F. (2021). What affects university image and students' supportive attitudes: The 4Q Model. *Journal of Marketing for Higher Education*, 1-18. https://doi.org/10.1080/08841241.2021.1927933

Erreyes, H.M.B., y García, T.M.B. (2017). Las competencias linguísticas español-inglés en el sistema educativo ecuatoriano. *REFCalE: Revista Electrónica Formación y Calidad Educativa, 5*(2), 33-42.

Escudero, J. M., & Portela, A. (2015). La formación continuada del profesorado. Políticas, programas, aprendizajes docentes e impacto en la enseñanza y los aprendizajes del alumnado. En M.A. Flores y F. Ilídio (Eds.), *Formaçao e trabalho docente* (pp. 25-37). Santo Tirso: De Facto Editores.

Esfandiari, S., & Tavassoli, K. (2019). The comparative effect of self-assessment vs. peer-assessment on young EFL learners' performance on selective and productive reading tasks. *Iranian Journal of Applied Linguistics*, *22*(2), 1-35.

Esterhazy, R., & Damşa, C. (2019). Unpacking the feedback process: an analysis of undergraduate students' interactional meaning-making of feedback comments. *Studies in Higher Education, 44*(2), 260-274. https://doi.org/10.1080/03075079.2017.1359249

European Comission (2012). *Language competences for employability, mobility and growth, Strasbourg.*https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52012SC0372

Fajar, R. (2019). The effect of peer assessment towards students' speaking skill at semester I I of Stie muhammadiya Jambi 2018-19. *Celtic: A Journal of Culture, English Language Teaching, Literature and Linguistics, 6*(2), 26-32.

Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. London: Routledge. https://doi.org/10.4324/9780203220993

Fauzan, U. (2016). Enhancing speaking ability of EFL students through debate and peer assessment. *EFL Journal, 1*(1), 49–57. https://doi.org/10.21462/eflj.v1i1.8

Fergurson, M. (2013). Praise: What does the literature say? What are the implications for teachers? *Kairaranga, 14*(2), 35–39.

Ferguson, P. (2011). Student perceptions of quality feedback in teacher education. *Assessment & Evaluation in Higher Education, 36(1), 51-62.* https://doi.org/10.1080/02602930903197883

Fernández-Sanjurjo, J., Fernández-Costales, A. & Arias Blanco, J. M. (2019). Analysing students' content-learning in science in CLIL vs. non-CLIL programmes: Empirical evidence from Spain. *International Journal of Bilingual Education and Bilingualism*, *22*(6), 661–674. https://doi.org/10.1080/13670050.2017.1294142

Fielding, M. (2001). Beyond the rhetoric of student voice: New departures or new constraints in the transformation of 21st century schooling? *FORUM, 43(*2), 100-109. https://doi.org/10.2304/forum.2001.43.2.1

Fielding , M. (2004). Transformative approaches to student voice: Theoretical underpinnings, recalcitrant realities. *British Educational Research Journal, 30*(2), 295-311. https://doi.org/10.1080/0141192042000195236

Finlay, L. (2012). Five lenses for the reflexive interviewer. *The SAGE Handbook of Interview Research: The Complexity of the Craft, 2*, 317–333.

Fisher, D., & Frey, N. (2007). *Checking for understanding: Formative assessment techniques for your classroom*. Alexandria: Association for Supervision and Curriculum Development.

Foley, S. (2013). Student views of peer assessment at the international school of Lausanne. *Journal of Research in International Education, 12*(3), 201-2013. https://doi.org/10.1177/1475240913509766

Fraile, J., Panadero, E., & Pardo, R. (2017). Co-creating rubrics: The effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Studies in Educational Evaluation, 53,* 69-76. https://doi.org/10.1016/j.stueduc.2017.03.003

Fuentes-Diego, V., & Salcines-Talledo, I. (2018). Estudio sobre la Implementación de la Evaluación Formativa y Compartida en un Ciclo Formativo de Grado Superior. *Revista Iberoamericana de Evaluación Educativa, 11*(2), 91-112. https://doi.org/10.15366/riee2018.11.2.005

García Laborda, J. (2006). Analizando críticamente la selectividad en inglés, Todos los estudiantes españoles tienen las mismas posibilidades. *Tesol Spain*, *30*, 9–12.

García Laborda, J. (2012). De la selectividad a la prueba de acceso a la universidad: Pasado, presente y un futuro no muy lejano. *Revista de Educación*, 357, 17-27. https://www.researchgate.net/publication/264240647_De_la_Selectividad_a_la_Prueba_de_Acceso_a_la_Universidad_Pasado_presente_y_un_futuro_no_muy_lejano

García Laborda, J., Díez-Bedmar, M. B., & Martín Monje, E. (2018). Differences and similarities between oral and written competence in spanish pre-university students: A correlational study. *Onomázein Revista de Lingüística Filología y Traducción*, *42*, 172–196. https://doi.org/10.7764/onomazein.42.05

García Laborda, J., & Fernandez Alvarez, M. (2011). Teachers' opinions towards the integration of oral tasks in the Spanish university examination. *International Journal of Language Studies*, *5*(3), 1-12.

García Laborda, J., & Litzler Jerman, M. (2015). External validation of the foreign language speaking tasks of the High School Leaving Exam. *International Journal of Language Studies*, *9*(2), 97–112.

García Sampedro, M. (2021). Teaching spoken English in Spain: Classroom practices and teachers' perceptions. *Journal of Education Culture and Society*, *12*(2), 583-596. https://doi.org/10.15503/jecs2021.2.583.596

Garnet, J., & Cavaye, A. (2015). Recognition of prior learning: Opportunities and challenges for higher education. *Journal of Work-Applied Management, 7*(1), 28-37. https://doi.org/10.1108/JWAM-10-2015-001

Gaynor, J. W. (2020). Peer review in the classroom: Student perceptions, peer feedback quality and the role of assessment. *Assessment & Evaluation in Higher Education*, *45*(5), 758–775. https://doi.org/10.1080/02602938.2019.1697424

Gee, J. P. (2018). Affinity spaces: How young people live and learn on line and out of school. *Phi Delta Kappan*, *99*(6), 8-13. https://doi.org/10.1177%2F0031721718762416

Gelle, J. (2013). *Bilingüismo: Estudio de casos en centros bilingües de educación secundaria en la provincia de Almería* [Archivo PDF] http://repositorio.ual.es/bitstream/handle/10835/1983/863.pdf?sequence=1

Gergen, M., & Jones, K. (2008). A conversation about performative social science. *Forum Qualitative Social Research, 9*(2). https://www.qualitative-research.net/index.php/fqs/article/download/376/819?inline=1#gcit

Ghaicha, A. (2016). Theoretical Framework for Educational Assessment: A Synoptic Review. *Journal of Education and Practice, 7*(24), 212-231.

Gipps, C., & Stobart, G. (1993). *Assessment: A teachers' guide to the issues.* Hodder & Stoughton: London

Gleeson, D., & Knights, D. (2006). Challenging dualism: Public professionalism in 'troubled times'. *Sociology*, *40*(2), 277–295. https://doi.org/10.1177/0038038506062033

Goldberg, G. L. (2014). Revising an engineering design rubric: A case study illustrating principles and practices to ensure technical quality of rubrics. *Practical Assessment, Research & Evaluation, 19*(8), 1-12. https://doi.org/10.7275/vq7m-e490

Goodman, J. (2011). *Assessment practices in an independent school: The spirit and the letter* (Thesis). Kings College London.

Gordon, E.W. (2020). Toward assessment in the service of learning. *Educational Measurement: Issues and Practice*, *39*(3), 72-78.

Gordon, E.W., McGill, M. V., Sands, D. I., Kalinich, K. M., Pellegrino, J. W., & Chatterji, M. (2014). Bringing formative classroom assessment to schools and making it count. *Quality Assurance in Education*, *22*(4), 339–352. https://doi.org/10.1108/QAE-07-2014-0034

Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal*, *115*(4), 523-547. https://doi.org/10.1086/681947

Greenstein, L. (2010). *What teachers really need to know about formative assessment*. Alexandria, Virginia: ASCD.

Gregersen, T. S. (2003). To err is human: A reminder to teachers of language-anxious students. *Foreign Language Annals, 36(1), 25-32.* https://doi.org/10.1111/j.1944-9720.2003.tb01929.x

Grenfell, M., & Harris, V. (1999). *Modern languages and learning strategies: In theory and practice*. London: Routledge.

Grimm, N. (1986). Improving students' responses to their peers' essays. *College Composition and Communication*, *37*(1), 91-94. https://doi.org/10.2307/357386

Hadwing, A., & Oshige, M. (2011). Self-regulation, coregulation, and socially shared regulation: Exploring perspectives of social in self-regulated learning theory. *Teachers College Record*, *113*(2), 240–264.

Hafner, J., & Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education*, *25*(12), 1509–1528. https://doi.org/10.1080/0950069022000038268

Hall, K., & Burke, W. (2003). *Making formative assessment work: Effective practice in the primary classroom.* Buckingham: Open University Press.

Hamodi, C., López-Pastor, V. M., & López-Pastor, A. T. (2017). If I experience formative assessment whilst studying at university, will I put it into practice later as a teacher? Formative and shared assessment in Initial Teacher Education. *European Journal of Teacher Education*, *40*(2), 171-190.

Hammersley, M. (1992). Some reflections on ethnography and validity. *International Journal of Qualitative Studies in Education, 5*(3), 195-203. https://doi.org/10.1080/0951839920050301

Hamzah, M. H., & Ting, L. Y. (2009). *Teaching speaking skills through group work: A case study in SMK Damai Jaya* [Thesis, Fakulti Pendidikan, University Teknologi Malaysia]. http://eprints.utm.my/id/eprint/10255/2/Lu_Yee_Ting.pdf

Hanson, F. A. (1993). *Testing Testing*. Berkeley: University of California Press.

Harlen, W., & James, M. (1997). Assessment and Learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, *4*(3), 365–379. https://doi.org/10.1080/0969594970040304

Harmer, J. (1991). *The practice of English language teaching*. London: Longman.

Harris, M., & McCann, P. (1994). *Assessment*. Oxford: Heinemann.

Harrison, C. (2005). Teachers developing assessment for learning: Mapping teacher change. *Teacher Development*, *9*(2), 255-264. https://doi.org/10.1080/13664530500200264

Hattie, J. (2009). The black box of tertiary assessment: An impending revolution. *Tertiary Assessment and Higher Education Student Outcomes: Policy, Practice and Research*, 259–275.

Hattie, J. (2012). Know thy impact. *Educational Leadership*, *70*(1), 10–23.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

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

Hlas, A. C. (2016). Secondary teachers' language usage: Beliefs and practices. *Hispania, 99* (2), 305–319.

Hodgen, J., & Marshall, B. (2005). Assessment for learning in English and mathematics: A comparison. *The Curriculum Journal, 16*(2), 153-176. https://doi.org/10.1080/09585170500135954

Holley, L. C., & Steiner, S. (2005). Safe space: Student perspectives on classroom environment. *Journal of Social Work Education*, *41*(1), 49-64.

Holloway, I., & Galvin, K. (2016). *Qualitative research in nursing and healthcare*. London: John Wiley & Sons.

Hornero, A. M., Mur-Dueñas, P., & Plo, R. (2013). Oral skills in the spotlight: EFL in secondary education in a Spanish local context. *Synergy*, *9*(2), 111–124.

Hortigüela Alcalá, D., Pérez Pueyo, Á., & Moreno Doña, A. (2016). ¿Cómo enseñamos a los futuros docentes?: Análisis documental y contraste de percepciones entre alumnos y profesores. *Estudios Pedagógicos (Valdivia)*, *42*(3), 207-221. https://doi.org/10.4067/S0718-07052016000400011

House, J. (2007). What is an 'Intercultural Speaker'?. In Soler, E. A., & Jordá, M. P. S. (Eds.), *Intercultural language use and language learning*, (pp. 7-21). Dordrecht, Netherlands: Springer.

Hsu, T. C. (2019). Using a concept mapping strategy to improve the motivation of EFL students in google hangouts peer-tutoring sessions with native speakers. *Interactive Learning Environments*, *27*(2), 272–285. https://doi.org/10.1080/10494820.2018.1463268

Huebert, J. P., & Hauser, R. M. (1999). *High-takes testing for tracking, promotion, and graduation*. Washington, DC*:* National Academy Press.

Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.

Hughes, G. (2014). *Ipsative assessment: Motivation through marking progress*. Hampshire, UK: Palgrave Macmillan.

International Baccalaureate organization (IBO). (2011). *Historia del IB* [PDF file].https://ibo.org/globalassets/digital-toolkit/presentations/1711-presentation-history-of-the-ib-es.pdf

International Baccalaureate Organization (IBO). (2014). *Programme Standard and practices* [PDF file]. https://www.ibo.org/globalassets/publications/become-an-ib-school/programme-standards-and-practices-en.pdf

International Baccalaureate Organization (IBO). (2018a). *Language B guide* [PDF file]. http://www.cosmopolitanschool.de/wp-content/uploads/2018/10/dp-language-B-guide.pdf

International Baccalaureate Organization (IBO). (2018b). *El IB por país. España*. https://www.ibo.org/es/about-the-ib/the-ib-by-country/s/espana/

International Baccalaureate Organization (IBO). (2019). *What is an IB Education*?[PDF file] https://www.ibo.org/globalassets/what-is-an-ib-education-en.pdf

Ion, G., Sánchez Martí, A., & Agud Morell, I. (2019). Giving or receiving feedback: Which is more beneficial to students' learning? *Assessment & Evaluation in Higher Education*, *44*(1), 124–138. https://doi.org/10.1080/02602938.2018.1484881

Irons, A. (2007). *Enhancing learning through formative assessment and feedback*. London: Routledge.

James, M., & Gipps, C. (1998). Broadening the basis of assessment to prevent the narrowing of learning. *The Curriculum Journal, 9*(3), 285-297. https://doi.org/10.1080/0958517970090303

Jhangiani, R. S. (2016). The impact of participating in a peer assessment activity on subsequent academic performance. *Teaching of Psychology, 43*(3), 180-186. https://doi.org/10.1177/0098628316649312

Jiménez Raya, M. (2007). Developing professional autonomy: A balance between license and responsibility. *Independence, Newsletter of the IATEFL Learner Autonomy Special Interest Group*, *40*, 32–33.

Jiménez Raya, M. (2008). Learner autonomy as an educational goal in modern languages education. In M. Jimenez Raya & T. Lamb (Eds), *Pedagogy for autonomy in language education: Theory, practice and teacher education*, 3-15. Dublin: Authentik.

Johnston, T. C. (2015). Lessons from MOOCS: Video lectures and peer assessment. *Academy of Educational Leadership Journal*, *19*(2), 91-98.

Jones, J. (2014). *Developments in formative assessment: A retrograde step for teaching and learning?* London: Routledge.

Jones, J. (2017) Teachers as emergent critical researchers of practice. In Christie, C. and Conlon, C (Eds.), *Success stories from secondary foreign language classroom* (pp. 2-18), London: Trentham Book with UCL/ IOE Press.

Jones, J., & Coffey, S. (2006). *Issues in teaching primary MFL 5-11*. London: David Fulton Publishers.

Jones, J., & Wiliam, D. (2008). *Modern foreign languages inside the black box: Assessment for learning in the modern foreign languages classroom*. London: GL Assessment.

Jones, S. M. (2015). Making me feel comfortable: Developing trust in the nurse for Mexican Americans. *Western Journal of Nursing Research*, *37*(11), 1423-1440.

Joo, S. H. (2016). Self and peer assessment of speaking. *Studies in Applied Linguistics and TESOL*, *16*(2), 68-83.

Junta de Andalucía (2005). *Plan de Fomento del Plurilingüismo*. Seville: Junta de Andalucía

Junta de Andalucía (2016). *Plan Estratégico de Desarrollo de las Lenguas en Andalucía. Horizonte 2020*. Sevilla: Junta de Andalucía. https://bit.ly/2OXiYK5

Junta de Andalucía (2019). *La Educación en Andalucía. Inicio de Curso 2018-2019. Datos y Cifras Curso 2018/2019*. Sevilla: Junta de Andalucía https://bit.ly/2ReABHn

Kafele, B. K. (2016). *The teacher 50: Critical questions for inspiring classroom excellence*. ASCD.

Kappe, F. (2008). How reliable is peer assessment? Two empirical studies on assessments by students. *Active Learning in Higher Education 13*(2), 129-142.

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

Kim, M. (2009). The impact of an elaborated assessee's role in peer assessment. *Assessment & Evaluation in Higher Education, 34*(1), 105-114. https://doi.org/10.1080/02602930801955960

Kirk, J., and Miller, M.L. (1986). *Reliability and validity in qualitative research*. Sage. Newbury Park, CA.

Kitano, K. (2001). Anxiety in the college Japanese language classroom. *The Modern Language Journal*, *85*(4), 549–566. https://doi.org/10.1111/0026-7902.00125

Kizlik, B. (2012). *Measurement, assessment, and evaluation in education*. Florida: Boca Ratón: Robert Kizlik and Associates.http://www.adprima.com/measurement.htm

Klenowski, V. (2009). Assessment for learning revisited: An Asia-pacific perspective. *assessment in education: Principles, Policy & Practice*, *16*(3), 263-268. https://doi.org/10.1080/09695940903319646

Knoblauch, H. (2013). Qualitative methods at the crossroads: Recent developments in interpretive social research. *Forum*: *Qualitative Social Research*, 14(3). https://www.qualitative-research.net/index.php/fqs/article/view/2063/3646

Ko, E. (2015). *Design, development, evaluation of mobile application WikiTalki to promote English speaking skills in forma high school context* [Paper, University of Seoul]. Ewa Research Repository. https://dspace.ewha.ac.kr/handle/2015.oak/212788

Kuckartz, U., & Rädiker, S. (2019). *Analyzing qualitative data with MAXQDA: Text, audio, and video*. London: Springer.

Kuhn, D. (2015). Thinking together and alone. *Educational Researcher*, *44*(1), 46-53. https://doi.org/10.3102/0013189X15569530

Kukulska-Hulme, A., Lee, H., & Norris, L. (2017). Mobile learning revolution: Implications for language pedagogy. In Shapelle, C. A. & Sauro, S. (Eds.), *The hand- book of technology and second language teaching and learning* (pp. 217–233). Oxford: Wiley-Blackwell.

Kurt, G., & Atay, D. (2007). The effects of peer feedback on the writing anxiety of prospective Turkish teachers of English. *Journal of Theory and Practice in Education*, *3*(1), 12–13.

Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. London: Sage Publications.

Kvale, S., & Brinkmann, S. (2009). *InterViews: learning the craft of qualitative research interviewing. London*: Sage Publications.

Lafuente, A. L., Vázquez, M. M., Martínez, I. P., & Pizarro, M. A. (2020). Up for debate: English tests in the Spanish university entrance exams. *Advances in English and American Studies: Current Developments*, 209–214.

Lam, R. (2010). A peer review training workshop: Coaching students to give and evaluate peer feedback. *TESL Canada Journal*, *27*(2), 114-127. https://doi.org/10.18806/tesl.v27i2.1052

Lamb, M. (2017). The motivational dimension of language teaching. *Language Teaching, 50*(3), 301–346. https://doi.org/10.1017/S0261444817000088

Lamb, T., & Reinders, H. (2005). Learner independence in language teaching: A concept of change. *An International Perspective on Language Policies, Practices and Proficiencies*, *4*(1), 225-239.

Langan , A. M., Wheater, C. P., Shaw, E. M., Haines, B. J., Cullen, W. R., Boyle, J. C., Penney, D., Oldekop, J. A., Ashcroft, C., Lockey, L., & Preziosi, R. F. (2005). Peer assessment of oral presentations: Effects of student gender, university affiliation and participation in the development of assessment criteria. *Assessment & Evaluation in Higher Education*, *30*(1), 21–34. https://doi.org/10.1080/0260293042003243878

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Leathy, S., & Wiliam, D. (2015). *Embedding formative assessment: Practical techniques for K-12 classrooms*. USA: Hawker Brownlow Education.

Lechermeier, J., & Fassnacht, M. (2018). How do performance feedback characteristics influence recipients' reactions? A state-of-the-art review on feedback source, timing, and valence effects. *Management Review Quarterly*, *68*(20), 145-193. https://doi.org/10.1007/s11301-018-0136-8

Lee, C., & Wiliam, D. (2005). Studying changes in the practice of two teachers developing assessment for learning. *Teacher Development*, *9*(2), 265-283. https://doi.org/10.1080/13664530500200244

Leenknecht, M., Wijnia, L., Köhlen, M., Fryer, L., Rikers, R., & Loyens, S. (2021). Formative assessment as practice: The role of students' motivation. *Assessment & Evaluation in Higher Education*, *46*(2), 236-255.

Leong, L. M., & Ahmadi, S. M. (2017). An analysis of factors influencing learners' English speaking skill. *International Journal of Research in English Education*, *2*(1), 34–41. https://doi.org/10.18869/acadpub.ijree.2.1.34

Ley Orgánica 10/ 2002, de 23 diciembre, de calidad de la educación (2002). *Boletín Oficial del Estado, 307,* de 24 de diciembre de 2002, 45188 a 45220. https://www.boe.es/eli/es/lo/2002/12/23/10

Ley orgánica 2/2006, de 3 mayo, de educación (2006). *Boletín Oficial del Estado, 106,* de 4 de mayo de 2006. https://www.boe.es/eli/es/lo/2006/05/03/2/con

Ley Orgánica 3/2020, de 29 diciembre, de modificación de la ley orgánica 2/2006, de 3 mayo, de Educación. *Boletín Oficial del Estado, 340,* de 30 diciembre 2020, 122868 a 122953. https://www.boe.es/eli/es/lo/2020/12/29/3

Ley Orgánica 1/1990, de 3 octubre, de ordenación general del sistema educativo (1990). *Boletín Oficial del Estado, 238,* de 4 de octubre, 28927 a 28942. https://www.boe.es/buscar/doc.php?id=BOE-A-1990-24172

Ley Orgánica 8/2013, de 9 diciembre, para la mejora de la calidad educativa (2013). *Boletín Oficial del Estado*, *295*, 10 de diciembre. https://www.boe.es/buscar/act.php?id=BOE-A-2013-12886

Ley Orgánica 3/2020, de 29 diciembre, de modificación de la ley orgánica 2/2006 (2020). *Boletín Oficial del Estado*, de 3 mayo.

Li, Y., & Dervin, F. (2018). *Continuing professional development of teachers in Finland*. Cham, Switzerland: Palgrave Macmillan.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. London: Sage Publications.

Linn, R. L., & Grolund, N. E. (2000). *Measurement and Assessment in Teaching* (8th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

Liu, S. H. J., & Lan, Y. J. (2016). Social constructivist approach to Web-Based EFL learning: Collaboration, motivation, and perception on the use of Google Docs. *Educational Technology & Society, 19*(1), 171–186.

Liu, X., & Li, L. (2014). Assessment training effects on student assessment skills and task performance in a technology-facilitated peer assessment. *Assessment & Evaluation in Higher Education*, *39*(3), 275–292. https://doi.org/10.1080/02602938.2013.823540

Liu, X., Li, L., & Zhang, Z. (2018). Small group discussion as a key component in online assessment training for enhanced student learning in web-based peer assessment. *Assessment & Evaluation, 43*(2), 207-222. https://doi.org/10.1080/02602938.2017.1324018

Loo, S. Y. (2016). *Training of teachers of occupation-related programmes*. New York: Nova Science Publishers.

López Bertomeo, E., & González Olivares, Á. L. (2018). Las evaluaciones externas y el rendimiento académico de los alumnos como instrumento de mejora educativa. *Tendencias Pedagógicas*, *31*, 309-328 https://doi.org/10.15366/tp2018.31.017

López Pastor, V. M., & Pérez Pueyo, A. (2017). *Evaluación formativa y compartida en educación: Experiencias de éxito en todas las etapas educativas* [monografía, Universidad de León]. BULERIA. http://buleria.unileon.es/handle/10612/5999

Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

Luque Agulló, G., Bueno-Alastuey, M., & Ramos Álvarez, M. M. (2016). Uso selectivo de estrategias según la destreza oral: un análisis de libros de texto de enseñanza del inglés usando la modelización categórica: Selective strategy use depending on oral skill: An analysis of EFL course books using categorical modelization. *Revista de Educación*, 374: 10-72.

Lyrigkou, C. (2021). *The role of informal second language learning in the spoken use of discourse markers by Greek adolescent learners of English* (Doctoral dissertation, Open University (United Kingdom)).

Madrid Fernández, D., Ortega-Martín, J. L., & Hughes, S. P. (2019). CLIL and language education in Spain. In K. Tsuchiya, & M. Pérez Murillo (Eds.), *Content and Language Integrated Learning in Spanish and Japanese Contexts* (pp. 11–35). Palgrave Macmillan., Cham. https://doi.org/10.1007/978-3-030-27443-6_2.

Magee, B. (2000). *The Story of Philosophy*. London: Dorling Kindersley.

Malan, M., & Stegmann, N. (2018). Accounting students' experiences of peer assessment: A tool to develop lifelong learning. *South African Journal of Accounting Research*, *32*(2–3), 205–224. https://doi.org/10.1080/10291954.2018.1487503

Mann, C.J. (2003). Observational research methods. Research design II: Cohort, cross sectional and case-control studies. *Emergency Medicine Journal, 20*(1), 54-60.

https://doi.org/10.1136/emj.20.1.54

Marks, L., Meek, S., Huser, C., & Blakemore, L. (2016). *Playing the numbers game: Students as sssessors in a MOOC context.* 9th Annual University of Glasgow Learning and Teaching Conference. Glasgow.

Martínez Serrano, B. (2017). El papel de la formación del profesorado en el Plan Estratégico de Desarrollo de las Lenguas en Andalucía. Horizonte 2020. *Revista Digital de Educación y Formación del profesorado.* CEP de Córdoba, [Archivo PDF]. http://revistaeco.cepcordoba.es/wp-content/uploads/2017/04/Serrano.pdf

Mason, J. (2002). *Researching your own practice: The discipline of noticing*. London: Routledge-Falmer.

McClure, C., & McAndrews, L. E. (2016). Going native to reach the digital natives: New technologies for the classroom, *International Textile and Apparel Association Annual Conference Proceedings 73*(1). Iowa State University Digital Press. https://www.iastatedigitalpress.com/itaa/article/id/3003/

McConlogue, T. (2015). Making judgements: Investigating the process of composing and receiving peer feedback. *Studies in Higher education*, *40*(9), 1495-1506. https://doi.org/10.1080/03075079.2013.868878

MacIntyre, P.D., & Gregersen, T. (2012). Emotions that facilitate language learning: The positive broadening power of the imagination. *Studies in Second Language Learning and Teaching*, *2*(2), 193–213. https://doi.org/10.14746/ssllt.2012.2.2.4

McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy and Practice*, *10*(2), 209-220. https://doi.org/10.1080/0969594032000121289

McLain, T. R. (2018). Integration of the video response app flipgrid in the business writing classroom. *International Journal of Educational Technology and Learning*, *4*(2), 68-75.

McMahon, T. (2010). Combining peer-assessment with negotiated learning activities on a day-release undergraduate-level certificate course (ECTS level 3). *Assessment & Evaluation in Higher Education*, *35*(2), 223–239. https://doi.org/10.1080/02602930902795919

Miller, R. L., King, J. A., Mark, M. M., & Caracelli, V. (2016). The oral history of evaluation: The professional development of Robert Stake. *American Journal of Evaluation*, *37*(2), 287-294.

Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centred framework for feedback literacy. *Assessment & Evaluation in Higher Education*, *45*(4), 527–540. https://doi.org/10.1080/02602938.2019.1667955

Moya Guijarro, A.J and Ruiz Cordero, M.B (2019). Un estudio sobre la diferencia de nivel de competencia lingüística entre alumnos pertenecientes a centros bilingües y o bilingües en la comunidad autónoma de Castilla-La Mancha (España*). Estudios Filológicos*, (62) 269-288.

Mur-Dueñas, M. P., Plo, R. & Hornero, A. (2013). Spanish secondary school students' oral competence in EFL: self assessment, teacher self assessment and assessment tasks. *Miscelánea: A Journal of English and American Studies*, (47),103–124 https://dialnet.unirioja.es/servlet/articulo?codigo=5826342

Murillo-Zamorano, L. R. & Montanero-Fernánez, M. (2014). *Coevaluación Iterativa con Rubrica de Exposiciones en la Educación Superior.* http://hdl.handle.net/11268/3636

Murillo-Zamorano, L. R. & Montanero- Fernández, M. (2018). Oral presentations in higher education: A comparison of the impact of peer and teacher feedback. *Assessment & Evaluation in Higher Education, 43*(1), 138-150. https://doi.org/10.1080/02602938.2017.1303032

Nguyen, T.T.H., & Walker, M. (2016). Sustainable assessment for lifelong learning. *Assessment & Evaluation in Higher Education*, *41*(1), 97-111. https://doi.org/10.1080/02602938.2014.985632

Nicol, D. (2009). Assessment for learner self-regulation: Enhancing achievement in the first year using learning technologies. *Assessment & Evaluation in Higher Education*, *34*(3), 335–352. https://doi.org/10.1080/02602930802255139

Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, *35*(5), 501–517. https://doi.org/10.1080/02602931003786559

Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218. https://doi.org/10.1080/03075070600572090

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education*, *39*(1), 102–122. https://doi.org/10.1080/02602938.2013.795518

Nilson, L. B. (2003). Improving student peer feedback. *College Teaching*, *51*(1), 34–38. https://doi.org/10.1080/87567550309596408

Ofsted. (2019). *Education Inspection Framework.* GOV.UK https://www.gov.uk/government/collections/education-inspection-framework

Orden de 1/2021, de 15 enero, por la que se desarrolla el currículo correspondiente a la etapa de Bachillerato en la Comunidad Autónoma de Andalucía (2021). Boletín Oficial de la Junta de Andalucía,7, de 18 de enero de 2021, 224-454.

Ortega, D. P., & Minchala, O. E. (2017). Assessing students in an authentic and ongoing manner in the English classroom. *Theory and Practice in Language Studies*, *7*(3), 159. https://doi.org/10.17507/tpls.0703.01

Ozogul, G., & Sullivan, H. (2009). Student performance and attitudes under formative evaluation by teacher, self and peer evaluators. *Educational Technology Research and Development*, *57*(3), 393–410.

Palacios-Hidalgo, F. J., Huertas-Abril, C. A., & Gómez-Parra, M. E. (2021). Language teaching and bilingual education policies in Scotland and Andalusia: A language-use-based comparative study. *Research in Comparative and International Education*, *16*(2), 140-159.

Palomeque, N.G. (2015). El desarrollo de la competencia comunicativa: una mirada hacia la formación de maestros. *Revista de la Facultad de Educación*, 22, 65-70.

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 247-266). New York: Routledge.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*(422), 1-28. https://doi.org/10.3389/fpsyg.2017.00422

Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: A roadmap of where we are, how we got here, and where we are going. *The Australian Educational Researcher, 45*(1), 13-31. https://doi.org/10.1007/s13384-018-0258-y

Panadero, E., & Lipnevich, A. A. (2021). A review of feedback models and typologies: towards an integrative model of feedback elements. *Educational Research Review, 35, 100416.* https://www.sciencedirect.com/science/article/pii/S1747938X21000397?via%3Dihub

Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2019). Using formative assessment to influence self-and co-regulated learning: the role of evaluative judgement. *European Journal of Psychology of Education*, *34*(3), 535-557.

Panadero, E., & Brown, G. (2017). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology of Education*, *32*(1), 133–156. https://doi.org/10.1007/s10212-015-0282-5

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74–98. https://doi.org/10.1016/j.edurev.2017.08.004

Panadero, E., Jonsson, A., & Strijbos, J. W. (2016). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. In *Assessment for learning: Meeting the challenge of implementation* (pp. 311-326). Springer, Cham.

Patchan, M. M., Schunn, C. D., & Clark, R. J. (2011). Writing in natural sciences: Understanding the effects of different types of reviewers on the writing. *Journal of Writing Research*, *2*(3), 365–393. https://doi.org/10.17239/jowr-2011.02.03.4

Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing*, *19*(2), 109–131. https://doi.org/10.1191/0265532202lt224oa

Pavón, V. & Ellison, M. (2013). "Examining teacher roles and competences in content and language integrated learning (CLIL)", *Linguarum Arena*, 4, 65-78.

Pavón, V., Lancaster, N., & Callejas, C. B. (2020). Keys issues in developing teachers' competences for CLIL in Andalusia: training, mobility and coordination. *The Language Learning Journal*, *48*(1), 81-98.

Penuel, W. R., & Shepard, L. A. (2016). Social models of learning and assessment. In Rupp A. A. and Leignton J.P, (Eds), *The handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 146-173).

Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning processes. Towards a wider conceptual field. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 85–102. https://doi.org/10.1080/0969595980050105

Petersen, J. B., Townsend, S. D., & Onaka, N. (2020). Utilizing Flipgrid Application on Student Smartphones in a Small-Scale ESL Study. *English Language Teaching*, *13*(5), 164-176.

Pintrich, P. R., & Zusho, A. (2002). Student motivation and self-regulated learning in the college classroom, in J.C. Smart (Ed), *Higher education: Handbook of theory and Research*, *17*, pp. 55–128). New York: Agathon Press.

Plo, R., Hornero, A., & Dueñas, P. (2014). Implementing the teaching/learning of oral skills in secondary education in Aragón: Gauging teachers' attitudes, beliefs and expectations. *International Journal of English Studies*, *14*(1), 55-77. https://doi.org/10.6018/ijes/14/1/175041

Plo, R., & Pérez-Llantada, C. (2010). A longitudinal attitude survey on English oral skills: Classroom, curriculum, learning and pedagogical implications. *Studium*, *16*(1), 309–326.

Popham, W. J. (2003). *Test better, teach better: The instructional role of assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Popham, W. J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Potter, J., & Shaw, C. (2018). The virtues of naturalistic data. In U. Flick (Ed.), *The SAGE handbook of qualitative data collection* (pp. 182-199). London: SAGE.

Powney, J., & Watts, M. (1987). *Interviewing in Educational Research*. London: Routledge.

Archive of European Integration (2002, March 16). Barcelona European Council 15-16 March 2002. Presidency conclusions. Barcelona. http://aei.pitt.edu/id/eprint/43345

Pressey, S.L. (1962). Basic unresolved teaching-machine problems. *Theory into Practice, 1*(1), 30-37. https://doi.org/10.1080/00405846209541773

Price, M., & O'Donovan, B. (2006). Improving performance through enhancing student understanding of criteria and feedback. In C. Bryan and K. Clegg (Eds.), *Innovative Assessment in Higher Education* (pp.100–109). London: Routldege.

Purpura, J. E. (2016). Second and foreign language assessment: Second and foreign language assessment. *The Modern Language Journal*, *100*, 190-208. https://doi.org/10.1111/modl.12308

Radford, B. W. (2014). *The effect of formative assessments on language performance*. Brigham Young University, Provo, Utah.

Ramsden, P. (1992). *Learning to teach in higher education*. London, UK: Routledge.

Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: Disentangling the construct of formative assessment. *Language Testing*, 17(2), 215-243. https://doi.org/10.1191/026553200673636547

Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato. *Boletín oficial del Estado*, 3, de 2 enero de 2015.https://www.boe.es/boe/dias/2015/01/03/pdfs/BOE-A-2015-37.pdf

Reinholz, D. (2016). The assessment cycle: a model for learning through peer assessment. *Assessment & Evaluation in Higher Education, 41*(2), 301-315.

Renukadevi, D. (2014). The role of listening in language acquisition; the challenges and strategies in teaching listening. *International Journal of Education and Information Studies*, *4*(1), 59–63.

Resnik, J. (2019). Struggling for recognition: Access to higher education through the international baccalaureate. *Critical Studies in Education, 60*(3), 340-357. https://doi.org/10.1080/17508487.2016.1263222

Resolución 2/ 2022, de 18 de febrero, conjunta de la Dirección General de Ordenación y Evaluación Educativa y de la Dirección General de Planificación de Centros, por la que se dictan instrucciones para regular la compatibilidad de las enseñanzas de Bachillerato y el Programa del Diploma del Bachillerato Internacional en centros docentes de titularidad pública en Andalucía para el curso 2022/2023, Junta de Andalucía, 1-22, [Archivo PDF]. https://www.adideandalucia.es/normas/resoluc/Resolucion18febrero2022BachilleratoInternacional.pdf

Rieckmann, M. (2017). *Education for Sustainable Development Goals: Learning objectives*. Unesco Publishing, Barcelona.

Robinson, V. M., & Timperley, H. S. (2007). The leadership of the improvement teaching and learning: Lessons from initiatives with positive outcomes for students. *Australian Journal of Education, 51*(3), 247-262.

Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers*. London: Blackwell Publishers.

Roldán Tapia, R. (2012). The shaping of Spanish CLIL, *Encuentro*, 21, 71-79.

Rotsaert, T., Panadero, E., Schellens, T., & Raes, A. (2018). "Now you know what you're doing right and wrong!" Peer feedback quality in synchronous peer assessment in secondary education. *European Journal of Psychology of Education, 33*(2), 255-275. https://doi.org/10.1007/s10212-017-0329-x

Rowntree, D. (1987). *Assessing students: How shall we know them?* London: Routledge.

Rubio Alcalá, F. D., & Martínez Lirola, M. (2008). La comunicación oral del inglés en España: influencia de los aspectos educativos y mediáticos en el éxito del aprendizaje. *Revista de Enseñanza Universitaria*, 32, 51-63.

Rubio Alcalá, F. D., & Martínez Lirola, M. M. (2010). English as a foreign language in the EU: Preliminary analysis of the difference in proficiency levels among the member states. *European Journal of Language Policy*, *2*(1), 23.

Rubio Alcalá, F. D., & Tamayo Rodríguez, L. (2012). Estudio sobre prácticas docentes en la evaluación de lengua inglesa en la ESO. *Profesorado: Revista de Curriculum y Formación del Profesorado*, *16*(1), 295–316.

Rudduck, J., & Flutter, J. (2004). *How to improve your school: Giving pupils a voice*. London: Continuum.

Ruegg, R. (2015). The relative effects of peer and teacher feedback on improvement in EFL students' writing ability. *Linguistics and Education, 29*, 73-82.

Ruiz Lázaro, J., González Barberá, C., & Gaviria Soto, J. L. (2021). Las pruebas de inglés para acceder a la Universidad. Una comparación entre Comunidades Autónomas. *Educación XX1*, *24*(1), 233-270, http://doi.org/10.5944/educXX1.26746

Saddler, B., & Andrade, H. (2004). The writing rubric. *Educational Leadership*, *62*(2), 48–52.

Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119-144. https://doi.org/10.1007/BF00117714

Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing*, *25*(4), 553-581. https://doi.org/10.1177/0265532208094276

Salaberri, S. (1995). *Classroom language*. Oxford: Macmillan Heinemann Publishers.

Saldaña, J. (2011). *Fundamentals of qualitative research*. New York: Oxford University Press.

Sanmarti, N. (2010). *Evaluar para aprender*. Barcelona: GRAÓ.

Sanmarti, N. (2018). *La finalidad principal de la evaluación es la regulación tanto de la enseñanza como del aprendizaje. 10 ideas clave. Evaluar para aprender*. Barcelona: GRAÓ

Sanmarti, N. (2020). *Evaluar y aprender: Un único proceso.* Barcelona: Octaedro.

Santos Estrada, P. (2016). *Expresión oral y lenguaje figurado en la clase de inglés, 1º de bachillerato: una propuesta de mejora*. Repositorio institucional de la universidad de Oviedo. https://digibuo.uniovi.es/dspace/handle/10651/38418

Sargeant, J. M., Mann, K. V., van der Vleuten, C. P., & Metsemakers, J. F. (2009). Reflection: A link between receiving and using assessment feedback. *Advances in Health Sciences Education, 14*(3), 399–410. https://doi.org/10.1007/s10459-008-9124-4

Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning & Technology*, 13, 96–120.

Scriven, M. (1967). The methodology of evaluation. In R. W Tyler (Ed.), *Perspectives on Curriculum Evaluation* (pp. 39-83). Chicago: Rand McNally.

Searby, M., & Ewers, T. (1997). An evaluation of the use of peer assessment in higher education: A case study in the school of music, Kingston University. *Assessment & Evaluation in Higher Education, 22*(4), 371-383.

Seedhouse, P & Jenks, C. J. (2015). International perspectives on ELT classroom interaction*:* an introduction. In C. J. Jenks y P. Seedhouse (Eds.), *International Perspectives on ELT Classroom Interaction* (pp.1-9). London: Palgrave Macmillan. https://doi.org/10.1057/9781137340733_1

Selcuk, H., & Jones, J. (2022). Turkish EFL Learner Perceptions of Using a Social Network Environment for Collaborative Writing: Creating a Trustful Affinity Space. *International Journal of Smart Education and Urban Society*, *13*(1), 1-14.

Selcuk, H., Jones, J., & Vonkova, H. (2019). The emergence and influence of group leaders in web-based collaborative writing: Self-reported accounts of EFL learners. *Computer Assisted Language Learning, 34*(8), 1040-1060. https://doi.org/10.1080/09588221.2019.1650781

Sendziuk, P. (2010). Sink or swim? Improving student learning through feedback and self-assessment. *International Journal of Teaching and Learning in Higher Education*, *22*(3), 320–330.

Serrano, R., & Muñoz, C. (2007). Same hours, different time distribution: Any difference in EFL? *System*, *35*(3), 305-321.

Shen, B., Bai, & Xue, W. (2020). The effects of peer assessment on learner autonomy: An empirical study in a Chinese college English writing class. *Studies in Educational Evaluation,* 64. https://doi.org/10.1016/j.stueduc.2019.100821

Shih, R. C. (2011). Can Web 2.0 technology assist college students in learning English writing? Integrating Facebook and peer assessment with blended learning. *Australasian Journal of Educational Technology*, 27(5), 829-845.

Shin, J. Y., Dixon, L. Q., & Choi, Y. (2020). An updated review on use of L1 in foreign language classrooms. *Journal of Multilingual and Multicultural Development*, *41*(5), 406-419.

Short, D., Becker, H., Cloud, N., & Hellman, A. B. (2018). *The 6 principles for exemplary teaching of English learners: Grades K-12*. TESOL Press.

Silverman, D. (2005). Instances or sequences? Improving the state of the art of qualitative research. *Forum Qualitative Socialforschung/Forum: Qualitative Social Research*, *6*(3). thttps://doi.org/10.17169/fqs-6.3.6

Silverman, D. (2013). *Doing qualitative research*: *A practical handbook*. London:Sage.

Siqin, T., van Aalst, J., & Chu, S. K. W. (2015). Fixed group and opportunistic collaboration in a CSCL environment. *International Journal of Computer-Supported Collaborative Learning*. https://doi.org/10.1007/s11412-014-9206-7

Soler, H. H. (1999). Is the English test in the Spanish university entrance examination as discriminating as it should be? *Estudios Ingleses de la Universidad Complutense*, *7*, 89–107.

Song, Y., Hu, Z., Gehringer, E. F., Morris, J., Kidd, J., & Ringleb, S. (2016). Toward better training in peer assessment: Does calibration help? *Computer-Supported Peer Review in Education*. Raleigh, NC, U.S. http://ceur-ws.org/Vol-1633/ws1-paper10.pdf

Spolsky, B., & Hult, F. M. (Eds.). (2010). *The handbook of educational linguistics*. Oxford: Blackwell.

Stewart, W. (2013, July 26). Is Pisa fundamentally flawed? *The Times Educational Suplement.* https://www.proquest.com/openview/b44c4949d5cee5372691c94cb6376499/1?pq-origsite=gscholar&cbl=366363

Stiggins, R. (2005). *Student-involved assessment for learning*. NJ, Prentice Hall.

Stiggins, R. (2002). Assessment crisis: The absence of assessment *for* learning. *Phi Delta Kappan*, *83*(10), 758–765. https://doi.org/10.1177/003172170208301010

Stobart, G. (2008). *Testing Times: The uses and abuses of assessment*. Abingdon, England: Routledge.

Stobart, G., & Gipps, C. (1997). *Assessment: A teacher's guide to the issues*. London: Hodder and Stoughton.

Stoll, L., & Fink, D. (1996). *Changing our schools: Linking school effectiveness and school improvement*. Buckingham: Open University Press.

Storch, N. (2013). *Collaborative writing in L2 classrooms: New perspectives on language and education*. Bristol: Multilingual Matters.

Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. London: Sage Publications.

Susinos-Rada, T., Calvo-Salvador, A., Rodríguez-Hoyos, C., & Saiz-Linares, A. (2019). ICT for inclusion. A student voice research project in Spain. *Magis. Revista Internacional de Investigación en Educación, 11*(23), 39–54. https://doi.org/10.11144/Javeriana.m11-23.iisv

Swaffield, S. (2011). Getting to the heart of authentic Assessment for Learning. *Assessment in Education: Principles, Policy and Practice,18*(4), 433-449. https://doi.org/10.1080/0969594X.2011.582838

Sykes, J. M. (2009). Learner requests in Spanish: Examining the potential of multiuser virtual environments for L2 pragmatics acquisition. In Lomicka, L. & Lord, G. (Eds.), *The next generation: Social networking and online collaboration in foreign language learning* (pp. 199–234). San Marcos, TX: CALICO.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467-481.

Tamayo-Beritán, AL. y Fernández- Terry, J.A. (2017). La importancia de diseñar instrumentos eficaces para evaluar la habilidad oral en el Idioma Inglés. *Educación y Sociedad, 15*(1), 121-135.

Tang, Y., & Hew, K. F. (2017). Is mobile instant messaging (MIM) useful in education? Examining its technological, pedagogical, and social affordances. *Educational Research Review*, 21, 85–104. https://doi.org/10.1016/j.edurev.2017.05.001

Tanner, H., & Jones, S. (2003). Self-Efficacy in mathematics and students' use of self-regulated learning strategies during assessment events. *International Group for the Psychology of Mathematics Education*, *4*, 275–282.

Taras, M. (2008). Summative and formative assessment: Perceptions and realities. *Active Learning in Higher Education, 9*(2), 172-192. https://doi.org/10.1177/1469787408091655

Tasker, T. Q., & Herrenkohl, L. R. (2016). Using peer feedback to improve students' scientific inquiry. *Journal of Science Teacher Education, 27*(1), 35-39. https://doi.org/10.1007/s10972-016-9454-7

Taylor, M. A. (2017). *Formative for whom? How formative assessment is used in the English classroom of an upper secondary school in Sweden* [Thesis, Umea University]. DiVa portal in Umea University.

Thoilliez, B., & Rappoport, S. (2018). El Programa del Diploma de Bachillerato Internacional institutos públicos españoles: Tres miradas sobre las voces de sus antiguos alumnos. *Journal of Supranational Policies of Education*, *7*, 83–104.

To, J., & Panadero, E. (2019). Peer assessment effects on the self-assessment process of first-year undergraduates. *Assessment & Evaluation in Higher Education*, *44*(6), 920–932. https://doi.org/10.1080/02602938.2018.1548559

Tognini, R., Philp, J., & Oliver, R. (2010). Rehearsing, conversing, working it out: Second language use in peer interaction. *Australian Review of Applied Linguistics*, *33*(3), 28-1.

Tokuno, K. A. (Ed.). (2008). *Graduate students in transition: Assisting students through the first year* [Monograph series, University of South Carolina].National resource center for the first-year experience and students in transition.

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249–276. https://doi.org/10.3102/00346543068003249

Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. S. R. Segers, F. J. R. C. Dochy, & E. C. Cascallar (Eds.), *Optimizing new modes of assessment: In search of qualities and standards* (pp. 55–87). Dordrecht, Netherlands: Kluwer Academic Publishers.

Topping, K.J. (2009). Peer assessment. *Theory Into Practice*, *48*(1), 20-27. https://doi.org/10.1080/00405840802577569

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and instruction*, *20*(4), 339-343.

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal*, *27*(5), 615–631. https://doi.org/10.1080/01411920120095780

Tsagari, D. (2011). Washback of a high-stakes English exam on teachers' perceptions and practices. *Selected Papers on Theoretical and Applied Linguistics*, *19*, 431–445.

Tseng, W.T., Chang, Y.J., & Cheng, H.F. (2015). Effects of l2 learning orientations and implementation intentions on self-regulation. *Psychological Reports*, *117*(1), 319–339. https://doi.org/10.2466/11.04.PR0.117c15z2

Tsiplakides, I., & Keramida, A. (2009). Helping students overcome foreign language speaking anxiety in the English classroom: Theoretical issues and practical recommendations. *International Education Studies*, *2*(4), 39-44. https://doi.org/10.5539/ies.v2n4p39

Tunstall, P., & Gipps, C. (1996). 'How does your teacher help you to make your work better?' Children's understanding of formative assessment. *The Curriculum Journal*, *7*(2), 185–203. https://doi.org/10.1080/0958517960070205

Ur, P. (1996). *A course in language teaching*. Cambridge: Cambridge University Press.

Van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning*, *4*(3), 259–287. https://doi.org/10.1007/s11412-009-9069-5

Van del Pol, J., Volman, M., and Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. *Educational Psychology Review*, 22(3), 271-296. https://doi.org/10.1007/s10648-010-9127-6

Van Duijvenvoorde, A. C. K., Zanolie, K., Rombouts, S. A. R. B., Raijmakers, M. E. J., & Crone, E. A. (2008). Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *Journal of Neuroscience, 28* (38), 9495-9503.https://doi.org/10.1523/JNEUROSCI.1485-08.2008

Van Lier, L. (1996). *Interaction in The Language Curriculum: Awareness, Autonomy, and Authenticity*. Harlow: Pearson Education.

Van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, *20*(4), 270–279. https://doi.org/10.1016/j.learninstruc.2009.08.004

Vercellotti, M. L., & McCormick, D. E. (2021). Constructing analytic rubrics for assessing open- ended tasks in the language classroom. *The Electronic Journal for English as a Second Language, 24*(4), 1-19.

Vonkova, H., Jones, J., Moore, A., Altinkalp, I., & Selcuk, H. (2021). A review of recent research in EFL motivation: Research trends, emerging methodologies, and diversity of researched populations. *System*, *103*. https://doi.org/10.1016/j.system.2021.102622

Walker, M. (2009). An investigation into written comments on assignments: Do students find them usable? *Assessment and Evaluation in Higher Education, 34*(1), 67-68. https://doi.org/10.1080/02602930801895752

Wanner, T. & Palmer, E. (2018). Formative self-and peer assessment for improved student learning: the crucial factors of design, teacher participation and feedback. *Assessment & Evaluation in Higher Education, 43*(7), 1032-1047. https://doi.org/10.1080/02602938.2018.1427698

Washington-Clark, S. E. (2019). *Digging deeper into formative assessment data: A case study of middle school language arts teachers' perspectives about formative assessment* [Doctoral Dissertation, Capela University]. ProQuest. https://www.proquest.com/openview/b00862fdee171b9864a1beee6977245d/1?pq-origsite=gscholar&cbl=18750&diss=y

Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education, 31*(3), 379-394. https://doi.org/10.1080/02602930500353061

Webb, M., & Jones, J. (2009). Exploring tensions in developing assessment for learning. As *Assessmen in Education: Principles, Poolicy and Practice, 16*(2), 165-184. https://doi.org/10.1080/09695940903075925

Webb, N.L. (1992). Assessment of students' knowledge of mathematics: Steps toward theory. *Handbook of Research on Mathematics Teaching and Learning*, 661-683.

Weeden, P., Winter, J., & Broadfoot, P. (2002). *Assessment*. London: Routledge/Falmer.

Wei, L. (2018). Translanguaging as a practical theory of language. *Applied Linguistics*, *39*(1), 9-30.

Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. *Handbook of Formative Assessment*, 18–40.

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

Wiliam, D. (2017). Learning and assessment: A long and winding road? *Assessment in Education: Principles, Policy & Practice, 24*(3), 309-316. https://doi.org/10.1080/0969594X.2017.1338520

Wiliam, D., & Leahy, S. (2016). *Embedding formative assessment*. London: Hawker Brownlow Education.

Willis, J. (2011). Affiliation, autonomy and assessment for learning. *Assessment in Education: Principles, Policy & Practice, 18*(4), 399-415.

Wilson-Thompson, B. (2005). *Factors influencing teachers' choice and use of tasks for formative assessment of Mathematics in grades 2-6* [Unpublished master's research report, University of the Witwatersrand]. Wits Institutional Repository on DSPACE [PDF file].https://core.ac.uk/download/pdf/39664106.pdf

Winet, D. (2016). Mobile instant messaging in the ESL writing class. *The Electronic Journal for English as a Second Language*, 20(3), 1-6.

Woodrow, L. (2017). Motivation in language learning. In Breeze R. , Sancho Guinda, C. (Eds.), *Essential Competencies for English- Medium University Teaching. Educational:* Vol. 27. Educational Linguistics (pp. 235-248) Springer, Cham. https://doi.org/10.1007/978-3-319-40956-6_16

Wu, J. G., & Miller, L. (2020). Improving English learners' speaking through mobile-assisted peer feedback. *RELC Journal, 51*(1), 168-178. https://doi.org/10.1177/0033688219895335

Xiao, Y., & Yang, M. (2019). Formative assessment and self-regulated learning: How formative assessment supports students' self-regulation in English language learning. *System*, *81*, 39–49. https://doi.org/10.1016/j.system.2019.01.004

Xu, Y., & Carless, D. (2017). 'Only true friends could be cruelly honest': Cognitive scaffolding and social-affective support in teacher feedback literacy. *Assessment & Evaluation in Higher Education*, *42*(7), 1082–1094. https://doi.org/10.1080/02602938.2016.1226759

Yan, Z., and Boud, D. (2022). Coceptualising assessment as learning. In Z. Yan & L. Yang (Eds.). *Assessment as learning: Maximising opportunities for student learning and achievement* (pp. 11-24). London: Routledge.

Yin S., Chen F. and Chang H. (2022) Assessment as learning: How does peer assessment function in students' learning? *Frontiers in Psychology*. 13:912568. doi: 10.3389/fpsyg.2022.912568

Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching*, *49*(4), 461–493. https://doi.org/10.1017/S0261444816000161

Zhao, M., Mu, B., & Lu, C. (2016). Teaching to the test: Approaches to teaching in senior secondary schools in the context of curriculum reform in china. *Creative Education*, *07*(01), 32–43. https://doi.org/10.4236/ce.2016.71004

Zhu, Q., & Carless, D. (2018). Dialogue within peer feedback processes: Clarification and negotiation of meaning. *Higher Education Research & Development*, *37*(4), 883-897.

Zimmerman, B. J., and Schunk, D. H. (2011). *Handbook of Self-Regulation of Learning and Performance.* New York : Routledge.

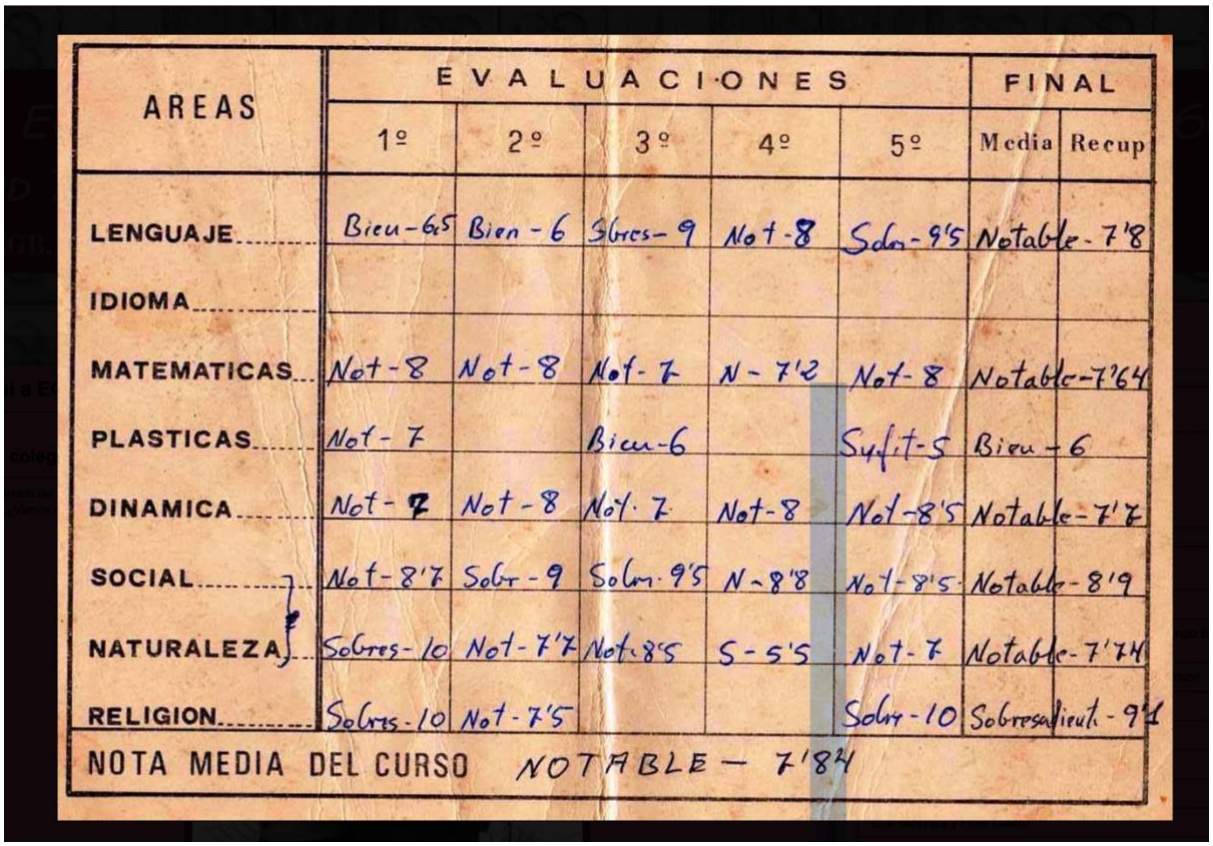
Zou, Y., Schunn, C. D., Wang, Y., & Zhang, F. (2018). Student attitudes that predict participation in peer assessment. *Assessment & Evaluation in Higher Education*, *43*(5), 800–811. https://doi.org/10.1080/02602938.2017.1409872

Zumeta, W. M. (2011). What does it mean to be accountable?: Dimensions and implications of higher education's public accountability. *The Review of Higher Education*, *35*(1), 131-148.

Zyoud, M. (2016). Theoretical perspective on how to develop speaking skill among university students. *Pune Research Scholar an International Multidisciplinary Journal*, *2*(1), 1–10.

CHAPTER 8 : APPENDICES

APPENDIX 1: REPORT IN THE 1970S IN SPAIN



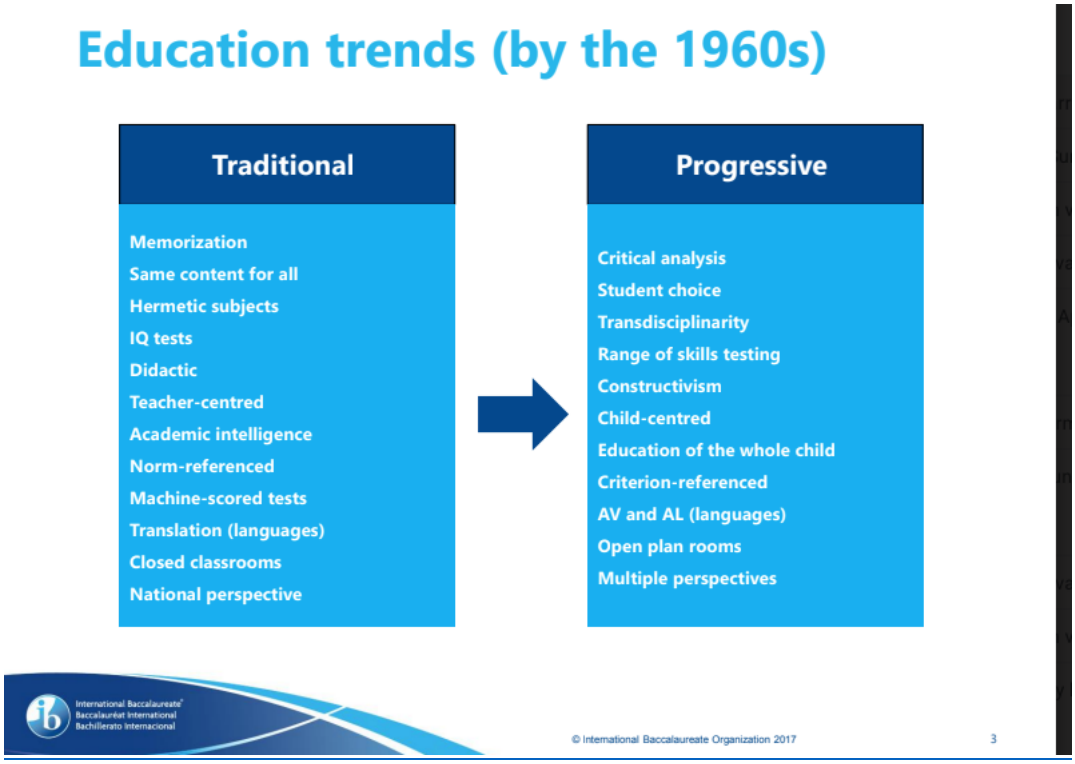
AREAS	EVALUACIONES					FINAL	
	1º	2º	3º	4º	5º	Media	Recup
LENGUAJE	Bien-6,5	Bien-6	Sbres-9	Not-8	Sobr-9'5	Notable-7'8	
IDIOMA							
MATEMATICAS	Not-8	Not-8	Not-7	N-7'2	Not-8	Notable-7'64	
PLASTICAS	Not-7		Bien-6		Suf.t-5	Bien-6	
DINAMICA	Not-7	Not-8	Not-7	Not-8	Not-8'5	Notable-7'7	
SOCIAL	Not-8'7	Sobr-9	Sobr-9'5	N-8'8	Not-8'5	Notable-8'9	
NATURALEZA	Sobres-10	Not-7'7	Not-8'5	S-5'5	Not-7	Notable-7'74	
RELIGION	Sobres-10	Not-7'5			Sobr-10	Sobresalient-9'1	
NOTA MEDIA DEL CURSO NOTABLE — 7'84							

http://yofuiaegb.blogspot.com/search/label/Material%20escolar%20el%20colegio%20en%20los%20años%2060%20y%2070

APPENDIX 2: AN INSIGHT TO THE INTERNATIONAL BACCALAUREATE (IB)

The International Baccalaureate® (IB) is a non-profit educational foundation, motivated by its mission, focused on the student. Our three programmes for students aged 3 to 19 help develop the intellectual, personal, emotional and social skills to live learn and work in a rapidly globalizing world. Founded in 1968, we currently work with 3,195 schools in 140 countries to develop and offer three challenging programmes to over 919,000 students aged 3 to 19 years (IB online), International Baccalaureate Organization https://www.ibo.org .

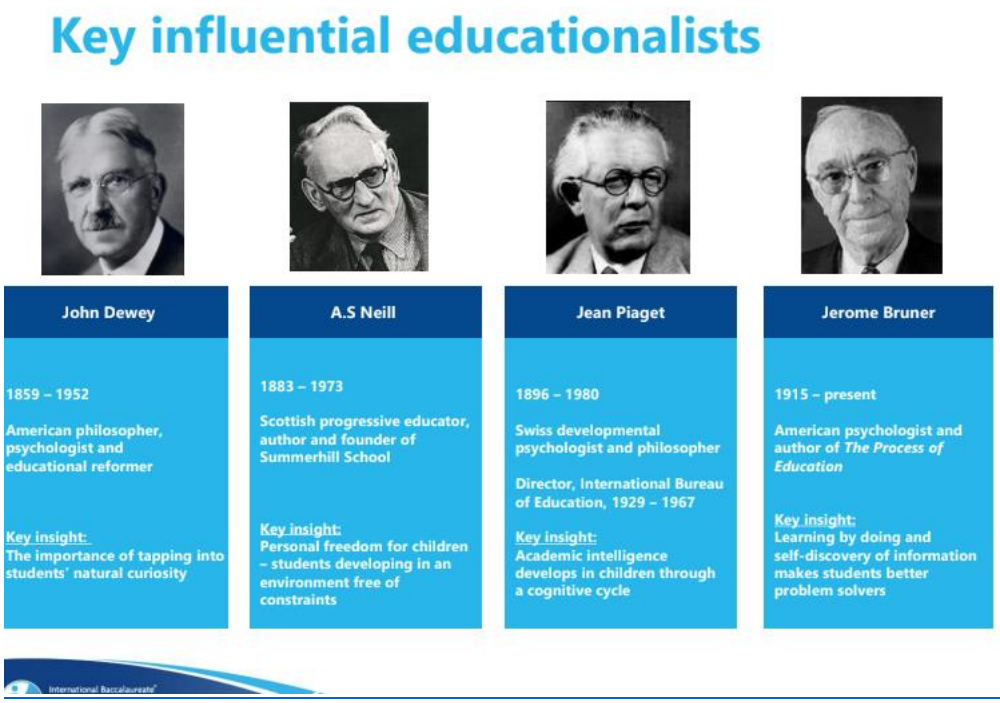
APPENDIX 3: EDUCATION TRENDS



International Baccalaureate (2017): The history of the IB

https://www.ibo.org/globalassets/digital-toolkit/presentations/1711-presentation-history-of-the-ib-en.pdf

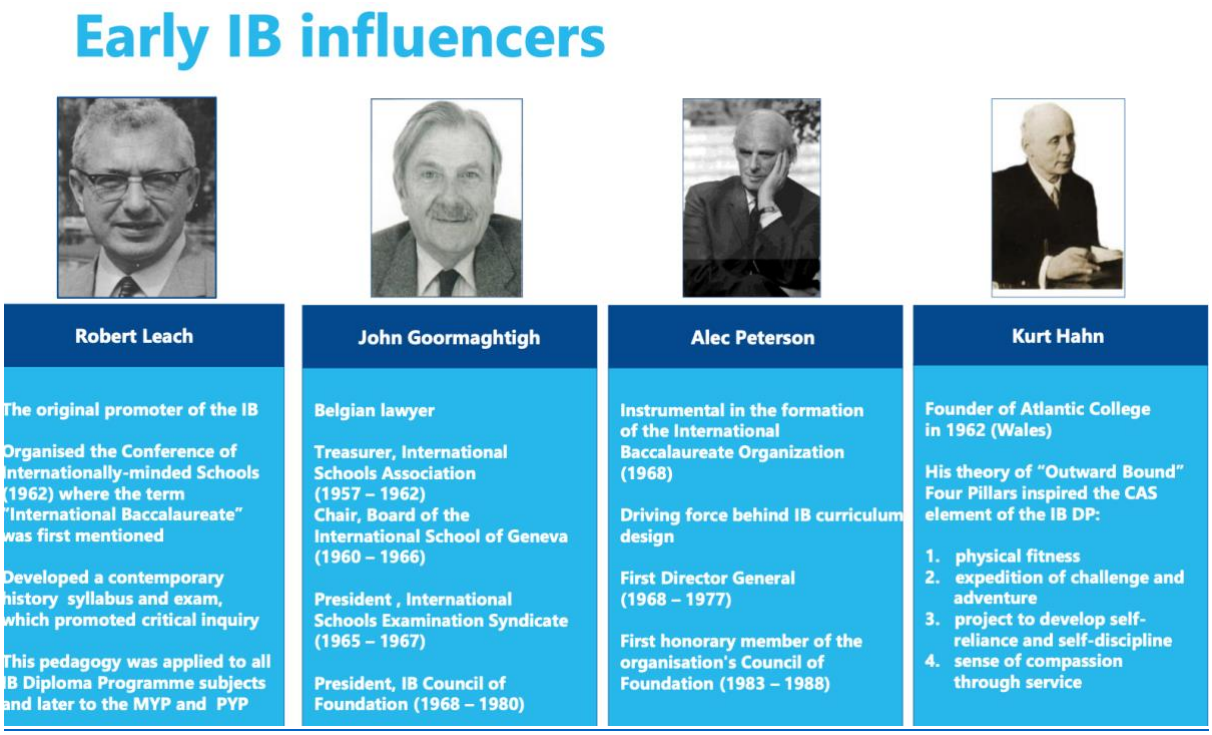
APPENDIX 4: KEY INFLUENTIAL EDUCATIONALISTS



International Baccalaureate (2017): The history of the IB

https://www.ibo.org/globalassets/digital-toolkit/presentations/1711-presentation-history-of-the-ib-en.pdf

APPENDIX 5: EARLY IB INFLUENCERS



International Baccalaureate (2017): The history of the IB

https://www.ibo.org/globalassets/digital-toolkit/presentations/1711-presentation-history-of-the-ib-en.pdf

APPENDIX 6: INTERNAL ASSESSMENT (IA) CRITERIA – HIGHER LEVEL (HL)

Criterion A: Language

To what extent is the vocabulary appropriate and varied?
To what extent are the grammatical structures varied?
To what extent does the accuracy of the language contribute to effective communication?
To what extent do pronunciation and intonation affect communication?

Marks	Level descriptor
0	The work does not reach a standard set by the descriptors below
1-3	**Command of the language is limited** • Vocabulary is sometimes appropriate to the task. • Some basic grammatical structures are used, with some attempts to use more complex structures. • Language contains errors in both basic and more complex structures. Errors interfere with communication. • Pronunciation and intonation are generally clear but sometimes interfere with communication.
4-6	**Command of the language is partially effective** • Vocabulary is generally appropriate to the task, and varied. • A variety of basic and some more complex grammatical structures is used. • Language is mostly accurate for basic structures, but errors occur in more complex structures. Errors at times interfere with communication. • Pronunciation and intonation are generally clear.
7-9	**Command of the language is effective and mostly accurate** • Vocabulary is appropriate to the task, and varied, including the use of idiomatic expressions. • A variety of basic and more complex grammatical structures is used effectively. • Language is mostly accurate. Occasional errors in basic and in complex grammatical structures do not interfere with communication. • Pronunciation and intonation are mostly clear and do not interfere with communication.
10-12	**Command of the language is mostly accurate and very effective** • Vocabulary is appropriate to the task, and nuanced and varied in a manner that enhances the message, including the purposeful use of idiomatic expressions. • A variety of basic and more complex grammatical structures is used selectively in order to enhance communication. • Language is mostly accurate. Minor errors in more complex grammatical structures do not interfere with communication. • Pronunciation and intonation are very clear and enhance communication.

Criterion B1: Message - literacy extract

How well does the candidate engage with the literary extract in the presentation?

Marks	Level descriptor
0	The work does not reach a standard set by the descriptors below
1-2	**The presentation is mostly irrelevant to the literary extract.**

	• The candidate makes superficial use of the extract. Observations and opinions are generalized, simplistic and mostly unsupported.
3-4	**The presentation is mostly relevant to the literary extract.** • The candidate makes competent use of the literary extract. Some observations and opinions are developed and supported with reference to the extract.
5-6	**The presentation is consistently relevant to the literary extract and is convincing.** • The candidate makes effective use of the extract. Observations and opinions are effectively developed and supported with reference to the extract.

Criterion B2: Message - conversation

How appropriately and thoroughly does the candidate respond to the questions in the conversation?
To what depth are the questions answered?

Marks	Level descriptor
0	The work does not reach a standard set by the descriptors below
1-2	**The candidate consistently struggles to address the questions.** • Some responses are appropriate and are rarely developed. • Responses are limited in scope and depth.
3-4	**The candidate's responses are mostly relevant to the questions.** • Most responses are appropriate and some are developed. • Responses are mostly broad in scope and depth.
5-6	**The candidate's responses are consistently relevant to the questions and show some development.** • Responses are consistently appropriate and developed. • Responses are broad in scope and depth, including personal interpretations and/or attempts to engage the interlocutor.

Criterion C: interactive skills - communication

How well can the candidate express ideas?
How well can the candidate maintain a conversation?

Marks	Level descriptor
0	The work does not reach a standard set by the descriptors below
1-2	**Comprehension and interaction are limited.** • The candidate provides limited responses in the target language. • Participation is limited. Most questions must be repeated and/or rephrased.
3-4	**Comprehension and interaction are mostly sustained.** • The candidate provides responses in the target language and mostly demonstrates comprehension. • Participation is mostly sustained.
5-6	**Comprehension and interaction are consistently sustained.** • The candidate provides responses in the target language and demonstrates comprehension. • Participation is sustained with some independent contributions.

Marking criteria:

Criterion A	Language	12 marks
Criterion B1+B2	Message	6+6=12 marks
Criterion C	Interactive skills	6 marks
	Total	**30 marks**

What I'm doing well...	**What I need to improve on...**

Adapted from the Language B guide (First assessment 2020)

http://www.cosmopolitanschool.de/wp-content/uploads/2018/10/dp-language-B-guide.pdf

APPENDIX 7: EXAMPLE OF EXTRACT THAT STUDENTS ARE GIVEN TO PERFORM THE INTERNAL ASSESSMENT ORAL TASK

EXTRACT 1: LOUIS SACHAR: HOLES

Two nights later, Stanley lay awake staring up at the star-filled sky. He was too happy to fall asleep.

He knew he had no reason to be happy. He had heard or read somewhere that right before a person freezes to death, he suddenly feels nice and warm. He wondered if perhaps he was experiencing something like that.

It occurred to him that he couldn't remember the last time he felt happiness. It wasn't just being sent to Camp Green Lake that had made his life miserable. Before that he'd been unhappy at school, where he had no friends, and bullies like Derrick Dunne picked on him. No one liked him, and the truth was, he didn't especially like himself.

He liked himself now.

He wondered if he was delirious.

186

He looked over at Zero sleeping near him. Zero's face was lit in the starlight, and there was a flower petal in front of his nose that moved back and forth as he breathed. It reminded Stanley of something out of a cartoon. Zero breathed in, and the petal was drawn up almost touching his nose. Zero breathed out, and the petal moved toward his chin. It stayed on Zero's face for an amazingly long time before fluttering off to the side.

Stanley considered placing it back in front of Zero's nose, but it wouldn't be the same.

It seemed like Zero had lived at Camp Green Lake forever, but as Stanley thought about it now, he realized that Zero must have gotten there no more than a month or two before him. Zero was actually arrested a day later. But Stanley's trial kept getting delayed because of baseball.

He remembered what Zero had said a few days before. If Zero had just kept those shoes, then neither of them would be here right now.

APPENDIX 8: CONSENT FORM SENT TO THE PARTICIPANTS' PARENTS

Barlovento, s/n
Urb. Almerimar
El Ejido. 04711 Almería

T +34 950 49 72 73
sek-alboran@sek.es
www.sek.es

Estimadas familias,

Dña. María Dolores Salmerón Carmona, profesora de Inglés en Bachillerato y del Programa del Diploma, se halla cursando el Doctorado en Educación en la Escuela Internacional de Doctorado de la Universidad de Almería (EIDUAL).

Dentro de su trabajo de investigación precisa incluir muestras y evidencias reales de trabajos de alumnos, que serán reflejados siempre desde el anonimato, y tratados desde el punto de vista pedagógico. Para poder hacerlo, se necesita la autorización de los padres/madres/tutores por lo que, si quieren colaborar con este trabajo, les pedimos firmen la autorización que se adjunta abajo.

Agradeciendo su colaboración, reciban un cordial saludo.

Luis Carlos Jiménez, Director

Yo, padre/madre/tutor _____ con DNI-_____, doy mi autorización para que se pueda incluir alguna actividad de Inglés del alumno/a_____ en el trabajo de Doctorado de Dña. María Dolores Salmerón Carmona.

APPENDIX 9: INTERNAL ASSESSMENT PRACTICE: PEER-ASSESSMENT PROFORMA (1ST ONE)

Work in pairs: during the presentation of your partner take notes in the spaces provided to give them later some feedback. Write not applicable if you are unable to infer or does not apply to the extract your partner has chosen.

How did the students present their opinion on the **characters**?

How did the students present their opinion on **the event of the extract?**

How did the students present their opinion **on the ideas and themes of the extract?**

How did the students make any **connection with the themes studied in class**? Identities, experiences, etc.

How did the students **centre t**he discussion on the extract and the work?

Command of the language

How easily can one work out what meaning is intended, or how much of an 'effort of translation' one has to make?

How **rich or complex or subtle** is the meaning conveyed - are there **nuances or** implications or **irony**?

Vocabulary

How **appropriate** is the vocabulary?

How **varied**?

How **nuanced**?

Do they use **idiomatic expressions**?

Grammar

Do they use **complex** grammatical structures?

Do errors in **basic** grammatical structures **interfere with communication?**

Do errors in **complex** grammatical structures **interfere with communication**?

Pronunciation and intonation

Do pronunciation and intonation **interfere with communication?**

Do pronunciation and intonation **enhance communication**?

Is the presentation consistently relevant to the literary extract and convincing?

Does the candidate make effective use of the extract to develop and support observations and opinions?

Message conversation

How much does your partner contribute beyond a simple relevant response? Are your partner's responses relevant to your questions and do they show some development?

Are the responses broad in scope and depth including personal interpretations and/or attempts to engage you? How?

How much does your partner interact actively with your ideas (e.g., asking questions back in order to clarify the initial question, or disagreeing)?

How much does your partner contribute to the flow of the conversation – do they understand your questions and ideas in the first place, and then how much do they have to contribute?

APPENDIX 10: STUDENTS' OWN QUESTIONS

Learner 9

2. What is the role of new technologies in climate change? Do you think that we need to use them and encourage investigation and development of other new technologies?
3. To what extent do you think that globalization is positive for our society?
4. In your opinion, what are the bases of a good relationship?
5. Does concrete art need to be more recognized than abstract art?
6. Do you feel there is pressure in our current society to portray a certain type of identity across social media?
7. How do you think that massive ideologies influence society?
8. To what degree does your society class when you born, determine your sense of identity?

Learner 2

1. How is an identity built in a society?
2. Does following trends create an identity in a society?
3. To what extent does society influence when choosing its own identity?
4. To what extent does society influence when choosing its own identity?
5. Why can we have false memories of childhood?
6. What is more important for memory, store data or create an identity?
7. How do fictional memories occur?
8. How do memory and personal history interact?
9. Why do researchers recommend being skeptical of our own memories?

Learner 11

1. Would you say that the topic of immigration is controversial? How so?

What do you think are the fundamental parts that make up a person's identity?

How much of your identity is made up by physical characteristics like your appearance?

In what ways can education help to alleviate some of the issues that minority groups face?

Tell me about the differences between marriage culture in different countries. (e.g. arranged marriage, forbidden same sex marriage)

In what ways can charity cause more problems for developing countries than actually help them?

Which generation do you think is most concerned about environmental issues? Why?

To what degree does the colour of your skin determine your sense of identity?

How would you define what art is?

To what extent is your identity defined by your beliefs, traditions, culture?

Learner 6

Which is more important to a person's identity, their interior characteristics (personality, upbringing etc.) or their exterior characteristics (appearance)?

Do you feel there is pressure in our current society to portray a certain type of identity across social media?

Do you believe that migration has a beneficial impact on countries? Why/why not?

To what extent is our definition of 'beauty' shaped by the media?

Why do you think people feel the need to take unnecessary risks in sports?

Learner 4

How do you think Stanley developed his personality during the extract and how do you think it affected to the successes of this extract?

Do you that this extract could in some way show if whether the camp was helpful or not for Stanley?

To what extent does identity matter on the development of the extract?

What is the author trying to do in this extract in order to give some sense to the following events of the book

APPENDIX 11: EXAMPLE OF THE WAY INTERVIEWS WERE CODED

Learner 4 feedback for Learner 6

Yellow: Benefits for the peer assessed and the assessor

P (praise)

CF (critical feedback)

SF (Specific feedback)

Blue: Constructing their own learning

RL (Reflection of their own learning, students' acceptance of their own mistakes)

IC (Internalisation of the criteria)

I (Interactions)

Pink: Emotional responses (How the students feel practising peer-assessment)

PA (Positive attitude)

R (Reluctance)

I'm gonna evaluate L6's internal assessment. In general, it was fine (P) but there are a few things that weren't really good (CF) and so, she didn't get one of the highest marks in my opinion and this is why, so for example in criterion A I think she used really good language and she used accurate vocabulary and a wide variety of nouns, adverbs, adjectives and everything. Also, I think she used complex grammatical structures and they were generally great, they were never bad but sometimes I think she didn't use them in the right place, she like kind of disorganised them and use them where when they didn't really fit so there weren't used in the right context (IC)Also, I think communication, her communication was really good because at every moment I did understand her that's really good for her and so as Criterion A wasn't perfect, but it was very good I think I would give her a 9/12. (IC)

So then in criterion B I think the ideas were explained well, everything was clear(P) but it wasn't really really good because sometimes you didn't really get what she was talking about(CF) but besides that more or less it was good and she really used the literary extract, quoted it, that's really good and that made me for example understand what she was talking(P) about and she made some points that she thought for herself which were really interesting (PA)and that really fit it to the context but I also have to say that the use of the extract wasn't really extended, she used it but it wasn't a really big thing, so that's why I think she didn't get the highest mark in this one, so I think she would get in criterion B1 a 4/6 (SF) (IC).

Then criterion B2 I think she did really good, the ideas were relevant to the conversation, there weren't any, there weren't any misunderstandings, I think I understood everything, and the ideas were developed the responses were consistent, appropriate and well developed by her and I think that's really important in this criterion and that's why I think she deserves 5/6 (IC) because she really interacted with the interlocutor (I, P) so, that's why I think she gets 5/6 (IC)

Then I think that in criterion C she didn't really do that well, she wasn't as good as in the other criteria, (IC) that's mostly because she wasn't really in the mood of interacting (R)with the person asking the questions, sometimes I got that feeling just listening to the extract, just listening to her oral assessment because sometimes I think she spoke really slow and that kind of made it boring and so this loses the interest for the extract and I think that's a really bad thing because you always wanna keep people interacting with you interested in what you are saying so that they look at your work like more enthusiastic and so you get a better mark, so I think that didn't interfere with communication but it interfered the understanding and the focus over the extract, you weren't kind of interesting with that, so that is why in criterion c I would give her a 4/6 (SF, IC).And that's it, so in general I think it was good but I think there are things that should be improved and that's why I said this recording, so yea that's it. (PA)

Learner 6 feedback for Learner 2

L6: I think that he did very well, (P). I think he also committed the same mistakes as me , because we didn't use like a structure. The gap year told us to go from the book, a summary of

the book, a summary of the extract, characters and opinions, and I couldn't stop to go from more to less. And I think neither of us did that, so I think that's a thing we should do (RL) But apart from that, he used very well vocabulary, he talked about the book, the extract and he also related identities with the characters of the book (P) And that's something that I really liked about that (PA) IC

TEACHER: was there a bit of interaction?

L6: Some of it, but not everything.

TEACHER: So now you know exactly how you have to do it?

BOTH **L6** AND **L2**: Yes. (RL)

Learner 11 feedback for Learner ***5***

L11: like he, like presented his opinion on the characters, but also like the story as a whole, the extract why, like the importance of the extract within the whole story. So, in that aspect he did it really well. (SF) (IC) What like the thing, the only thing that I see is that he should extend a bit more like in the presentation, because it was, it did n't reach like 4 minutes. The presentation as a whole it didn't reach 4 minutes. (SF) (C) (IC)

TEACHER: But, did he express his opinion about the characters? The event? Did he interpret the …?

L11: Yes like, in that aspect he did it really well. (P)Like, he talked about was, like the extract there was like two characters I think, three, and he talked about like every single one, the importance in the extract in the whole story and yes, I think that. (SF) IC

TEACHER: Any kind of images? Did he talk about anything relevant? Did he also explain the beginning, I mean, the story itself? Would you have understood the story if you hadn't read the book?

L11: Yes, because the gap year student, I think she understood, and she didn't know about this book. But like firstly he like, put the extract in a context before explaining the extract, because if not we would be lost. (SF) IC

TEACHER: What about the grammar and the language? I mean, obviously his pronunciation is good, but did you have to make any effort to understand him or not?

L11: No, you can understand him really perfect. (P)

L5: I think I didn't use any expressions (RL)

TEACHER: And in the end when you started to ask him questions, was there any kind of interaction?

L11: Yes, but I don't remember what I asked. (I, T)

L5: You asked about the importance about this extract at the end of the story. (I, T)

L11: Ah, yeah, I remember I asked him which like, if he would (I, T, EQ)

L5 INTERRUPTS: If I would change the end of the story (I, T)

L11: And like, by that, we established like a little conversation. (IC, I)

TEACHER: So, because what you have to do is try to make the conversation as likely and even if you want to ask any questions. My job or your job was like to make you develop your ideas as much as possible. So, in the end, some advice for the next time?

L11: Yes, yes that like to extend a bit more because he didn't reach the 4 minutes. Like for example me, I extended a lot. (IC, RL)

TEACHER: So, what happened with her **L5**, could you give me your feedback? How did she do?

L5: Very good (P) I liked a lot her summary (PA)before like, the context before her extract and then she connected it really well. Then she gave her opinion on Zero, Stanley and the group of kids that were with them (SF)

L11: Like I developed too much. (RL)

TEACHER: You developed your ideas too much?

L11: Yes, because I like, the time was longer than 4 minutes for sure. (IC)

L5: She ran out of time (SF).

TEACHER: So, when he asked you the questions, you developed your ideas too much?

BOTH **L11** AND **L5**: Not in that time, before the questions. (SF)

TEACHER: It doesn't matter if you develop your ideas with the questions, that's good.

L11: Like he had to stop me because I was… (SF, IC, I)

TEACHER: talking and talking. Do you think now that you understand exactly what you have to do?

L11: Yes, because before doing it I should like, time what I want to say and practice before, because I over-talk. (IC, RL)

TEACHER: And you?

L5: I think I need to spend longer in the summary, but the connection of the ideas and answering the questions was quite good. (IC, RL)

TEACHER: Thank you

APPENDIX 12: EXAMPLE PAGE OF THE DIARY

STAGE 1

28th November: 55' Although I had already explained the students what the internal assessment of the subject English B HL oral assessment consists of at the beginning of the Academic year, I do it again. It is explained in the chart below taken from the International Baccalaureate Organization Language B guide, p 50 (first assessment 2020).

,

Internal Assessment: HL

Supervised preparation time	The student is shown two extracts of up to approximately 300 words each: one from each of the two literary works studied during the course. The student chooses one of the extracts and prepares a presentation focused on the content of the extract. During this time, the student is allowed to make brief working notes.	**20 minutes**
Part 1: Presentation	The student presents the extract. The student may place the extract in relation to the literary work, but must spend the majority of the presentation discussing the events, ideas and messages in the extract itself.	**3–4 minutes**
Part 2: Follow-up discussion	The teacher engages with the student on the content of the extract that the student has presented, expanding on observations that the student has provided in the presentation.	**4–5 minutes**
Part 3: General discussion	The teacher and student have a general discussion using one or more of the five themes of the syllabus as a starting point.	**5–6 minutes**

This time the focus will be on Part 1: Presentation

Each student is told to choose an extract of about 300 words that they consider relevant in the book. They start looking for it. Once they have it, they have to send it to me. Then they can start preparing the presentation individually. Once they have prepared it, they have to record themselves in flipgrid (an online programme). They are allowed to have the extract with when doing the presentation and a paper with 10 bullet points at the most. They are not allowed to read what they have prepared.

29th November: 45': The students who are prepared start recording themselves with flipgrid. Others are not so good with technology and record themselves with their mobiles. Those who don't finish in the class have to do it for homework.

3- 4 December: 1 hour and 50' The students start listening to their classmates' videos on flipgrid. Some of them didn't upload them on flipgrid but they sent them to me. It is surprising that the 2 students who are usually shier to speak recorded their presentations on flipgrid. They write some feedback (open comments) on Onenote (another online programme) because Flipgrid only allows you to give feedback recording yourself. They are not very used to giving oral feedback so to record themselves giving feedback to their classmates was a bit weird for them. I could have insisted but I didn't see the need and thought that written feedback would be also very valid.

APPENDIX 13: PROFORMA (RUBRICS): FORM THAT STUDENTS COMPLETE ABOUT THEIR PEERS' WORK AND RETURN TO THEM AS FEEDBACK (2ND)

A: Language

How appropriate and varied is the vocabulary?

Do they use complex grammatical structures?

Do errors in grammatical structures interfere with communication?

Do pronunciation and intonation interfere with communication?

B1: How well the students engage with the literary extract

B2: Message: conversation

Are your partner's responses relevant to your questions and do they show development?

Do your partner's answers include personal interpretations and attempts to engage you?

C: Interactive skills:

How much does your partner interact actively with your ideas asking questions back in order to clarify the initial question, disagreeing, etc?

Do they understand your questions? How much do they contribute to the flow of the conversation?