ENGLISH AS A FOREIGN LANGUAGE THROUGH WHOLE BRAIN TEACHING IN PRIMARY SCHOOL

Edward Alvar Lockhart Domeño

Edward A. Lockhart

ENGLISH AS A FOREIGN LANGUAGE THROUGH WHOLE BRAIN TEACHING IN PRIMARY SCHOOL

DOCTORAL DISSERTATION

Supervised by Dr. Joaquín Romero Gallego
Department of English and German Studies

UNIVERSITAT ROVIRA i VIRGILI

Tarragona
2016

UNIVERSITAT ROVIRA I VIRGILI
ENGLISH AS A FOREIGN LANGUAGE THROUGH WHOLE BRAIN TEACHING
IN PRIMARY SCHOOL
Edward Alvar Lockhart Domeño

UNIVERSITAT
ROVIRA i VIRGILI

Departament d'Estudis Anglesos i Alemanys

Av. De Catalunya, 35
43002 Tarragona, Spain
Tel. 977 559 755

I hereby certify that the present study, *English as a Foreign Language through Whole Brain Teaching in Primary School,* submitted by Edward A. Lockhart Domeño in partial fulfillment of the requirements for the degree of Doctor of Philosophy, has been carried out under my supervision at the Department of English and German Studies (Universitat Rovira i Virgili).

Tarragona, June 10, 2016

Doctoral dissertation advisor

Dr. Joaquín Romero Gallego

UNIVERSITAT ROVIRA I VIRGILI
ENGLISH AS A FOREIGN LANGUAGE THROUGH WHOLE BRAIN TEACHING
IN PRIMARY SCHOOL
Edward Alvar Lockhart Domeño

To you, dad. As I once told you, the good things I have are because of you ...

UNIVERSITAT ROVIRA I VIRGILI
ENGLISH AS A FOREIGN LANGUAGE THROUGH WHOLE BRAIN TEACHING
IN PRIMARY SCHOOL
Edward Alvar Lockhart Domeño

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Edward A. Lockhart

June 2016

UNIVERSITAT ROVIRA I VIRGILI
ENGLISH AS A FOREIGN LANGUAGE THROUGH WHOLE BRAIN TEACHING
IN PRIMARY SCHOOL
Edward Alvar Lockhart Domeño

Acknowledgements

I would like to acknowledge *Universitat Rovira i Virgili* and the Department of English and German Studies because they trusted me with a *Martí i Franqués* grant that enabled me to focus on this long and hard task.

My deepest gratitude to my PhD supervisor, Joaquín Romero, for his trust, support, help and orientation all throughout the process. Some difficult times I had the feeling that he believed in me more than myself.

This dissertation would have not been possible without the collaboration of the *Escola Joan Ardévol*. Enric Masdeu, the principal, was always helpful and willing to help. Nuria Capella, the English teacher, was willing to innovate and to incorporate Whole Brain Teaching into her practice. The kids from 4th grade were patient and collaborative.

Mercè Bernaus and Tim Murphey are experts who have also given me altruistic guidance that helped me to keep going in moments when I was stuck.

Ana Jacas, Crucita Bretón, Jackeline Katzeb, Jessica J. Lockhart and Louis Ortega, thank you for helping me out with the pronunciation accuracy tests.

Francesc Esteve crossed my path when I most needed help with statistics and allowed me to re-engage the whole project,which had been blocked way too long. Diego Parga has also been a great help whenever a statistical doubt arose.

Special thanks to José Luis Becerril, who introduced me to LaTeX and made the editing of this dissertation more complex but also much more entertaining and instructive.

Of course, I also want to thank my family: my wife, Núria, and my daughter, Leyre (her especially, for sometimes I had to sacrifice witnessing part of her childhood). But also my mother, Menchu, and my sister, Wendy (my tandem). They have understood the hours of dedication and they have supported me all throughout the process.

I might be forgetting some important people here. If this is the case, forgive me, because it was not intentional.

Abstract

The purpose of this dissertation is to explore how the use of the different tools and techniques inside the Whole Brain Teaching methodology affects the process of learning English as a foreign language in mainstream primary education. Whole Brain Teaching is a methodology that was designed to teach subjects through the mother tongue of the students. Nevertheless, the extensive use of gestures, the frequent pair work, the motivational tools and the fact that the learners often have to repeat what the teacher explains, suggest that this methodology could enhance the teaching and learning process of English as a foreign language both in terms of general language acquisition and the motivation of the learners. Two research hypotheses are formulated. The first one explores whether Whole Brain Teaching can enhance the process of language acquisition regarding the language skills of the learners. The second research hypothesis aims to see how the use of Whole Brain Teaching techniques affect the motivation of the learners.

These hypotheses were tested in a mainstream primary school with an experimental design that involved a control and an experimental group. During three months, both groups received the same content, but only the experimental group received it through Whole Brain Teaching techniques. The study was performed from a mixed-methods perspective, with a quantitative part that measured the changes in the English command of the learners and the changes in their motivation, and a qualitative part that explored the whole process and the impressions of the teacher to support or

question the quantitative results and to give a better insight of what happened in the process. The changes in the English command were measured through listening, reading and writing tests designed by the publisher of the textbooks used in the school and through a speaking test to check the fluency, use of vocabulary and pronunciation of the learners. The changes in motivation were assessed through an adaptation of the mini-AMTB (Attitude Motivation Test Battery). The mini-AMTB was adapted to suit the characteristics of nine year-old learners and was piloted in the same school with two other groups. Nevertheless, the results after the experiment suggested that the test was not as reliable as it had seemed and it was piloted again, this time in a pre-test/post-test situation that proved the unreliability of the test.

The results partially support the first hypothesis, showing significant differences regarding the listening and the reading and writing skills of the learners. Regarding the speaking skills, however, that was not the case. As far as the second hypothesis is concerned, the quantitative results are, although non-significant, negative for the experimental group. On the other hand, the qualitative part of the study and the reviewed literature question those results.

Enough evidence has been found to support the potential of the Whole Brain Teaching method for the instruction of English as a foreign language in primary school.

Table of contents

UNIVERSITAT ROVIRA I VIRGILI
ENGLISH AS A FOREIGN LANGUAGE THROUGH WHOLE BRAIN TEACHING
IN PRIMARY SCHOOL
Edward Alvar Lockhart Domeño

List of figures

UNIVERSITAT ROVIRA I VIRGILI
ENGLISH AS A FOREIGN LANGUAGE THROUGH WHOLE BRAIN TEACHING
IN PRIMARY SCHOOL
Edward Alvar Lockhart Domeño

List of tables

Chapter 1

Introduction

The main objective of this dissertation is to test whether Whole Brain Teaching, a methodology designed for improving the classroom management of mainstream teachers, has the potential of improving the classes of English as a foreign language in a setting of primary education. It will hopefully provide enough evidence on how the different tools and techniques from the method affect the speed and quality of language acquisition and how they affect some intrinsic acquisition factors in the learners.

1.1 Literature Review

Most authors (see Richards & Rodgers, 2001) agree that we are in the post-method era. After a whole century seeking the perfect method that would allow teachers to be certain they were teaching in the best possible way, researchers have realized that this perfect method does not exist. Each has positive things and not so positive ones. Some of them will work under certain circumstances and with certain types of learners but not with others. Brown (2002) talks about having a principled (more eclectic) approach rather than following a strict method. He mentions twelve principles: automaticity, meaningful learning, the anticipation of reward, intrinsic motivation,

strategic investment, language ego, self-confidence, risk taking, the language-culture connection, the native language effect, interlanguage and communicative competence. In his work, he recommends having an approach that respects these principles using bits and pieces of various methods when they support our objectives rather than sticking to just one single method.

Other authors talk about different principles we should consider when teaching English. Gardner (1983) described seven different intelligences. This list has recently been updated to eight or even nine, if considering existential intelligence (Gardner, 1999). Gardner (1983) recommends that teachers take all of these into account when teaching. Felder and Henriques (1995) refer to five dichotomies that can help the teacher reach all the students in a more holistic way. Krashen and Terrell (1983) talk about the importance of providing comprehensible input so the students can acquire the language. They also mention the importance of motivation in the process of second or foreign language acquisition. All these principles seem to promote acquisition in most educational contexts. Following is a more detailed overview of the theories and studies that influence the teaching of English nowadays and that include some of those universal principles for teaching English as a Foreign Language (EFL henceforth) that were commented on before.

1.1.1 The Natural Approach

The Natural Approach is one of the most widely accepted theories in terms of language acquisition (Lightbown & Spada, 2006). According to this theory a learner must receive immense quantities of comprehensible input to acquire the language (Ellis, 1985). Specifically, the input has to be only one step above the interlanguage of the learners (i+1). This is known as the Input Hypothesis.

In an English teaching process, though, we need to consider more things than just the input. In fact, if it were the only thing to take into account, all the learners with the same interlanguage and in the same class (that is, with the same amount and quality of input) would acquire the language at the same rate. This does not happen because each student has a different attitude towards the foreign language acquisition. The Affective Filter Hypothesis inside the Natural Approach (Krashen, 1982) is one of the theories that better explains these individual differences. In this hypothesis Krashen talks about 'motivation and self-confidence' as some of the factors that will determine the degree of success (and rate) of acquisition.

Skehan (1989) places motivation as the second most important factor to predict the success in language acquisition. He states that motivation is only surpassed by aptitude. Motivation is a dynamic process that changes depending on external factors (Gass, 2008). According to Dörnyei (2005), it seems motivation is not something exclusively personal about the learner, but that it is affected also by the group dynamics inside the classroom. He also talks about the importance of 'devising motivational strategies' that will make teachers 'good enough motivators' . Some of the techniques he recommends as motivational strategies include 'a pleasant and supportive classroom atmosphere', 'a cohesive learner group with appropriate group norms', 'increasing the learners' expectancy of success', 'making learning stimulating', 'presenting tasks in a motivating way', 'promoting cooperation among the learners', amongst others.

The Natural Approach distinguishes between 'learning' and 'acquisition', also known as the Learning versus Acquisition Hypothesis. Krashen and Terrell (1983) refer to the first one as a conscious, cognitive process that is relatively fast but quickly forgotten. Things can be learned in an arbitrary order but with the drawback that if you want to use a learned form you need to invest time to retrieve it from your memory. They see language acquisition as the opposite: a subconscious process that takes a long time but

is quite permanent. Things are acquired in a natural order (very similar to the native speakers of the target language) and with the advantage that what you acquire comes out spontaneously whenever you need to use it. Some other authors (see Ellis, 1985) prefer to use either term interchangeably without a clear distinction between them because they do not see enough proof of the differences stated by Krashen and Terrell.

Another important hypothesis in this approach is the Natural Order Hypothesis. According to Lightbown and Spada (2006, p. 37) it is 'based on the finding that, as in first language acquisition, second language acquisition unfolds in predictable sequences'. Regarding this hypothesis, Ellis (1985, p. 9) says that 'Both Error Analysis and the longitudinal studies show that there are striking similarities in the ways in which different L2 learners learn an L2 [...]. This route resembles that reported for L1 acquisition but is not identical with it'.

The fifth hypothesis of the Natural Approach is the Monitor Hypothesis. According to Krashen (1988), if a student learns a form instead of acquiring it, she will need time to retrieve it when she wants to use it. This will happen whenever she feels the need to use that specific form. Ellis (1985) indicates that this theory has been attacked and that 'it is seen as too narrow, in the sense that the learner is clearly able to edit his performance using implicit as well as explicit knowledge' (Ellis, 1985, p. 179).

In line with the Natural Approach theory, Terrell (1986) talks about the difference between passive vocabulary (the one that enables the learner to understand but that she cannot use unless with time and/or help) and active vocabulary (the one that the learner can both understand and use autonomously). He describes the process of acquiring the former as 'binding' the language and the latter as 'accessing' the language. Terrell considers that both processes are subconscious. He states that most binding of words happens through an acquisition process while only some would be explained by learning. He states that in the accessing process only acquisition is involved and that

it is essential that the word to be accessed be in the passive language of the learner. Accessing, thus, would be the process of transferring the passive vocabulary into active. In the process of communication both the passive and the active vocabulary are essential, but the latter is basic when trying to produce. Fluent users of the language need immense quantities of active vocabulary.

The Natural Approach has been challenged by both psychologists and linguists (Lightbown & Spada, 2006, p. 38). Psychologists and linguists argue that this method cannot be tested empirically and that it is too imprecise (White, 1987). Some other authors, like Higgs and Cliffort (reported by Ellis, 1997) accuse Krashen and Terrell's theory of causing an early fossilization of the language due to its communicative tendency and its criticism toward a conscious learning of grammar.

1.1.2 The Communicative Approach

Some authors consider the Natural Approach one of the methods and approaches inside a wider group: the communicative approach (Richards & Rodgers, 2001). This approach was formerly known as 'Communicative Language Teaching' but its general suggestions and principles turned it into an approach. Richards and Rodgers (2001, p. 151) say that its 'general principles [...] are today widely accepted around the world'. They talk about three elements that teachers should consider when selecting or designing activities: the communication principle, the task principle and the meaningfulness principle.

According to Lightbown & Spada (2006, p. 110) the communicative approach 'places the emphasis on interaction, conversation, and language use, rather than on learning about the language'. They also say that the focus in this approach is on fluency rather than on accuracy. To achieve this, the teacher adapts the input to make

it comprehensible. This adaptation can come through 'the use of contextual cues, props, and gestures' (2006, p. 113).

Teachers only correct the mistakes that can harm the communicative process. They rely on the negotiation of meaning for the students to realize their errors and to reformulate them in a correct way. Teachers tend to use pair-work and group-work so students can have more chances to communicate and, thus, to negotiate meaning. To do so, the activities proposed in class involve 'authentic and meaningful communication' (Richards & Rodgers, 2001, p. 172). This view of not correcting mistakes has been criticized by many authors (see Doughty & Long, 2003), identifying it as one of the possible causes of fossilization.

One of the theories that argued against a strict communicative approach was the one developed by Cummins (1979) in which he distinguished between 'Basic Interpersonal Communicative Skills' and 'Cognitive/Academic Language Proficiency'. The former seems to be acquired through contact with the language in communicative situations while the latter seems to need a more formal learning setting and does not depend only on communication. Ellis (1997, pp. 52,53) goes even further and says that 'The studies constitute evidence only that full grammatical competence does not seem to develop in communicative classrooms', although he also comments that 'the problem may rest with the learners rather than with the learning environment'.

1.1.3 Total Physical Response

Total Physical Responce (TPR henceforth) was designed by James Asher in the 1970s. It is a grammar-based method that relies on the use of the imperative form of the verb and the link of that with physical actions or gestures (see Section 1.2, Gestures in Language Teaching). According to Richards & Rodgers (2001), Asher listed three learning hypotheses:

- The mind is genetically prepared to learn languages with a natural path to acquire them (very similar between L1 and L2). In this path the learners learn through listening before speaking, they need to respond physically to oral commands and they are able to speak easily if they have developed good listening comprehension skills.

- The brain has different learning functions depending on the hemisphere. TPR is more right-brain directed while most second language acquisition (SLA henceforth) methods are left-brain oriented. Language functions are basically controlled by the left hemisphere (Taylor & Taylor, 1990) while the right hemisphere shows a superiority in dealing with visual stimuli and spatial manipulation (Beaumont, 2008). They use movement because 'right-hemisphere activities must occur before the left hemisphere can process language for production' and movement is controlled by the right-hemisphere.

- Students will learn more if they have a low level of stress. To do this, teachers try to somehow replicate the conditions of first language acquisition in which students were relaxed and enjoyed positive experiences.

In TPR, learners are encouraged to physically act out the commands that the teacher gives. For those actions to be realistic, the teacher will use realia and/or pictures and will relate the command to these. Students are not forced to produce and they can start speaking when they feel prepared.

TPR is thought to work mainly with children, only at beginner levels and only with the imperative form of verbs. Asher (2014), in the TPR website, claims that these three statements are false. He states that not only does TPR work with adult learners, it also works for all levels and for any verb tense and grammar feature.

1.1.4 Neuro-Linguistic Programming

Neuro-Linguistic Programming (NLP henceforth) did not start as a a teaching or a learning method, but rather as a line of research to try to identify why certain people were especially able to influence others. Bandler & Grinder (1979) identified four important aspects:

- *Rapport*: the positive and harmonious connection between people or inside a group.
- *Outcomes*: successful therapists had clear objectives and followed them.
- *Flexibility*: if something was not working they were ready to try something different instead of continuing trying the same thing over and over.
- *Sensor acuity*: those therapists were able to identify what sensor preference their interlocutor had and adapted their speech and actions to it.

Although NLP was not designed for education, teachers soon saw that, if they applied some of the techniques, they managed to reach students in a more powerful way (Churches & Terry, 2007). One of the most extended NLP theories amongst educators is the one related to the perceptive channels (also known as VAK: Visual, Auditory, Kinesthetic). Teachers adapt their activities to include stimuli of one type or another or, often, of all three together. Moreover, sometimes teachers include other perceptive channels (smell and taste) that can enhance the feeling and, thus, the memorability of the learning.

Most of the recent educational research does not show any direct improvement of teaching by taking into consideration the perceptive preferences of the learner separately, especially when trying to teach a group mainly through their preferred channel. However, it seems that by using the sensory learning styles we improve the

group rapport, which eventually increases the motivation of the students (Churches & Terry, 2007).

1.1.5 Latest Trends

Scott Thornbury stirred the world of language teaching and methodology with his controversial 2000 article called 'A Dogma for EFL' in which he advocated going back to the pre-method era and avoid depending so much on tools and techniques (Thornbury, 2000). He recommended putting the learners back in the center of the process. He advocated for allowing the students to talk about things that could be interesting to them. Furthermore, in 'Dogme' learners should become part of the planning of the lessons. They should provide materials and decide on the contents.

Thornbury's article and some of his conferences dealing with 'Dogme' have created heated discussions in the field of EFL. There are many teachers that have embraced Thornbury's ideas and others that see 'Dogme' as, at least, a risky approach especially for new, less-experienced teachers (Gill, 2000).

All this has restarted the discussion of whether we should teach through methods or, as 'Dogme' recommends, we should do so through core issues that go beyond methods or tools. In fact, Meddings and Thornbury, after being asked if they considered 'Dogme' a method, wrote in an online article (2003) that 'We see it more as a mindset, a way of being in the classroom'.

1.2 Gestures in Language Teaching

Gullberg & De Bot (2010, p. 163) said that '... gestures that convey speech-related meaning should improve language learners' comprehension and possibly also learning of language.'. This suggests that using gestures that transmit information about meaning

should enhance the comprehension of the input and, thus, improve the language learning process. On the other hand, they also explained that research exploring the effect of gestures in the output was hard to find. They said that there was still no strong empirical proof that gestures in the output had long-term effects in the language acquisition process.

Another study by McCafferty (2002) identified gestures as an important strategy to create Zones of Proximal Development. In other words, gestures help the learner bridge the gap of what they know and they do not and, according to McCafferty, they do this both when receiving input and when producing output. The author expected gestures to 'play a facilitating role' in real communicative situations. Another thing that McCafferty explained was that, in his experiment, 'the first use of gestures came as an explicit pedagogical tool' and, after that, the student started using this tool to support both his production and his comprehension. In other words, when teachers use gestures with a pedagogical aim, students are more bound to use them in their learning process.

Gregersen (Olivares-Cuhat) went a step further and recommended that teachers allow learners, especially beginners, to use gestures while they communicate in the L2. They also said (p. 205) 'that all learners at varying levels of proficiency could enhance the verbal meaning of their messages by using more speech-related gestures to complement and accent their spoken words.'. Gregersen encouraged teachers to keep verbal and non-verbal communication together rather than separating them, especially those that separate the visual channel from the auditory one, like listenings versus videos. One last recommendation in their article was to use role-playing and drama activities due to the relation between verbal and non-verbal communication in these activities.

1.3 Characteristics of Whole Brain Teaching

This section will explore the major characteristics of Whole Brain Teaching (WBT henceforth). First, the main techniques will be presented and then the existing research on the method will be reviewed and some conclusions will be drawn about the relationship between WBT and language acquisition.

1.3.1 Whole Brain Teaching Techniques

WBT is a generalist method (oriented to classroom teachers dealing with various school subjects— math, science, etc.) that was first developed in 1999. It offers a set of tools and techniques to improve classroom management and some other tools to improve the performance in instrumental areas such as mental arithmetics or language (from a mother tongue perspective). WBT is not an EFL method, but it has certain techniques and characteristics that respect some of the principles of language acquisition that were mentioned in the previous pages.

Below is a summary of some of these techniques and tools that are presented as the 'Big Six'[1] (Biffle, 2007, 2009a):

- *Class! Yes!* : is an attention grabbing routine. The teacher says *Class!* and the students reply *Yes!*, fall silent and focus their attention on the teacher. It is important to include some variety from time to time so the learners stay interested in this routine. The teacher will sometimes change the tone of voice, include endings to *Class!* or vary the speed or pitch of the word. Students have to answer in that same way. The author claims that this is a very effective way to get the attention of the whole group and that it saves time.

[1]Note that in the training seminars that the team of http://www.wholebrainteaching.org are now giving, the have added a seventh technique to the set of six: *Mirror*. This technique is presented at the end of this Section, when talking about other minor techniques used in the experiment.

- *Teach! OK!*: is a routine designed to maximize learning. This technique requires chunking the information and the explanations in shorter pieces. The teacher explains an activity or part of the content using gestures to illustrate the explanation (see Section1.2) while the students mime these gestures. Then, she says *Teach!*. All the students answer *OK!*, turn to their partners and tell each other what the teacher has just explained at the same time that they continue using the gestures. In this technique they recommend introducing some variety like some rhythmic clapping before saying *Teach!* that the students will have to repeat before saying *OK!*.

- *Switch*: this technique supplements the previous one and is intended to allow both learners in each couple to have time for speaking and time for listening to their partner. The students in the class would be divided in pairs. These would further be divided between number 1 and number 2. Before saying *Teach!*, the teacher says what number will start paraphrasing what she has just explained. All the students with that number start speaking and using the gestures while the others listen in silence but mirror the gestures. When the teacher says *Switch*, all the class says *Switch* and the students swap roles. The author says that this is a way to get all the students speaking for a certain amount of time and then listening for some other amount of time. Biffle (2007, p. 28) explains this by saying 'It ensured that the chronic talkers would do their share of listening and that chronic listeners would do their share of talking'.

- The *scoreboard*: according to Biffle (2007), it is the motivating tool inside the method. It uses little rewards and punishments in a playful way that makes the students want to earn *happy faces* and avoid *frowny faces*. Figure 1.1 is an example of a scoreboard. In this figure there is a header row with two 'smileys'

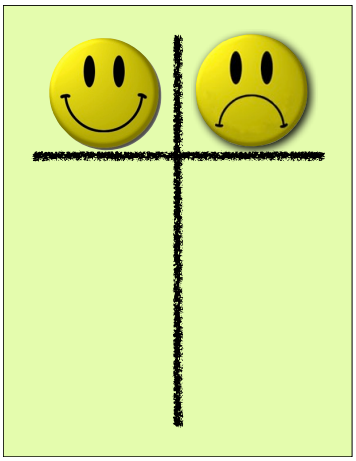


Fig. 1.1 An example of a scoreboard

(one happy and one sad) and a larger row underneath. This is used to note down the good or bad behavior (happy or sad face) of the students in the class.

Each time the teacher identifies a good attitude in one or more students, she might go to the scoreboard, get the attention of the class and, then, put a mark under the *happy face*. All the class will then celebrate the positive point with a *Mighty Oh Yeah* (this is, giving a clap and raising their fists in a celebrating way at the same time they say 'Oh, Yeah!'). If the attitude was something a student or a group of students could improve, the teacher will put the mark under the *frowny face*. This time, the students will celebrate the negative point with a *Mighty Groan* (this is, rubbing their eyes and pretending they are sad). In the first scenario, the teacher would probably mention who earned the class a positive point, but in the second, the teacher would never use names to avoid stigmatizing any student.

At the end of the class, the students will receive a prize if there are more positive points than negative or a punishment if it is the other way around. The size of the prize or the punishment will depend on the difference between one side and the other. The bigger the difference, the bigger the prize or punishment.

A possible prize or punishment the author recommends is in terms of minutes. They can leave 'x' minutes earlier or they must stay 'x' minutes longer depending on which side wins and with which difference. Biffle says that 'the smaller the reward you give, the more valuable it is' (2009a, p. 22). A very important rule inside the scoreboard is that the difference between one side and the other should never exceed three. If it did, the game would be too easy or too difficult and the learners would lose interest.

- *The 5 classroom rules* used in WBT are supposed to cover all the possible misbehaviors a teacher could find in a class. The 5 classroom rules are the following (Biffle, 2007):

  Rule 1 '*Follow directions quickly!*' is designed to keep the students active and on track and responding fast to the commands of the teacher.

  Rule 2 '*Raise your hand for permission to speak*' is used to prevent students from speaking out of time.

  Rule 3 '*Raise your hand for permission to leave your seat!*' is used to prevent students from standing up when they cannot.

  Rule 4 '*Make smart choices!*' is '*used [...] to cover every kind of disruptive behavior, in class and out.*'

  Rule 5 '*Keep your dear teacher happy!*' is designed to avoid arguments with disruptive students that might not agree with the decision of the teacher.

  Biffle (2007, p. 46) recommends rehearsing these rules often, especially at the beginning of the application of the method. Also, he explains that each of these rules can be reviewed any time the teacher detects someone having problems to follow them. The 5 rules are linked to the *Scoreboard* and are often used in conjunction with it, granting *happy* or *forwny* faces.

- *Hands and Eyes* is used when the teacher wants the full attention of the learners. When the teacher says *Hands and Eyes*, the learners repeat it and each of them puts their hands together and fixes their eyes on the teacher[2].

There are other techniques promoted by the authors of WBT that are not part of the 'Big Six'. One of them is the way to give commands to the students in a way they all do them faster and more efficiently. The teacher tells the students that, for instance, every time she says *Books!*, all the learners in the class will repeat 'Books' three times while they all take out their books. If the teacher says *Books, page forty-five*, the learners repeat 'Books, page forty-five' three times while they take out the books and open them on that specific page. The same could be used with instructions such as *Stand up*, *Sit down* or *Line up*. The authors claim that this technique works because even if some of the students were distracted when the teacher first gave the command, they will hear it when the rest of their classmates repeat it three times and start doing what the teacher said. Another small technique that was not part of the 'Big Six' is *Mirror*. When the teacher says *Mirror*, the learners say it once and start imitating the gestures the teacher is using. It is a command to emphasize and increase the use of gestures by the students. These techniques, though part of the 'Big Six', were used throughout the study, as will be explained further on.

1.3.2 Research on WBT

1.3.2.1 WBT in the press

WBT was developed very recently and, therefore, there is not much literature on the subject. In fact, most of it can only be found in blogs and internet sites without much scientific weight. We can also find some information about WBT in the press. One

[2]This technique was not implemented in the experiment because being only a three month experiment, the researcher thought that *Class! Yes!* would suffice.

example is a short article called 'Teachers learn ways to keep students' attention, but are brain claims valid?' (Higgins, 2012). In this article, the author interviewed Dan Willingham, a neuroscientist at the University of Virginia, and David Daniel, the editor of the 'Mind, Brain and Education' journal. Both experts agreed that the neuroscience claims stated by the authors of WBT are not based on actual research and sound more like marketing. The claims they were referring to can be found in the training manual written by Biffle (2009a, p. 48) and are the following:

1. Class!-Yes! activates the pre-frontal cortex, the reasoning center of the brain. Think of this area as a 'light switch' that must be turned on, repeatedly by Class-Yes, for the rest of the brain to process information.
2. Five classroom rules because they engage seeing, hearing, saying, doing and the limbic system engage the pre-frontal cortex, Broca's area, Wernicke's area, the limbic system, hippocampus, visual cortex and motor cortex [sic].
3. Teach-Okay is the most powerful of Power Teaching's[3] learning activities. Students have their pre-frontal cortex engaged, activate Broca's area as they listen, Wernike's area as they speak, the visual and the motor cortex as they see and make gestures. This whole brain activity powerfully stimulates the hippocampus to form long term memories.
4. The Scoreboard keys directly into the limbic system's emotions and the amygdala which registers pleasure (Mighty Oh Yeah) and pain (Mighty Groan!) as students accumulate rewards and penalties. [...]
5. Switch! helps students fully develop both their listening (Broca's area) and their speaking (Wernicke's area) abilities.

[3]Power Teaching was the original name of the method. In year 2009 it was changed into WBT

The article finishes with an interview with David Brobeck, an assistant professor of graduate education at Walsh University in Ohio. He explains their plan to launch a project for training students on how to use WBT and do action research about it. This may explain the hows and whys of the results of WBT.

On July 1, 2015, PBS Newshour broadcasted a video where they also analyzed the method from a neuroscience perspective[4]. The video showed a WBT class, an interview with the teacher and an interview with Daphna Sohamy, a neuroscientist from Columbia University. In the interview with the teacher, she claimed that WBT is targeting to activate certain parts of the brain, like the pre-frontal cortex or the motor cortex. On the other hand, when Dr. Sohamy was asked about WBT, she explained that 'the brain learns when things are surprising and interesting'. She claimed that the WBT techniques can be surprising and interesting at the beginning, but the learners will end up getting used to them and will gradually lose interest. This is explained, she said, by the dopamine neurons only being triggered when something new or unexpected happens. Dr. Sohamy does agree that we learn better when there is an 'emotional or social significance'. The video ends by pointing out three facts that neuroscientists claim: stress can damage neurons, benefits of brain-training games have not been proven and physical exercise is good for the body and the brain. Some days after the video was broadcasted, the creators of WBT asked all their followers in Facebook and Twitter to go into the PBS network and answer back to their criticism. Figure 1.2 shows this request.

1.3.2.2 WBT scientific studies

As explained in the previous section, WBT is a very recent method and it has not yet been researched in depth. Also, the validity of most of the research related to WBT

[4]http://www.pbs.org/newshour/bb/teachers-tap-brain-science-boost-learning/ (Last visited: April 2016)

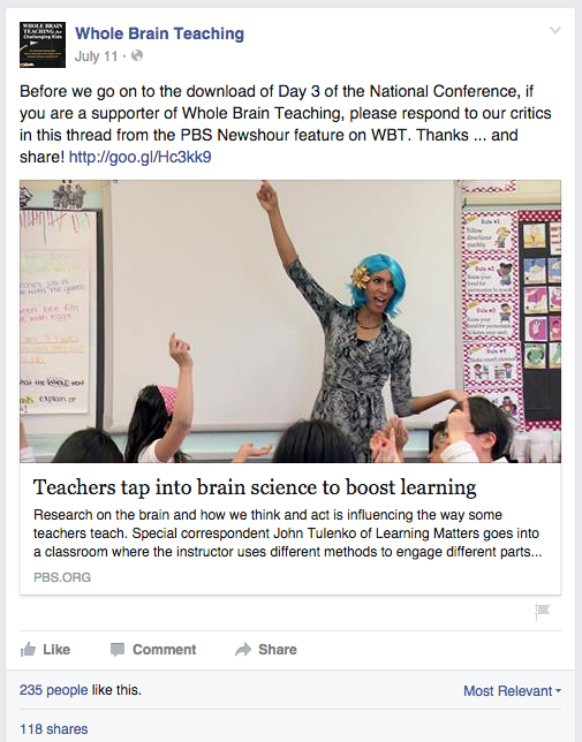


Fig. 1.2 WBT Facebook page asking supporters to respond to the critics of the method

found at the time of writing this dissertation was very limited. Only one online journal article related to WBT was found (Wirani, Setiyadi). This study intended to explore how to implement WBT in class, check if it increased the learners' participation in speaking activities and explore the feelings of the students when using WBT. The authors of this article claim that *Power Teaching*[5] increased the participation of the students in speaking activities and say that most of the thirty-three students valued WBT positively.

The rest of the research studies were undergraduate or master's thesis, so their validity has not been proved by the review process of professional journals or PhD dissertations. Therefore, the claims that we can find in these studies are not trustworthy enough to be of much weight in this research. Also, the online journal article lacks

[5]WBT changed the name from *Power Teaching* to *Whole Brain Teaching* in the summer of 2009, 5 years before the publication of the article

rigour as can be seen by the fact that it is based on 4 techniques of *Power Teaching* that do not correspond to reality. They talk about 'controlling, exploring, pursuing, and preserving', terms that are not related to WBT (or *Power Teaching*). Therefore, their claim that the speaking participation increased after using WBT, cannot be taken into consideration.

Of the other research, the following studies could be highlighted:

- Azusa Pacific University in California did a thorough case study on the impact of WBT at San Jacinto Elementary (Armijo, 2009). This case study included a quantitative part that compared the results of the 4 years previous to the implementation of WBT with the year when it was first implemented. It also contained a qualitative part where the author explored the opinions of both teachers and students involved in WBT.

  In the quantitative part the author claimed that there was an 'increase in the Academic Performance Index (API henceforth) score over the past five years' (Armijo, 2009, p. 76). This was true, but it is difficult to say that it was only due to the application of WBT, especially because between the years 2006-2007 and 2007-2008, the increase in the API score was almost as high as in the year of the introduction of WBT (43 points against 46 points). This might indicate that the district was already improving and not that the implementation of WBT made an impact on things. The lack of an experimental design makes it difficult to know if WBT was the variable that determined the improvement in the API scores. The quantitative study also gave some information about the opinion of both teachers and students. Although the sample was a small one (10 teachers and 36 students), there seemed to be strong positive opinions about the use of WBT, both by the teachers and by the students. Teachers saw WBT as a positive method and valued the use of gestures in the teaching-learning process. Students

enjoyed WBT as a method, they found that they remembered more information taught in class and they identified *Teach! OK!* as a positive strategy.

The quantitative data collected by the researcher was used to confirm the claims extracted by the qualitative part of the study. Armijo (2009, p. 79), explained that 'Whole Brain Teaching strategies have a positive impact on student achievement'. She also explained that 'The teachers and students agree Whole Brain Teaching has a positive impact on student engagement' (Armijo, 2009, p. 80).

- Lockhart (2009) studied how WBT affected a group of primary students in three subjects: mathematics, Catalan language and science. The results showed an improvement in the performance of the students, especially in the language-related subjects. This research also included a quantitative part and a qualitative one. In the latter, the researcher interviewed the teacher, who highlighted how the behavior, performance and attitude of the students all had improved since the application of the method. Nevertheless, the design of the research was not robust enough (not an experimental design and comparing different content), which suggested the need for further research.

- Szott & Molitoris (2010) analyzed WBT from the teacher training perspective. It is an inquiry based research related to their internship at Pennsylvania State University Professional Development School. Their main question is 'How can Whole Brain Teaching (WBT) impact my classroom environment?' (p. 5) but they also wanted to know how WBT affected their teacher presence, the 'on task versus off task behavior' (that is, students focused on the task or distracted with other things) and student participation. As in the other pieces of research, they stated that WBT seemed to improve the performance and implication of the students.

- The master's thesis by De Jager (2012) focused on teacher development and mentoring to improve the classroom management capabilities of new teachers. The study was an action research that combined a quantitative analysis with a qualitative one. It involved five beginner teachers in a peer mentoring group in South Africa. The action research was divided into two spirals as described by Carr & Kemmis (1986) (cited by De Jager (2012)): '[...] the project proceeds through a spiral of cycles of planning, acting, observing and reflecting, with each of these activities being systematically and self-critically implemented and interrelated [...]'

  In the first spiral, the researcher conducted a study in her own teaching practice and then mentored the five teachers. In the second spiral, the five teachers started teaching by implementing the techniques they had been mentored with. All the teachers involved in the study except one liked WBT. Paradoxically, the group of students with more diverse feelings about the method was the one this teacher taught. However, De Jarger claims that the majority of the learners, even in that specific group, liked the method.

1.3.2.3 Other published materials related to WBT

Another type of published material related to WBT is the one that can be found inside the authors' webpage. There is a wide variety of ebooks with instructions about how to apply the method and/or some of the tools. We can also find some videos that serve as examples of the application of the method or the different tools.

Unfortunately, these ebooks and videos do not have any scientific weight. They lack a bibliography and references supporting their claims. They do not state where the ideas come from. They do not provide any empirical data to back their theory. In fact, in a personal communication with Chris Biffle, one of the creators of WBT, he

explained that 'the method was created after being 'involved for years in trial and error teaching', rather than creating the method starting from other pedagogical theories' (Lockhart, 2009, p. 10).

1.3.3 WBT and its relation to language acquisition

As mentioned in Section 1.1, the EFL field is in a post-method era. It is not usual to research methods nowadays, but in this case the intention is not to test WBT as an isolated method, but to integrate it with other methods and techniques in a wider approach to language acquisition. The aim is to check if the process of acquisition can be improved by applying the WBT routines and some of its techniques. Obviously, every method has positive and negative points and WBT is no different. Consequently, some of its tools and techniques will be used but others that are not adequate to the characteristics of the learners and teacher in the experiment will be left out.

The SLA and EFL literature recommend that teachers keep their students motivated at the same time as they provide them with huge quantities of comprehensible input. TPR talks about the importance of having a 'stress-free environment'. NLP recommends improving the rapport inside the group by, for example, using different perceptive channels (including the kinesthetic one). The most extended approaches (e.g. The Communicative Approach) and the most up-to-date trends (e.g. Dogme) recommend that teachers not only allow students to speak and communicate, but they also facilitate this.

WBT seems to do all this. One of the things that appeared in the research related to WBT is that teachers and students seemed more motivated. The use of fun and variety (variations in *Class! Yes!* or *Teach! OK!* or the *Mighty Oh Yeah* or the *Mighty Groan*), combined with more structured techniques like the *Scoreboard*, seem to target and increase the motivation of the learners. This seems to happen because students

are more active all the time. Biffle (2009b) says the students should have more fun following the rules than breaking them, and the same could apply to learning the content. The students should be able to enjoy themselves more when trying to learn the content of the lesson than when not paying attention.

Breaking up the content in smaller chunks and delivering it with gestures seems to make input more comprehensible. Besides, there is repetition of this input through *Teach! OK!*. Input becomes more comprehensible by the possibility of negotiating the meaning of what was not understood.

Students work in pairs. This enables the learners to be more relaxed because their productions are going to be heard only by their partner and not by the whole class. With this type of grouping, they can make mistakes without feeling embarrassed about them and, at the same time, they can receive nonintrusive corrective feedback through the negotiation of meaning with their colleague and with the support of a teacher that has more time to go around the class monitoring the productions of her students.

Techniques such as *Class! Yes!*, *Teach! OK!* or The Scoreboard seem to target the creation of a good group rapport. All the students perform certain actions at the same time (i.e. say *Yes!*, do the same gesture as the teacher...), which can help the teacher gain rapport with the group fast (Churches & Terry, 2007). In fact, the combination of the students hearing the teacher's explanation, seeing her gestures and doing the gestures themselves at the same time (and afterwards), is an example of involving all three main perceptive channels, which is one of the basic steps for creating a good rapport with big groups.

With *Teach! OK!* and *Switch* students are given the chance to interact. The explanations of the teacher are short, so students spend a lot of class time talking to each other. Furthermore, as mentioned above, they have the chance to negotiate

the meaning and to understand all those things they might have missed with the explanation of the teacher.

1.4 WBT and the twelve principles of language learning by Brown

Brown (2002) recommended an approach to language teaching based on twelve principles rather than on any single method (see page 1). This section explores the links between these principles and the WBT method.

1. *Automaticity:* In WBT there is a lot of repetition through the *Teach! OK!* technique. This constant repetition provides the chance to have some meaningful controlled practice in pairs that should ease the way towards automaticity.

2. *Meaningful learning:* Although this will depend more on the content that the students are learning rather than on the method, the more fun the students have while learning, the more chances there are that the learning will be meaningful and memorable. One of the objectives of WBT, as mentioned in Section 1.3.3, is that the students have fun while they learn.

3. *Anticipation of reward:* The Scoreboard and the celebrations with the *Mighty Oh Yeah* or the *Mighty Groan* are examples on how WBT uses rewards.

4. *Intrinsic motivation:* This factor is not only mentioned by Brown, Krashen (1988) considers it one of the essential factors for language acquisition. He calls it a *relatively stable factor* and, thus, it takes a long time to change. It is still to be proven whether WBT addresses this issue.

5. *Strategic investment:* Learners (especially kids) will be more ready to make a *strategic investment* if they feel they are active in the class and they have fun.

Kids do not usually have the conscience of how important English (or foreign languages) is in life, so their *strategic investment* should be influenced by external factors. Some of these external factors in WBT could be techniques like *Teach! - OK!* or the use of gestures.

6. *Language ego:* Although this factor is more widely spread amongst adults, children also feel weird sometimes when using a foreign language. Giving them the chance to use the language in a smaller setting can help them overcome the problems associated with the *language ego*. Thus, *Teach! OK!* is a technique that could help scaffold those students with a bigger *language ego*.

7. *Self-confidence:* This is another factor also mentioned by Krashen (1988). This factor is based on the previous learning experiences of each individual. It is also a *relatively stable factor* that is difficult to change. The use of the *Scoreboard* could help improve the self-confidence of the learners, especially if the teacher respects the principles mentioned on page 12: praising individual students who do things right but avoiding pointing at those not doing things as they should.

8. *Risk taking:* Learners have to face reasonable challenges in the class. These challenges can be in the form of input or in the form of output. The use of gestures in WBT should make the challenge of receiving input somehow more achievable. Also, *Teach! OK!* should make the challenge of producing less threatening.

9. *Language-culture connection:* This principle mainly relates to the content being dealt with. This is not specifically catered for through WBT.

10. *Native language effect:* This is another principle that is not explicitly exploited through WBT.

11. *Interlanguage:* Being able to hear the errors that the students are making is a very valuable source of information for teachers. It allows them to know where their students stand in terms of language acquisition and what language features they must reinforce. In big classes, gathering this information can be difficult and slow. The use of *Teach! OK!* gives teachers the chance to move around the classroom monitoring the productions of the students and, thus, gathering very valuable information related to their interlanguage.

12. *Communicative competence: Teach! OK!* is key again regarding this principle. This technique respects lots of the principles of a communicative approach, thus, enhancing the communicative competence of the students:

    - *Small group work:* In this technique students are already working in pairs.
    - *Focus on language use:* The learners are supposed to be focusing on transmitting the message the teacher said rather than focusing on the formal aspects of the language.
    - *Focus on fluency:* The teacher walks around the class checking for mistakes that harm the communication, but should not be correcting other errors or mistakes. Also, the corrections will usually be done after the production stage and in the whole group, so the correction should not be harming the self-confidence of the student that made the mistake.

1.5 Research Hypotheses

All the previously mentioned literature related to English language teaching, gestures and WBT seem to point at the fact that the tools and techniques of WBT can be a good addition to the foreign language class. The classroom management techniques should make class time more efficient. The gestures should make the input more

comprehensible and, thus, the language acquisition more accessible. *Teach! OK!* should allow the learners to access the language faster, helping them transfer it into their active language better.

Also, none of the twelve principles recommended by Brown (2002) are negatively affected by WBT, as explored on Section 1.4. On the contrary, nine out of those twelve principles should see a benefit when using WBT, at least theoretically. One of the most prominent ones, the 'Comunicative Competence', is the main objective of most English language learners: to be able to communicate through the English language.

Another factor related to language acquisition that could potentially improve through the application of WBT would be the motivation of the students. Both Krashen (1988) and Brown (2002) insist on the importance of this factor in the language acquisition process. Several studies mentioned in Section 1.3.2.2 point towards an improvement of the motivation of the learners when using WBT. Therefore, this research was designed to explore two main hypotheses:

1. The use of the WBT method can enhance the process of language acquisition, allowing the learners to improve their language skills more significantly with the same amount of exposure and, thus, to have better results in listening and reading comprehension and written and oral production tests.

2. The use of the WBT method can improve the motivation of the learners. This hypothesized improvement could lead to a lowering of the affective filter, which is considered a major factor in language acquisition.

UNIVERSITAT ROVIRA I VIRGILI
ENGLISH AS A FOREIGN LANGUAGE THROUGH WHOLE BRAIN TEACHING
IN PRIMARY SCHOOL
Edward Alvar Lockhart Domeño

Chapter 2

Method

This chapter presents the design of the study, where it was applied, the materials and tools used as well as the description of the statistical analyses chosen to check the hypotheses that were presented in Section 1.5.

2.1 Design

The design for testing the hypotheses is mixed, with a quantitative part and a qualitative part. The design is also experimental with two groups (i.e., experimental and control) where the distribution of the learners amongst the groups had been prearranged randomly by the school. Both groups received the same content, but the experimental group received it through the WBT method and techniques, while the control group received it using the same techniques the teacher had been using until before the experiment. The choice of which group was the control one and which was the experimental one was also done randomly.

2.1.1 Quantitative research

The quantitative part of this study was designed to test the validity of the hypotheses mentioned previously. General language tests, speaking tests and motivation tests were

used to test them. Table 2.1 shows a visual representation of when each quantiative test was administered to the students. These tests will be described in more detail in the following subsections.

Table 2.1 Quantitative tests and their time of administration

	Pre-test	Mid-test 1	Mid-test 2	Post-test
Language test	x	x	x	x
Speaking test (small group)	x			x
Motivation test	x			x

2.1.1.1 General language tests

The general language tests in this study were designed by the publisher of the textbooks used in the school, Big Surprise 4 by Mohamed (2012). The purpose of using these tests was to check the first hypothesis in this dissertation: that the use of the WBT method can enhance the process of language acquisition, allowing the learners to improve their language skills more significantly with the same amount of exposure and, thus, having better results in listening and reading comprehension and written and oral production tests. These tests had two differentiated parts that will be contemplated as independent tests: a part assessing the listening comprehension and a part assessing the reading comprehension and the writing. These tests were run at four different times throughout the experiment (as seen in Table 2.1), and lasted an hour each.

- *Pre-test*: before the experiment began. It tested content delivered during the first term of the school year.
- *Mid-test 1*: at the end of the first unit taught during the experiment. It contained activities assessing the content of this unit.
- *Mid-test 2*: at the end of the second unit taught during the experiment. It contained activities assessing the content of this unit.

- *Post-test*: the last week of the experiment. It tested content delivered all throughout the experiment (during the second term of the school year).

The teacher administered and assessed all the language tests, both in the control and in the experimental group. After each test was administered, the learners could see the corrections. This fact did not bias the experiment because each of the tests was different. The results obtained from these exams will be described in Section 3.1.1. Each of these exams were assessed with marks. The higher the amount of marks, the higher the grade. As all the tests had a different number of marks the learners could get, the maximum possible grade varied from one test to another. These grades were transformed into base-100 so they could be compared.

2.1.1.2 Speaking test

Another quantitative measurement used in this study was an oral production test. The objective of this test was to assess the oral production skills, which had not been covered in the general language tests. Also, this test could shed some light on the differences between using or not using the method in terms of oral fluency, pronunciation and vocabulary used. This test, as seen in Table 2.1, was administered both at the beginning and at the end of the experiment. The difference with the language and motivation tests is that this one was not administered to all the learners in the experiment. Due to the difficulty of doing this kind of test with big groups, only six students in each class took the test. To have a fair representation of all the levels of English command inside the groups, the teacher divided all the students in each group into three levels of English command: good, medium and poor. Two students were randomly chosen from each of these sub-groups. The test consisted on the students describing an image[1] (see Figure 2.1) while they were being taped. The learners did

[1]Image found at http://www.dreamstime.com/stock-photography-kids-park-image2581032 and used with permission of the author

not receive any feedback from this test because they would have to take the same test again at the end of the experiment.

Fig. 2.1 Image used for the speaking test

This test intended to measure the difference in fluency and pronunciation shown between the pre-test and the post-test. The fluency would be analyzed by the length of the sentences, the length of the pauses between words and between sentences and the amount of different words each student used. The results obtained in the pre-test would then be compared to the ones obtained in the post-test to see the evolution. The first two analyses related to the fluency (i.e., length of sentences and length of pauses) could not be performed because the students' command of the language was not good enough to create whole sentences. The students would only point at the image and say isolated words or an adjective with a noun. Thus, only the evolution in the vocabulary and the pronunciation could be analyzed.

The test took between four and five minutes per learner. The recordings were done individually (one student with the researcher) in an adjacent classroom. The researcher would summon a learner, explain how the test worked and would start the recording

in the Voice Memos app of an iPhone using a microphone to improve the quality of the recording. The learner would then proceed to explain everything she saw in the image. If the student got blocked, the researcher would offer some support by pointing at something in the picture and asking the learner if she recognized what it was. If the learner still did not know the word, the researcher would offer three options from which the student would have to choose the correct one. The words that the learners could use autonomously were considered to be in their active vocabulary while the words that they were only able to use with the help of the researcher were considered passive vocabulary. If the learners were not able to choose the correct word out of the three options the researcher offered, it meant the word was still not acquired by the learner. The researcher would not transmit whether the learner's choice was right or wrong so the test could be used again in the post-test. The results obtained from these speaking tests will be explained in more detail in Section 3.1.2.

2.1.1.3 Motivation test

The test used for measuring the motivation of the learners in this experiment is an adaptation of the mini-AMTB[2] motivation test that was used by Bernaus (Wilson). It was applied to test the second hypothesis in this dissertation. The learners took the mini-AMTB for the first time right before starting the experiment. Just as in the speaking test, the learners did not receive any feedback from this test with the purpose of being able to use it again at the end. In the last session of the experiment the learners took the mini-AMTB again (see Table 2.1). The objective was to test the variations that they had experience throughout the experiment. The results from this test will be presented in Section 3.1.3.

[2]The mini-AMTB is a brief form of the Attitude Motivation Test Battery, which is widely used in the field of EFL.

2.1.2 Qualitative research

The rationale behind the qualitative research in this study is to have a clearer understanding of the results from the quantitative analysis and to be able to obtain information from the experiment that does not appear in an exclusively quantitative research. In order to accomplish this, there were two semi-structured interviews with the teacher. The first one was conducted one month and a half into the experiment, once the teacher had had a proper first contact with the method. The second one was done at the end of the experiment after three months of use of WBT. This intended to show differences after time. The interviews were semi-structured, to cover all the important factors directly related to the research (improvement or not of the language acquisition and increase or not in the motivation of the students) at the same time that they allowed the researcher to explore responses for a better insight on the quantitative results. The interviews lasted thirty-eight minutes and twenty seconds the first one and thiry-three minutes and thirty-two seconds the second one. These interviews will be further explored in Section 3.2.1.

Once the interviews were video recorded, they were analyzed to extract the most meaningful information. The first step to keep track of all the questions and the comments was to note down the most important items from the interview in a spreadsheet. The answers that drifted off topic or the questions that were not directly relevant to the topic were obviated. The spreadsheet was divided in 4 columns, each of them containing the time and the important question or comment. These 4 columns were the following:

- Question by the researcher.
- Positive comment from the teacher (e.g., 'They used to ask things in Catalan, they are now using more English').

- Negative comment from the teacher (e.g., 'The scoreboard is stressful for me. It forces me to pay a lot of attention to the time').

- Neutral comment from the teacher (e.g., 'I have realized that when you say a thing, the students have to do it. When I say class, you have to listen to me').

Once all the important information had been extracted from the videos, each of the comments given by the teacher was coded with a color. Each color represented a different category. Some of the things mentioned by the teacher could clearly fit into only one of those categories, but some others could fit into more than one. In those cases, they were attributed to two categories and assigned two colors.

Blue Comment related to the command of the language of the students.

Green Comment related to the motivation of the students.

Red Comment related to gestures.

Purple Comment related to specific WBT techniques.

Yellow Comment related to general classroom management.

Orange Comment related to the feelings of the teacher

2.1.3 Application of the method

The experimental group received the same content as the control group, but with WBT techniques for transmitting it and for managing the class. The content they received was what was included in their English textbooks. This can be seen in Table 2.2. Notice how both lessons have the same activities and the only difference is the use of the WBT techniques only seen on Table 2.2b.

Table 2.2 Lesson plan sample for each of the groups

(a) Lesson plan for the control group

SESSION 3

Activity	**Content**	**WBT Techniques**
Songs	Cowboy, Fifi and Food Chant	—
Story	The Hungry Cat	—
There is Some	Find the sentence *there is* in the story	—
AB Pages 22-24		—

(b) Lesson plan for the experimental group

SESSION 3

Activity	**Content**	**WBT Techniques**
Songs	Cowboy, Fifi and Food Chant	—
Story	The Hungry Cat	*Class! Yes!*
There is Some	Find the sentence *there is* in the story	*Teach! OK!*
AB Pages 22-24		*Teach! OK!*

Every class in both groups was videotaped. The researcher observed the videos of the experimental group to spot problems in the application of the WBT techniques. Every two weeks, the researcher, after observing the videos, would transmit feedback to the teacher so she could use the techniques more accurately. The videos were then stored to be used in case some questions arose throughout the research and there was a need to compare them between the two groups.

2.2 School

The research was done in the *Escola Joan Ardvol*. This is a two-section (two groups per grade) public school in Cambrils, a coast city in the province of Tarragona, in north-east Spain. It is a public school that depends on the Department of Education of the Catalan Government (*Departament d'Ensenyament de la Generalitat de Catalunya*). The school teaches the pre-primary stage (learners from three to five years old) and the primary stage (learners from six to eleven years old). Pre-primary education is divided in three grades (P3, P4 and P5), while primary education is divided in six

grades (from first grade to sixth grade). Primary education is also sub-divided in cycles: first-cycle, composed by first and second grade; second cycle, by third and fourth grade; and third cycle, by fifth and sixth grade. Although pre-primary school in Catalonia is nonobligatory education, most of the kids between those ages attend it. Primary education is an obligatory stage and English is one of its compulsory subjects. The groups in this school were formed by around twenty-five students per grade, so the school has a total of around four hundred and fifty students and around thirty teachers. The school is now more than fifty years old. It was first opened in 1960, becoming the first public school in Cambrils.

Escola Joan Ardèvol is a school that tries to innovate and find new and better ways for teaching. They transmit this in their webpage, in the project section (Escola Joan Ardèvol, 2016). One of the projects they had at the date of starting the experiment was devoted to the improvement of the teaching of English, but it was not related to WBT or to this research. This school was chosen due to the fact that it was a two-section school, and the students had been placed randomly in their group (without using performance or intelligence as a variable to distribute the groups). This way, the groups would already be randomized (as mentioned on page 29). Also, using one single primary school minimized the external variables because students all came from a similar sociocultural background, they were all the same age and most of them had had very similar previous English learning experiences.

2.3 Subjects

The most suitable group for the experiment in the school was the 4th grade (nine years old) because they would not have to pass the Basic Competencies Test[3] and because they did not have to go through the adaptation process to primary education of the

[3]Proves de Comptències Bàsiques: they are a whole battery of standarized tests that all students have to take at the end of primary education (sixth grade)

first cycle. This decision was agreed on with the school principal and the staff. The control and the experimental groups were chosen randomly so the experiment would not be biased. The characteristics of these 4th grade groups were the following:

- The control group had a total of twenty-five students (ten boys and fifteen girls). In this case, one of the girls did not attend classes regularly.
- The experimental group had a total of twenty-three students (twelve boys and eleven girls), although one of them left school early and one of them had attendance issues due to external factors and could not be used in the research.
- Both groups presented students from a variety of linguistic, geographic and cultural backgrounds.
- There were mixed socio-cultural backgrounds: students who were middle class and others that were lower class.
- Some of the students attended extra-curricular English classes (eleven in the experimental group and five in the control one).
- Most of the students had been learning English in school for five years.

The learners received three sessions of English per week. Of these, two were part of the experiment and involved the whole class. As mentioned in Section 2.1, the control group followed the content in the textbook without any use of WBT teaching techniques while the experimental group followed the same content but with some of the WBT teaching techniques (see Figure 2.2). For the third session, which was not part of the experiment, each group was divided in two and engaged in communicative activities with teachers that were not participating in the experiment. No WBT techniques were applied in these sessions in either of the groups.

2.4 Teacher

The teacher who participated in the experiment had more than twenty years of teaching experience. She was an English teaching specialist, but she was also certified to teach any primary school subject. She had two prior contacts with WBT. The first contact was a training course for in-service English teachers that was organized by the *Generalitat de Catalunya* and that was delivered by the researcher. It was based on storytelling, but techniques like *Class-Yes* or *Teach!-OK!* were used throughout the course.

The other contact with WBT was a tailor-made training course prior to the start of the experiment that was given in the school to her and to five other teachers interested in the method. It was a six session training course where the researcher transmitted the most important techniques within the method (the previously-mentioned Big Six) and taught the group of teachers how to apply it, both for English teaching and for other subjects. Those other five teachers started applying the method even before the beginning of the experiment, but did so with other groups in the school that were not going to be involved in the current study.

2.5 Materials

2.5.1 Textbook

The textbook used in both groups was *Big Surprise! 4* by Mohamed (2012). All the activities that the teacher used came from it. The teacher used an Interactive White Board (IWB henceforth) to present some of the information from the book. Sometimes the students had to use the IWB to correct some of the exercises in front of the whole group. The use of the IWB was the same in both the control and the experimental group.

2.5.2 Language Exams

The language exams used in the experiment were also designed by Oxford University Press. As mentioned on page 30, the pre-test exam was the one designed to assess the whole first term, the post-test exam was the one designed to assess the whole second term and the two mid-term exams were the ones for assessing the two units that the teacher covered throughout the experiment. The exams were part of resources inside the Big Surprise 4 teacher's book.

All four language exams were divided into two separate parts:

- Listening comprehension.
- Reading comprehension and writing.

For the pre-test, the reading and writing part had three questions. The first question had fifteen items the students had to answer (fifteen possible marks) between reading and matching (i.e., sentences and images) and writing. In the second question the learners had to write eight sentences related to some pictures (eight marks). And in the third question they had to complete a table about themselves and then write five sentences with that information (five marks). The reading and writing part had a total of twenty-eight marks. The listening exam also contained three questions. In the first one they had to listen to a recording related to some images, make a drawing related to the image and the recording and write the sentence that was in the recording (twelve marks). In the second question, they had to number some images, make a drawing related to those images and the listening and write what the recording said of each picture (nine marks). The third question consisted in listening to five sentences and circling the correct word out of two possible words in each of the sentences(five marks). The listening part had a total of twenty-six marks.

The mid-test 1 was the unit three exam. The reading and writing part had four questions. In the first question, the learners had to read some sentences, match them to a drawing and write a short sentence adding some extra information depending on what appeared in the drawing. They could get a total of five marks. In the second questions, the learners had to look at an image, read a sentence and write whether the sentence was true or false. If it was false, they had to rewrite it. They had to do this with four sentences and they could get up to four points. In the third question, the learners had to write a dialog between a mother and a son, based on some images that were marked as yes or no (four marks). In the fourth activity, the learners had to draw some food in certain places in the kitchen (i.e., jug, plate, fridge) and then they had to write a short paragraph describing where each food was (five marks). The listening exam consisted in one question with twelve marks. The learners had to listen to a recording, select the appropriate image amongst two and write the corresponding sentence.

The mid-test 2, was the unit four exam. The reading and writing part had four questions. The first question had four marks, although the teacher decided to give it five marks due to the fact that one of the questions only had a leading word and she understood it was not complete. In this question, the learners had to read a time from a clock, compare it with a drawing and write whether it was true or false and then write the correct (or the complete) time. The second question was a text with five gaps and some words that the learners had to put into the gaps (five marks). The third question consisted of five images depicting some actions. The learners had to write down each of those actions. The first sentence was written as an example (four marks). The fourth question involved writing a short letter about their day to a friend (five marks). The reading and writing part, after the modification of the teacher, had a total of nineteen marks. The listening exam contained one question in which the learners

had to listen to a recording and had to mark one of three different time options and finish a sentence depending of what the recording said. This activity and, thus, the listening part, had a total of twelve marks.

For the post-test, the exam contained questions related to the content worked on units four, five and six of the textbook. The reading and writing part had three questions. The first one (four marks) consisted in reading a statement and relating it with the corresponding picture (four marks). The second question consisted of four written statements and a grid with images where the students had to mark a tick or a cross depending on what the statements said (twelve marks). The last question was a fill-in the gaps activity where they had to write the information they could find in five pictures (five marks). The reading and writing was worth a total of twenty points. The listening test also had three questions. The first one consisted in listening to some numbered sentences and writing the corresponding number next to each of six pictures (six marks). The second question consisted in listening to some sentences and linking two images depending on what the sentences said (ten marks). In the last activity, the learners had to number objects in an image in the order that they appeared in the listening (four marks). The listening part had a total of twenty marks.

2.5.3 Speaking Test

In the speaking test described in Section 2.1.1.2, the students were presented with a picture that they had to describe. They had to use English to explain what they saw in the image while they were being recorded with an iPhone (the picture used can be seen on page 32). From the recordings of each student, the researcher isolated six words that were said both in the pre-test and in the post-test[4] and these words were evaluated by the judges. Each judge, thus, had to assess a total of one hundred

[4]one student in each group only had five words in common between the pre and the post-test, so in their case, there is a word in the pre-test and one in the post-test that do not match

and forty-four words. These six words per student were extracted using the computer application *Audacity*. They were then uploaded into a Google Form so five judges could analyze the 'native-like pronunciation' of each word through a 7-point Likert scale[5]. A screenshot of the form can be seen on Figure 2.2.

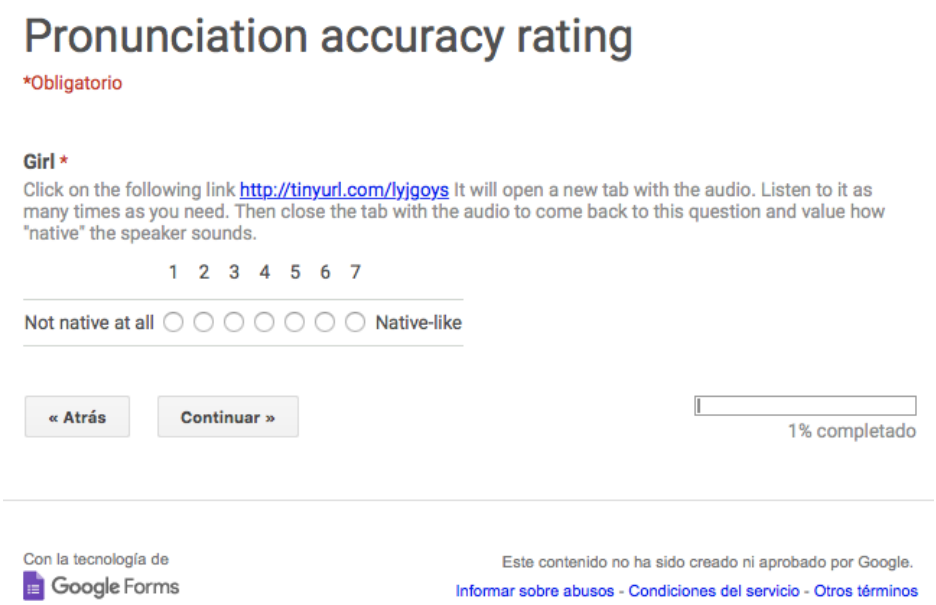


Fig. 2.2 Google form to allow the judges to access the recordings and measure them

The judges that had to use this form were chosen under certain requirements. They had to be native speakers of English, they had to have lived in Spain so they could all have a knowledge of the characteristics of the Spanish speakers of English and they had to have some kind of relation to teaching. The five judges selected for this experiment fulfilled all these characteristics, but they were from different parts of the world, which impeded the in-person completion of their evaluation. The Google Form presented above allowed them to do so from their homes. This might have affected the result of the judgements because of the lack of professional equipment to do the listening of the recordings. Each judge received an email with an explanation on how to proceed. In the email, which can be found in Appendix A, the judges were instructed to watch a

[5]The whole form can be found at http://goo.gl/forms/fkSA5nxWVe

tutorial that would explain their task. Then, they were encouraged to use a trial test to get used to the tool before moving on to the actual questionnaire.

2.5.4 mini-AMTB

The mini-AMTB test that was used to measure the motivation of the students was an adaptation of the one used by Bernaus (Wilson). The test is divided in two differentiated parts: an objective part where the learners have to answer questions about their background and a subjective part where the learners have to mark their opinions in several Likert scales regarding things like motivation, the teacher, how much their parents insist on the importance of English, etc.

Some of the questions in the objective part were modified to make the language more accessible to the children in the experiment (in Bernaus's study, the test was used with teenagers, not children). These changes appear in Table 2.3, where every change has been highlighted. Also, the original test and the adaptation, both in Catalan, can be found in Appendix B and C respectively.

Table 2.3 Adaptations to the objective questions of the mini-AMTB

Original	**Adapted**
How many languages, besides Catalan and Spanish, do you speak and understand **fluently**?	How many languages, besides Catalan or Spanish, do you speak and understand **well**?
Do either of your parents speak any **foreign** language?	Does your **father or mother** speak any language **other than Catalan or Spanish**?
How many **years** have you studied English **at school**?	How many **grades** have you studied English?
How many weeks in a foreign country?	**Have you ever been to a country where they only spoke English?**
How many years in an academy or language school?	**Have you ever attended extracurricular English classes or have you been to any English school?**

The instructions in the test were also changed. Instead of stating the purpose of the questionnaire, the adaptation offered an example of how the students should answer the test. This was designed to help them understand how a Likert scale worked because they had probably had little contact with them in the past. This change is illustrated in Table 2.4. The example tried to be one that they would be able to understand easily and would help them have a clear picture on how to answer a Likert scale.

Table 2.4 Adaptations to the instructions of the mini-AMTB

Original	**Adapted**
The objective of this questionnaire is to know what you think about some topics related to the acquisition of English. Following you have a series of statements followed by a scale that you will have to mark according to what you think.	Now you will answer some questions about English learning. You will have to put a mark at the place that better shows how you feel. For example, if you are a super fan of Barça and the sentence says: 'I like Barça:' `NOT A BIT __:__:__:__:__:_X_ A LOT` You will put a mark as close to 'A LOT' as possible, like in the example. On the other hand, if you don't like Barça at all, you will mark as close to 'NOT A BIT' as possible.

It was estimated that the Likert-scale questions would be beyond the comprehension ability of fourth graders, so they were rewritten to make them more accessible to them. Table 2.5 shows the differences between the original mini-AMTB and the adaptation. The changes made in these questions were in terms of more simplified structures (as in questions 1 or 4), less abstract ideas (as in questions 2 or 3) or with phrases closer to their reality (as in question 7).

Table 2.5 Adaptations to the Likert-scale questions of the mini-AMTB

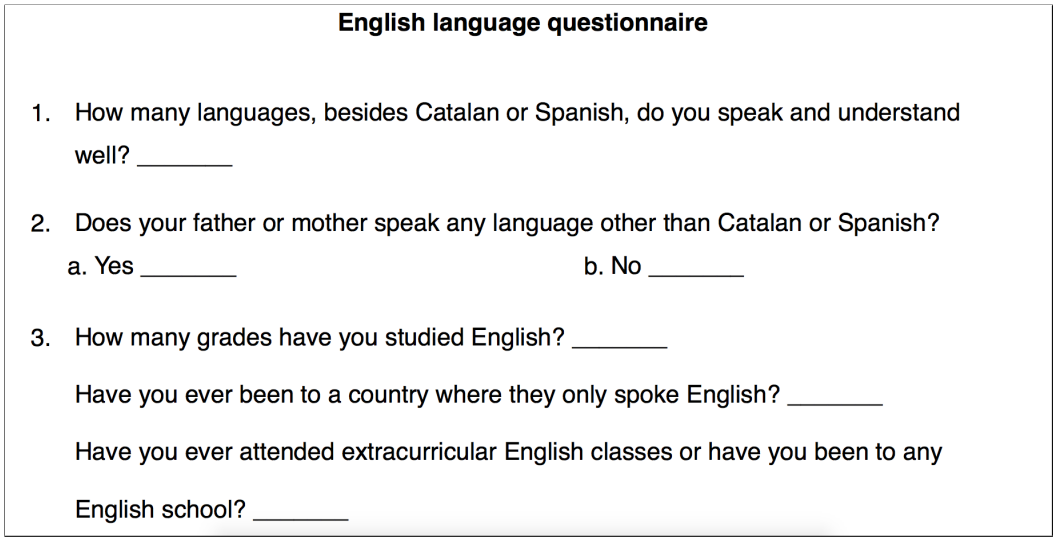
	Original	**Adapted**
1	My motivation for learning English for communicating with English speaking people is:	I want to learn English to be able to speak to people that only speak English:
2	My attitude toward the people that speak English is:	The people that only speak English are:
3	My interest toward foreign languages is:	I like foreign languages:
4	My desire to learn English is:	I want to learn English:
5	My attitude towards English learning is:	I like learning English:
6	My attitude toward my English teacher is:	I like learning English with my teacher:
7	My motivation for learning English for practical reasons (e.g., to get a good job) is:	I want to learn English to use it (to watch cartoons in English, for example):
8	I worry about speaking English outside of class:	I'm ashamed to speak English outside of class:
9	My attitude toward my English course is:	I like English classes:
10	I worry about speaking in my English class:	I'm ashamed to speak English in class:
11	My motivation to learn English is:	I enjoy learning English:
12	12. My parents really encourage me to learn English:	My parents tell me it's important to learn English:

2.5.4.1 Problems with the mini-AMTB

Before using the adaptation of the mini-AMTB, the test was administered to the two existing groups of third grade students in the same school. The objective was to validate the adaptation. In this validation, there were no problems and children seemed to understand the newly-formulated questions (i.e., the students did not ask for clarifications and the answers they gave seemed logical).

Understanding that the mini-AMTB worked fine, it was administered to both the control and experimental groups as a pre-test. In this case there were no issues either. The test was administered to the same students once more at the end of the experiment as a post-test.

When analyzing the data, some important contradictions where identified that led to question the validity of the tool. These contradictions were first spotted in the objective part of the questionnaire, where the students had to write an answer regarding their personal lives. The questions in this part can be seen on Figure 2.3.


English language questionnaire

1. How many languages, besides Catalan or Spanish, do you speak and understand well? _____

2. Does your father or mother speak any language other than Catalan or Spanish?
   a. Yes _____ b. No _____

3. How many grades have you studied English? _____

   Have you ever been to a country where they only spoke English? _____

   Have you ever attended extracurricular English classes or have you been to any English school? _____

Fig. 2.3 Objective questions in the adaptation of the mini-AMTB

Each test on its own (pre-test and post-test) had seemingly adequate answers to these questions, as had happened when piloting the adaptation. However, when seeing the evolution in the answers of the students, some important incongruences appeared. Some examples of these incongruences are:

- In the pre-test some students said they spoke and understood one language apart from Catalan or Spanish, while in the post-test they said they did not speak any other language.

- In the pre-test some students marked that their parents spoke other languages, but in the post-test they did not.
- In the pre-test some students had stated they had been studying English for seven years, but in the post-test they said they had been doing so for five years.
- Students that had been in English speaking countries before the pre-test, answered that they had not been in any English speaking country in the post-test.
- In the pre-test some students said that they had taken extracurricular English classes, but in the post-test they answered that they had not.

There were other surprising answers in the post-tests but were not considered incongruences because these things could have happened in reality. Some of these could be:

- In the pre-test some students said they spoke and understood one language apart from Catalan or Spanish, but in the post-test they said they spoke and understood two.
- In the pre-test some students marked that their parents did not speak any other language, but in the post-test they marked the opposite.
- Some students that had never been in English speaking countries before the pre-test, in the post-test they answered that they had.
- In the pre-test some students said that they had never taken extracurricular English classes, but in the post-test they replied they had.

As these incongruences were found, the adaptation of the mini-AMTB was piloted once again, this time in a second school with three separate groups of fourth grade students and also in a setting of pre-test–three months of teaching–post-test. In this

second school, the results were very similar to the ones obtained in the experimental school. These incongruences in the objective part (which will be further analyzed in Section 3.1.3.1) create serious doubts about the validity of the adaptation of the mini-AMTB as a measurement tool. Therefore, even though Section 3.1.3 will present the motivation results extracted from this questionnaire, these will not be definitive and will have to be supported with further research on the validity of the mini-AMTB or with other motivation measurement tools.

2.6 Statistical Analysis

Checking the distribution of the results was the first step in the analysis of the data obtained from the different tests run throughout the experiment (see Table 2.1 on page 30 for the full picture). The objective was to find out whether the distributions were normal, which would allow to use parametric tests like the ANOVA to show differences between the experimental and the control groups.

2.6.1 Normality Tests

The normal distribution of the different tests (general language tests, speaking tests and motivation tests) was tested to know if the data could be analyzed through parametric vs. non-parametric tests. Table 2.6 presents the skewness and Shapiro-Wilk data from the different language exams in the experiment.

Note that out of the skewness values for the control group shown in Table 2.6a, only the listening pre-test shows a value out of the normality range of -1/+1. On the other hand, the skewness values for the experimental group in Table 2.6b show a non-normal distribution in five different tests. Given those values and the small size of the smaples, the Shapiro-Wilk was ran to confirm the non-normal distribution. The results of this

Table 2.6 Normal distribution analysis in the different language exams

(a) Skewness for the cont. group

	Skewness
PretestList	-1.049
PretestRead	-.623
Midtest1List	-.670
Midtest1Read	-.012
Midtest2List	-.342
Midtest2Read	.475
PosttestList	-.675
PosttestRead	-.479

(b) Skewness for the exp. group

	Skewness
PretestList	-3.262
PretestRead	-1.533
Midtest1List	-1.629
Midtest1Read	-.814
Midtest2List	-.847
Midtest2Read	-.260
PosttestList	-1.053
PosttestRead	-1.847

(c) Shapiro-Wilk for the cont. group

	Shapiro-Wilk		
	Stat	df	Sig.
PretestList	.823	21	.002
PretestRead	.886	21	.019
Midtest1List	.827	21	.002
Midtest1Read	.950	21	.334
Midtest2List	.921	21	.092
Midtest2Read	.861	21	.007
PosttestList	.850	21	.004
PosttestRead	.926	21	.117

(d) Shapiro-Wilk for the exp. group

	Shapiro-Wilk		
	Stat	df	Sig.
PretestList	.683	18	.000
PretestRead	.731	18	.000
Midtest1List	.646	18	.000
Midtest1Read	.915	18	.105
Midtest2List	.874	18	.021
Midtest2Read	.882	18	.028
PosttestList	.822	18	.003
PosttestRead	.680	18	.000

analysis can be seen in Tables 2.6c and 2.6d. In this case, half of the tests show values below .05, which means that they are not normally distributed. The only ones that present normal distributions (Sig. $>$ .05) are the reading and writing mid-test 1 (for both groups), the listening mid-test 2 and the reading and writing post-test for the control group. The absence of normal distribution lead to the use of non-parametric tests, as will be explained below.

The statistical analysis performed to the results obtained in the speaking tests were also analyzed in terms of normality. In the case of the vocabulary, the skewness values

obtained in the different times and by the different groups, both in terms of passive and active vocabulary, can be observed in Table 2.7.

Table 2.7 Normality tests for the passive and active vocabulary found in the speaking pre-tests and post-tests

(a) Skewness for the passive and active vocabulary in the speaking pre-tests and post-tests

Vocabulary	Group	Test	Skewness
Passive	Control	Pre-test	-.513
		Post-test	.392
	Experimental	Pre-test	.523
		Post-test	.807
Active	Control	Pre-test	1.183
		Post-test	.043
	Experimental	Pre-test	1.942
		Post-test	1.311

(b) Shapiro-Wilk for the passive and active vocabulary in the speaking pre-tests and post-tests

			Shapiro-Wilk		
Vocabulary	Group	Test	Statitstic	df	Sig.
Passive	Control	Pre-test	.928	6	.566
		Post-test	.950	6	.737
	Experimental	Pre-test	.823	6	.093
		Post-test	.945	6	.700
Active	Control	Pre-test	.884	6	.287
		Post-test	.939	6	.694
	Experimental	Pre-test	.777	6	.036
		Post-test	.892	6	.331

Table 2.7a shows how both the control and the experimental group showed non-normal distributions in the active vocabulary. In the case of the control group, this non-normal distribution only happened in the pre-test, but in the case of the experimental group, it happened in both the pre-test and the post-test. This non-normal distribution was also checked through a Shapiro-Wilk (Table 2.7b), where most of the tests showed a normal distribution except for the active vocabulary for the pre-test of the experimental group, which showed a value below .05. This confirmed

the non-normal distribution observed in the skewness test. Even though it is only one result that is not normally distributed, this forces the use of non-parametric tests.

The results of the pronunciation part of the speaking test were also analyzed for normality. Table 2.8 presents the results of this normality tests both in terms of skewness and in terms of the Shapiro-Wilk test.

Table 2.8 Normality tests for the pronunciation in the pre-tests and post-tests

(a) Skewness for the pronunciation in the pre-tests and post-tests

Group	Test	Skewness
Control	Pre-test	-.669
	Post-test	1.959
Experimental	Pre-test	.265
	Post-test	-.130

(b) Shapiro-Wilk for the pronunciation in the pre-tests and post-tests

		Shapiro-Wilk		
Group	Test	Statitstic	df	Sig.
Control	Pre-test	.971	6	.902
	Post-test	.782	6	.040
Experimental	Pre-test	.870	6	.228
	Post-test	.996	6	.999

In this case, only the post-test in the control group showed a skewness value (Table 2.8a) that represented a non-normal distribution. The same happened in the case of the Shapiro-Wilk test (Table 2.8b). The other three values in both tests showed normality. Again, even though it was only one of the samples that had a non-normal distribution, this meant having to use non-parametric tests for the comparisons between the groups.

Analyses of normal distribution were also run in the motivation test. Table 2.9 shows the results in these tests regarding the motivation questions.

All the different motivation questions in these tests showed skewness values below -1, as can be seen in Tables 2.9a and 2.9b. This lack of normality was confirmed through

Table 2.9 Normality test for the motivation tests of both groups

(a) Skewness for the cont. group

	Skewness
Pre-test Mot 1	-2.560
Pre-test Mot 2	-2.455
Pre-test Mot 3	-1.374
Post-test Mot 1	-1.721
Post-test Mot 2	-1.966
Post-test Mot 3	-1.489

(b) Skewness for the exp. group

	Skewness
Pre-test Mot 1	-1.564
Pre-test Mot 2	-2.120
Pre-test Mot 3	-1.428
Post-test Mot 1	-1.598
Post-test Mot 2	-1.582
Post-test Mot 3	-1.412

(c) Shapiro-Wilk for the cont. group

	Shapiro-Wilk		
	Stat	df	Sig.
Pre-test Mot 1	.578	21	.000
Pre-test Mot 2	.637	21	.000
Pre-test Mot 3	.665	21	.000
Post-test Mot 1	.625	21	.000
Post-test Mot 2	.564	21	.000
Post-test Mot 3	.667	21	.000

(d) Shapiro-Wilk for the exp. group

	Shapiro-Wilk		
	Stat	df	Sig.
Pre-test Mot 1	.707	21	.000
Pre-test Mot 2	.630	21	.000
Pre-test Mot 3	.661	21	.000
Post-test Mot 1	.642	21	.000
Post-test Mot 2	.653	21	.000
Post-test Mot 3	.665	21	.000

the Shapiro-Wilk tests (Tables 2.9c and 2.9d), in which all the different questions for both groups and both for the pre-test and post-test were significant. In the case of the motivation test and given that all the questions presented similar values in terms of non-normal distribution, several attempts to normalize the data were made using Log10, LogN and Sqrt, but all of them were unsuccessful.

Table 2.10 shows the four attempts made over the results obtained from the experimental group in Question 3.

Table 2.10 Attempts to normalize the data from the motivation tests

		Shapiro-Wilk		
Exp Pre Motivation 3	Skewness	Statistic	df	Sig.
Original results	-1.428	.647	22	.000
RLog 10	1.104	.651	22	.000
RLog N	1.104	.651	22	.000
RSqrt	1.245	.652	22	.000

Not only did the Skewness continue to be outside of the 1/-1 range recognized as normal distribution, but the Shapiro-Wilk results were also significant in all four cases,

confirming the non-normal distribution. This, again, forced the use of non-parametric tests instead of their parametric counterparts like the t-test or the ANOVA. The specific tests used will be explained below, and the results obtained from them will be explored in Chapter 3.

2.6.2 Statistical Tests

The data obtained from the quantitative tests were analyzed statistically using a variety of techniques, including descriptive analysis, Mann-Whitney U tests and linear regressions. The descriptive statistics were used to give an overview of the results in each test and in each group. In case parametric tests could not be used, the non-parametric equivalent would be used instead, like the t-test and the Mann Whitney U. The latter was used to see if there were significative differences between the two groups in the experiment. To further explore the data, linear regressions were used as a way to compare groups and see their evolution in time. These last two tests (i.e., Mann Whitney U and linear regressions) were also used when exploring individual students, as will be dealt with in Section 3.1.2.2.1.

The data that will be taken into account in the descriptive statistics will be the number of learners (N), the average (Mean) and the standard deviation (Std. Deviation). The Mann-Whitney U will provide two sets of data: a rank table that will illustrate the ranks of each analyzed factor (Mean Rank) and another table stating the significance of that comparison (Asymp. Sig. (2-tailed)) between the two groups (i.e., experimental and control). In the linear regressions the β factor and the significance will be considered (β will refer to the variation over time and Sig. will determine if that time factor is significant). The results of all the previously mentioned statistics will be presented in detail in the following chapter.

The statistical analysis in this dissertation have been performed with three different applications:

Numbers: This MacOS program was used to perform a preliminary analysis of the quantitative data. The program was used to calculate means, standard deviations and percentages and to display all the information in graphic form in order to gain a visual understanding of the data. Some of these graphics have been used to illustrate some ideas that will be explored in Chapter 3.

SPSS: This was used for more complex calculations such as tests of normal distribution or others like the U Mann-Whitney or the linear regression. The tables extracted from these analysis were then retyped onto LaTeX. The graphs for the linear regressions were maintained in the same format that the SPSS created.

Google Spreadsheets: This application was used to gather the data from the judges for the oral production test. It was also used to calculate means and standard deviations of these data.

UNIVERSITAT ROVIRA I VIRGILI
ENGLISH AS A FOREIGN LANGUAGE THROUGH WHOLE BRAIN TEACHING
IN PRIMARY SCHOOL
Edward Alvar Lockhart Domeño

Chapter 3

Results

This chapter presents the results obtained from the different tests and analyses detailed in Chapter 2. The chapter is divided in two main sections: quantitative results and qualitative results. The former explores the outcomes of the different exams and tests. The latter analyzes the opinions and reactions shared by the teacher. In each of the subsections, the different tables and graphs analyzing the data are presented and then explained in detail. The first time a type of statistical study or a type of graphic is presented, its different elements are explained so it is better understood.

The results will be illustrated graphically using either linear regression graphics or bar charts. In the linear regression graphics, two important elements will be highlighted. On the one hand, a box in the middle of the regression line which includes a formula specifying the starting point of the regression line, a '+' symbol and the slope. The slope factor will determine whether the line shows a positive or negative incline. On the other hand, the R_2 value will also be included. This represents how far the different scatter points are from the actual regression line shown in the graphic. Low numbers in R_2 mean that the data does not correspond well to the visual representation in the graphic whereas numbers close to 1 mean a closer representation to the line shown[1].

[1]For a basic explanation of r-square, check https://en.wikipedia.org/wiki/Coefficient_of_determination

The bar graphs will be used to show a visual comparison of some factors such as the fluctuation of grades between judges in the speaking test or the variation between average evaluations and the standard deviations.

3.1 Quantitative results

3.1.1 General language tests

This section is going to present the quantitative results for the general language tests. It will be divided into several subsections that will explore the descriptive statistics and the other different statistical tests used to analyze the data. As mentioned in Section 2.6.1, the analysis of the normal distribution of the samples of the different language exams showed that some of these were not normally distributed. This forced the use of non-parametric tests.

3.1.1.1 Descriptive statistics

The results of the descriptive analysis can be found in Table 3.1, where Table 3.1a shows the descriptive statistics for the control group and Table 3.1b shows the ones for the experimental group.

One of the first things that can be observed in these statistics is how some of the learners missed some of the tests. In the case of the control group, Table 3.1a (N) shows how 23 learners took the pre-test, 22 mid-test 1 and 24 both mid-test 2 and the post-test. 21 students in this group took all the tests. Regarding the experimental group, Table 3.1b (N) shows that each test was taken by 22 learners (except the reading and writing pre-test, which was taken by 23). Only 18 learners in the experimental group took all the tests.

Table 3.1 Descriptive statistics of the language exams throughout the experiment

(a) Descriptives for the cont. group

	N	Mean	Std. Deviation
PretestList	23	85.652	16.696
PretestRead	23	74.652	24.737
Midtest1List	22	73.136	29.004
Midtest1Read	22	51.409	30.478
Midtest2List	24	57.916	30.375
Midtest2Read	24	39.166	31.736
PosttestList	24	73.333	28.155
PosttestRead	24	65.291	28.157
Valid N (list.)	21		

(b) Descriptives for the exp. group

	N	Mean	Std. Deviation
PretestList	22	89.045	22.491
PretestRead	23	76.913	34.646
Midtest1List	22	86.318	22.495
Midtest1Read	22	63.090	32.921
Midtest2List	22	73.909	23.139
Midtest2Read	22	50.545	32.529
PosttestList	22	84.772	12.861
PosttestRead	22	85.818	21.337
Valid N (list.)	18		

As the results of the tests had been transformed into base-100, the range in the grades could go from 0 to 100 points. These grades are represented by the average grade (Mean). The highest average grade corresponds to the experimental group, listening pre-test (89.045) while the lowest average grade corresponds to the control group, reading and writing mid-test 2 (39.166). In the control group, the listening pre-test and post-test (85.652 and 73.333 respectively) are higher than the mid-tests (73.136 for mid-test 1 and 57.916 for mid-test 2). The same happens with the reading and writing tests in the control group (74.652 for the pre-test and 65.291 for the post-test in contrast to 51.409 for mid-test 1 and the 39.166 for mid-test 2). In the case of the experimental group, the listening results show less fluctuation than the ones in the control group (ranging from 73.909 in mid-test 2 to 89.045 in the pre-test). Also, in the listening exams, contrary to what happened in the control group, mid-test 1 shows a higher average than the post-test (86.318 for the former and 85.818 for the latter). Regarding the reading and writing tests of the experimental group, the highest grades are in the pre-test and post-test (76.913 and 85.818 respectively), while mid-test 2

shows the lowest grade (50.545). This is the only case where the post-test has the highest grade in the four exams of its type.

Both groups show big standard deviations. In the control group it ranges from 16.696 to 31.736, while in the experimental group it ranges from 12.861 to 34.646. In the control group, the standard deviations in both pre-tests show lower values (16.696 for the listening test and 24.737 in the reading and writing) than any other test. In the mid-tests, the standard deviations increase over time both in the listening and in the reading and writing tests (they rise in the mid-test 2 up to 30.375 for the listening and 31.736 for the reading and writing) and then they slightly decrease in the post-tests (28.155 in the listening and 28.157 in the reading and writing). In the experimental group, in the first three listening tests, the standard deviation is quite constant (22.491, 22.495 and 23.139), but in the post-test, it goes down (to 12.861). Something similar happens in this group in the reading and writing tests. The first three exams have similar standard deviations (34.646, 32.921 and 32.529) and the post-test has a much lower one (21.337).

3.1.1.2 Mann-Whitney U test

The results of the Mann-Whitney U test performed on the data from the different language exams can be seen in Tables 3.2 and 3.3.

Table 3.2 Mann Whitney U test comparing the language exams of both groups

(a) Pre-test and Mid-test1

	PretestList	PretestRead	Midtest1List	Midtest1Read
Mann-Whitney U	169.000	192.000	166.500	186.000
Asymp. Sig. (2-tailed)	.070	.154	.060	.188

(b) Mid-test2 and Post-test

	Midtest2List	Midtest2Read	PosttestList	PosttestRead
Mann-Whitney U	181.000	212.000	232.000	139.500
Asymp. Sig. (2-tailed)	.066	.251	.473	.006

Table 3.2 shows only one statistically significant value, which can be found in the Asymptomatic Significance (2-tailed) of the Post-test, reading exam. In this case, it shows a value of .006, lower than the .05 which marks the significance level. The rest of the values are not significant, although some seem to show a tendency (the first three listening exams, which are close to the previously mentioned threshold).

Table 3.3 Rank table for the Mann-Whitney U test comparing the exams

	Group	N	Mean Rank
PretestList	Cont	23	19.35
	Exp	21	19.35
PretestRead	Cont	23	20.35
	Exp	22	25.77
Midtest1List	Cont	22	19.07
	Exp	22	25.93
Midtest1Read	Cont	22	19.95
	Exp	22	25.05
Midtest2List	Cont	24	20.04
	Exp	22	27.27
Midtest2Read	Cont	24	21.33
	Exp	22	25.83
Post-testList	Cont	24	22.17
	Exp	22	24.95
Post-testRead	Cont	24	18.31
	Exp	22	29.16

This information is complemented by the one found in Table 3.3 where it can be observed that all the mean ranks in the experimental group are higher than the control group, except for the listening pre-test, where both ranks mark 19.35. These ranks, combined with the significance values determine the significance tendency (whether it is the control group or the experimental one that performed it significantly better: the higher the rank, the better performance).

3.1.1.3 Linear Regression

Regression Lines were studied to further explore the evolution of the grades of each group. The results obtained will be further explained below and can be seen in Table 3.4 and Figure 3.1.

Table 3.4 Coefficient Time for the different exams and the different groups

(a) Listening, control group

	Unstandardized Coefficients		
Model	β	Std. Error	Sig.
(Const.)	85.409	7.045	.000
Time	-5.166	2.548	.046

(b) Listening, experimental group

	Unstandardized Coefficients		
Model	β	Std. Error	Sig.
(Const.)	89.818	5.502	.000
Time	-2.523	2.009	.213

(c) Reading and writing, control group

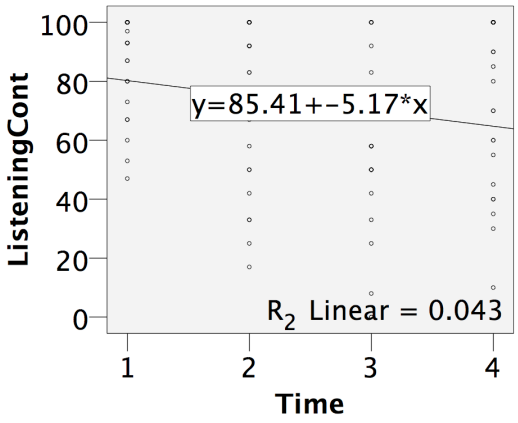
	Unstandardized Coefficients		
Model	β	Std. Error	Sig.
(Const.)	67.596	8.016	.000
Time	-3.963	2.899	.175

(d) Reading and writing, experimental group

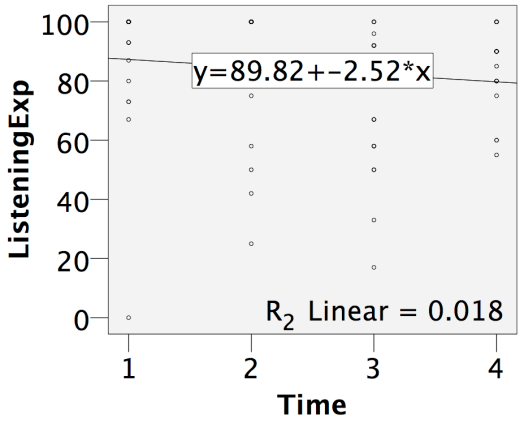
	Unstandardized Coefficients		
Model	β	Std. Error	Sig.
(Const.)	65.988	8.576	.000
Time	1.286	3.147	.684

The time coefficients in Table 3.4 show the possible variation that the students in each group would have in every time unit (understood as the time between each exam). The β factor under Unstandardized Coefficients shows the variation that is expected in a certain group and exam depending on the Time factor. The Standard Error shown besides it explains the margin of error that SPSS has calculated for that variation. The significance value (Sig.) determines if the time made a significant difference in the variation that happened in that specific group and exam. Something to highlight in Table 3.4a is that there is a significant value of .046. A significance below 0.05 in this type of test means that certain factor is key in the variation. If there is no significance (like in the other three subtables), this means that the variation might be explained by other factors not observed in the table.

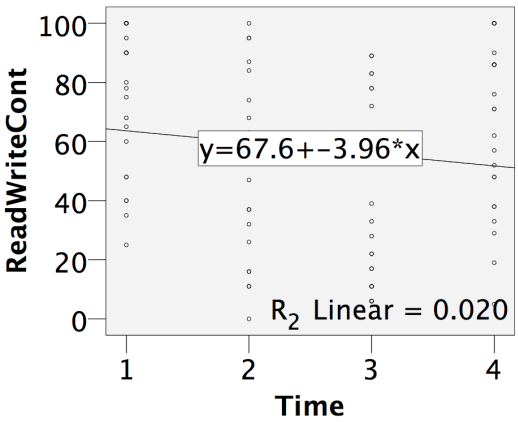
These results are visually represented in Figure 3.1. In this figure, the graphs on the left represent the control group, while the ones on the right represent the experimental group. Also, the top graphs represent the listening exams and the bottom ones represent the reading and writing.



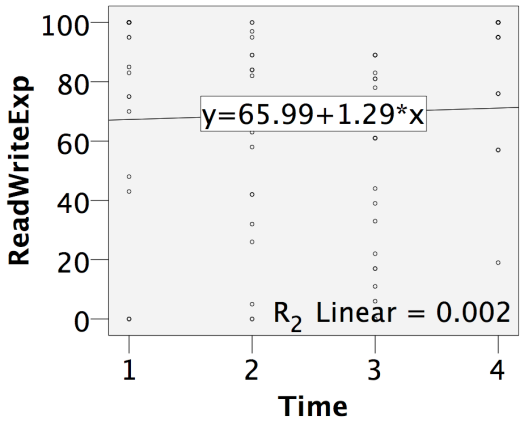
(a) Listening, control



(b) Listening experimental



(c) Reading & writing, control



(d) Reading & writing, experimental

Fig. 3.1 Regression lines of the different exams and the different groups

The line printed over the scatter plot projects the possible evolution of each group in each exam. Where the line starts on the left of the graphic will depend on the first number in the box over the line (for the listening exams in the control group—

Figure 3.1a— this would be 85.41)[2]. The positive or negative slope of the line is determined by the number following the + symbol[3] in that same box (in the previous example, it was -5.17). If it is a positive number, the line will go upwards whereas if it is a negative number, the line will go downwards. The R_2 value in these graphs will transmit how far the different scattered points are from the actual line represented on the graphic: the closer the R_2 value is to 1.000, the closer the scatter dots will be to the line. The only group that shows a positive regression line (the line goes upwards and, so, it has a positive value of +1.29 in the box, which corresponds to the β Time factor in Table 3.4d), is the experimental group, and this only happens in the reading and writing exams (Figure 3.1d). All the other lines (Figures 3.1a, 3.1b, and 3.1c) show negative slopes, but not all of them with the same degree (-5.17, -2.52 and -3.96). Both lines in the control group have a steeper negative slope (-5.17 and -3.96) that the one in the listening exams of the experimental group (-2.52). The separation between the real values and the regression line is bigger in Figure 3.1d with a value for the R_2 of only 0.002. The rest of the Figures also show low R_2 values, ranging from 0.018 to 0.043.

3.1.2 Speaking tests

As mentioned in Sections 2.1.1.2 and 2.5.3, the recordings in the speaking tests were analyzed to count for the vocabulary each student was able to use. Also, five judges evaluated parts of the oral production of six children from each class to measure the quality of those productions in terms of pronunciation. The following subsections will explore the results and the statistics extracted from these tests.

[2]Note that this number comes from the top β value in Table 3.4a
[3]Note that this number comes from the bottom β value also in Table 3.4a

3.1.2.1 Passive and active vocabulary

As mentioned in Section 2.1.1.2, the students in the speaking tests used some words autonomously and some others with the support of the researcher, who gave them three options to choose from. The words the students used with help were considered passive words while the ones they were able to use without help were accounted as active words (for a description of passive and active vocabulary, refer to Chapter 1.1.1). Table 3.5 shows the values and descriptive statistics for both types of vocabulary, both groups and both tests (i.e., pre-test and post-test).

Table 3.5 Descriptive statistics for the passive and active vocabulary found in the speaking pre-tests and post-tests

Vocabulary	Group	Test	Min	Max	Mean	Standard Deviation
Passive	Control	Pre-test	3	8	5.83	1.835
		Post-test	1	6	3.33	1.862
	Experimental	Pre-test	4	7	5.33	1.366
		Post-test	1	7	3.5	2.074
Active	Control	Pre-test	5	20	10.17	5.565
		Post-test	11	23	16.67	4.761
	Experimental	Pre-test	5	14	7.67	3.266
		Post-test	10	22	14.17	4.401

The table is divided into passive and active vocabulary and then into groups (i.e., control and experimental). This is then subdivided for each group into pre-test and post-test. Min refers to the minimum amount of words used by any one student, while max refers to the maximum amount of words. The mean is the average number of words used by the students in that specific group and test. The standard deviation refers to the dispersion in the number of words used by the students in that group and exam regarding the previously mentioned average.

Both the control and the experimental groups showed very similar numbers regarding the passive vocabulary. In the pre-test, the minimum passive words used by any one

student in the control group was 3, whereas in the experimental group it was 4. The minimum for both groups in the post-test was 1. In the case of the maximum number of words used by a student in each group, in the control group it went from 8 words in the pre-test down to 6 in the post-test while in the experimental group it remained at 7. The average number of passive words used by the students and the evolution of this average between the pre-tests and the post-tests was also very similar between groups. In the case of the control group, it had an average of 5.83 words per student in the pre-test, which went down to 3.33 in the post-test. The experimental group had 5.33 words per student in the pre-test (0.5 less than the control group) and 3.5 in the post-test (0.17 more than the other group). The results in the control group showed a steady standard deviation that went from 1.835 words in the pre-test to 1.862 in the post-test. In the case of the experimental group, although it showed a lower standard deviation than the one in the control group (1.366) in the pre-test, it ended rising above the other group's in the post-test (2.074).

Regarding the active language, both groups had very similar results in the minimum words used by any one student both for the pre-test and the post-test. In the case of the control group, these were 5 and 11 words respectively while in the experimental group they were 5 and 10. These numbers are not only higher than the ones regarding the passive vocabulary, they also showed a rising tendency (better results in the post-test than in the pre-test). Something similar happened with the maximum words said by any one student. In the case of the control group, they grew from 20 words to 23 and in the experimental group from 14 to 22. The average use of active words per student also rose in both groups. The control group showed an average of 10.17 in the pre-test and 16.67 in the post-test, while the experimental group went from a lower 7.67 to an also lower 14.17. The standard deviation of the control group went down from 5.565 to 4.761 while in the experimental group it rose from 3.266 to 4.401, both of which

are lower than the ones in the control group. The time factor related to the active vocabulary in each of the groups is shown in table 3.6.

Table 3.6 Effect of time in the active vocabulary acquisition in each of the groups

(a) Control group

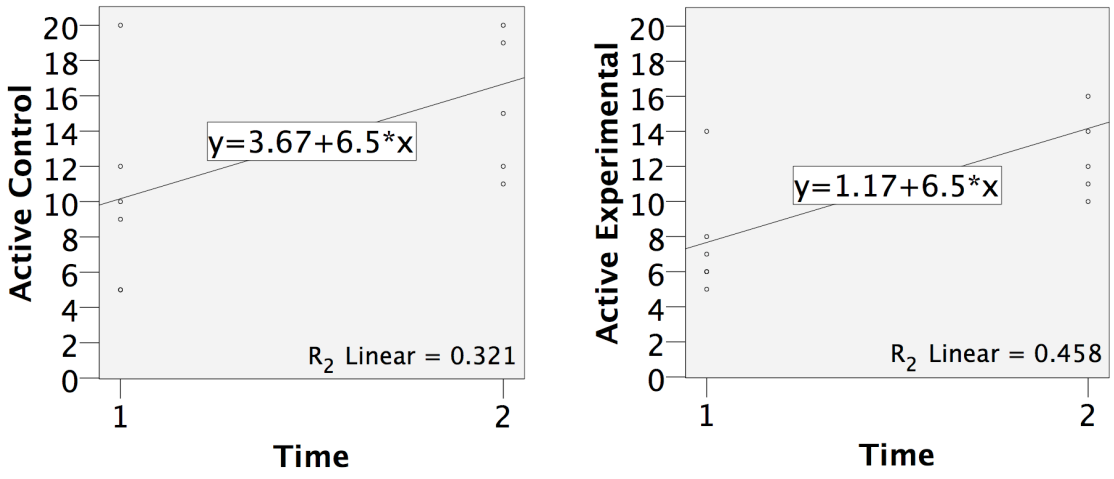
Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	3.667	4.727	.456
Time	6.500	2.990	.055

(b) Experimental group

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	1.167	3.537	.748
Time	6.500	2.237	.016

Table 3.6a shows a tendency towards significance in the control group with a significance value of .055, which is very close to .05. At the same time, Table 3.6b shows a significance value of .016, below .05 for the experimental group. Although both tables show a β value for Time of 6.500, the starting value of the control group was 3.667, while in the experimental group it was 1.167. The standard error is higher in the control group than in the experimental group both for the Constant (4.727 against 3.527) and Time factors (2.990 against 2.237).

The data in the previous tables is visually represented in Figure 3.2:



(a) Control group (b) Experimental group

Fig. 3.2 Regression lines of the active vocabulary in each group

Figure 3.2a presents an R_2 of 0.321 while the value in Figure 3.2b is 0.458. This means that the line shown in the second figure represents reality slightly better than the one shown in the first. Both regression lines are positive and both with a factor of +6.5. The regression line in the control group starts slightly below ten and rises up to around sixteen while the one in the experimental group starts between seven and eight and rises up to slightly above fourteen.

3.1.2.2 Pronunciation Accuracy

The following sections present the data related to the quality of the pronunciation of the students. It will be divided in two subsections. Subsection 3.1.2.2.1, dealing with the changes found in the pronunciation of each student, and subsection 3.1.2.2.2, dealing with some important differences found between the judges.

3.1.2.2.1 Evolution of the pronunciation of the students

In order to assess the evolution of the pronunciation of each student, the average grade of the different judges for each word has been taken into account. These average grades have been analyzed in terms of regression lines to try and find patterns and significance. A similar analysis has been done with the global average of the whole group. Table 3.7 presents the unstandardized coefficients, standard error and significance of these regression lines.

Table 3.7 Effect of time in the improvement in the pronunciation

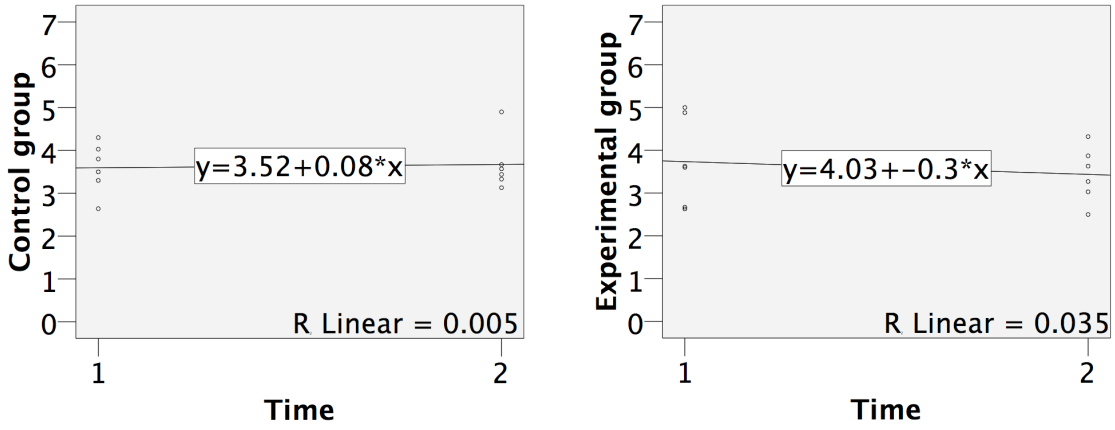
(a) Control group

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	3.517	.557	.000
Time	.078	.352	.828

(b) Experimental group

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	4.033	.784	.000
Time	-.298	.496	.561

Table 3.7a shows a positive tendency with a β of .078 and a starting value of 3.517 for the control group. In contrast, the experimental group, shown in Table 3.7b, starts with a higher value of 4.033, but has a negative tendency with a β of -.298. The standard error in both groups is very similar (.352 for the control group and .496 for the experimental regarding the Time factor). Neither of the groups shows any significance in Time (.828 and .561 respectively). Figure 3.3 shows the regression lines for both groups:



(a) Control group (b) Experimental group

Fig. 3.3 Regression lines for the pronunciation of each group

As shown in Figure 3.3, the line for the control group has a slightly positive tendency (+0.08) with a very low R_2 (0.005). On the other hand, the regression line for the experimental group has a negative tendency (-0.3) and the R_2 value is higher (0.035) than the one in the control group, although it is still very low.

Table 3.8 provides linear regression results for each student in both groups. The subtables in the left column illustrate the information regarding the students in the control group while the subtables in the right column illustrate the experimental group.

The levels of significance in Tables 3.8k and 3.8f, corresponding to student 6 in the control group and to student 3 in the experimental group respectively, stand out. Student 6 in the control group has a significance of .000 with a positive time coefficient

Table 3.8 Effect of time in the pronunciation

(a) Control group, student 1

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	4.467	.747	.000
Time	-.667	.473	.164

(b) Experimental group, student 1

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	5.440	.806	.000
Time	-.560	.510	.278

(c) Control group, student 2

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	1.840	.790	.024
Time	.800	.500	.116

(d) Experimental group, student 2

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	3.333	.752	.000
Time	.267	.475	.577

(e) Control group, student 3

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	5.033	.742	.000
Time	-.733	.469	.124

(f) Experimental group, student 3

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	6.367	.752	.000
Time	-1.367	.476	.006

(g) Control group, student 4

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	4.400	.802	.000
Time	-.367	.507	.472

(h) Experimental group, student 4

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	2.300	.677	.001
Time	.367	.428	.396

(i) Control group, student 5

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	3.667	.820	.000
Time	-.167	.519	.749

(j) Experimental group, student 5

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	2.767	.631	.000
Time	-.133	.399	.739

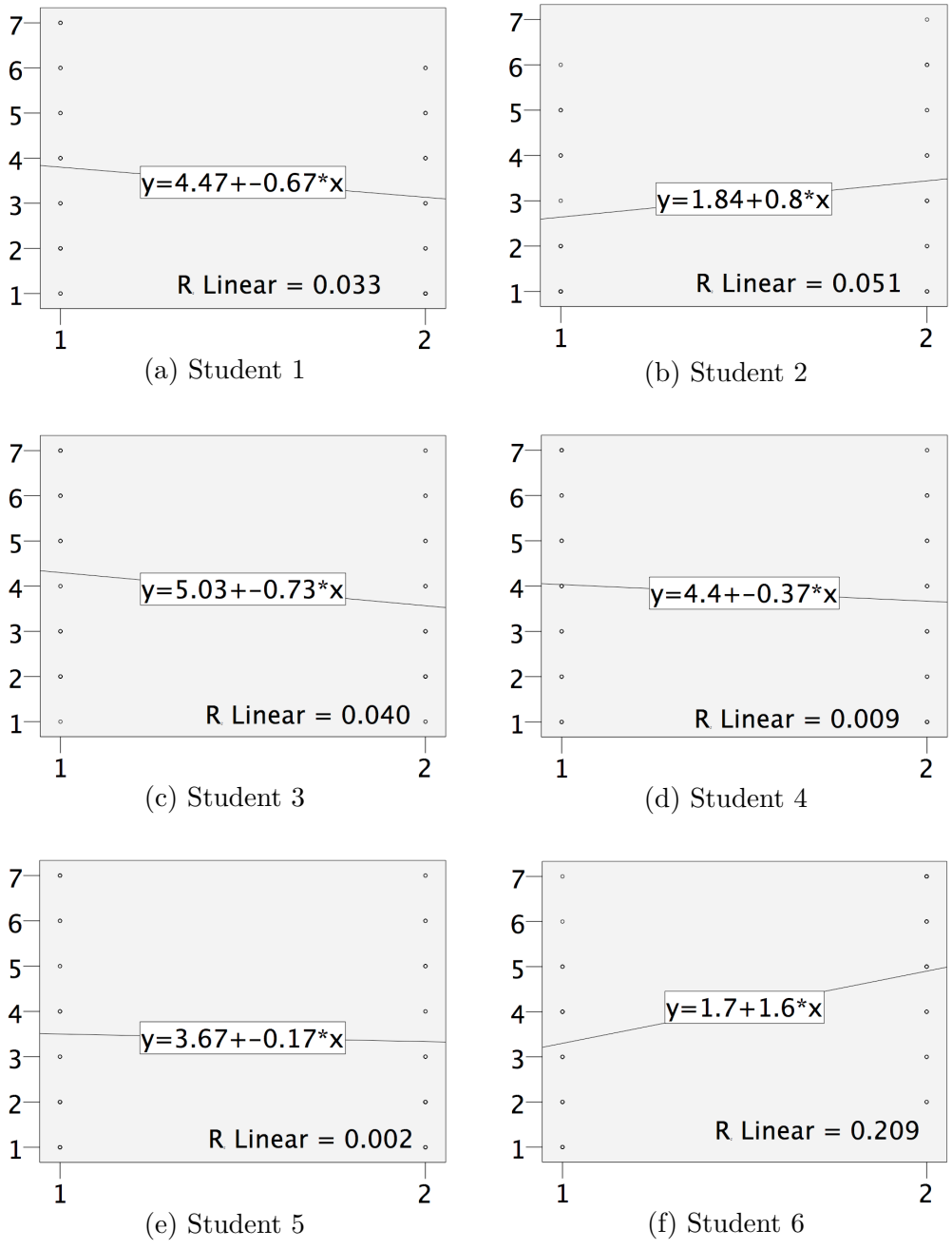
(k) Control group, student 6

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	1.700	.645	.011
Time	1.600	.408	.000

(l) Experimental group, student 6

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	4.000	.744	.000
Time	-.367	.471	.439

of 1.600 and a starting point (β) of 1.700. Student 3 in the experimental group presents a significance of .006. This learner has a starting point (β) of 6.367 and a negative time factor of -1.367. Figures 3.4 and 3.5 are the visual representation of the information presented in the previous table.

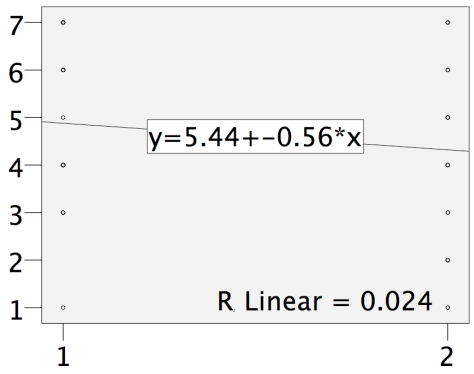


(a) Student 1 (b) Student 2

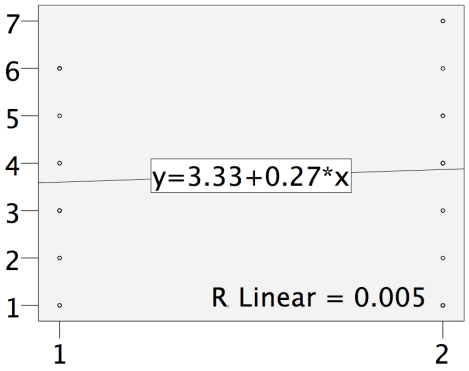
(c) Student 3 (d) Student 4

(e) Student 5 (f) Student 6

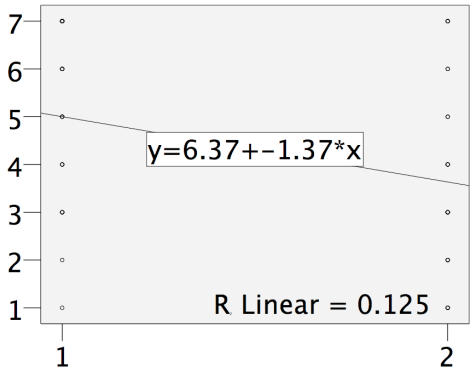
Fig. 3.4 Regression lines for the pronunciation of the students in the control group



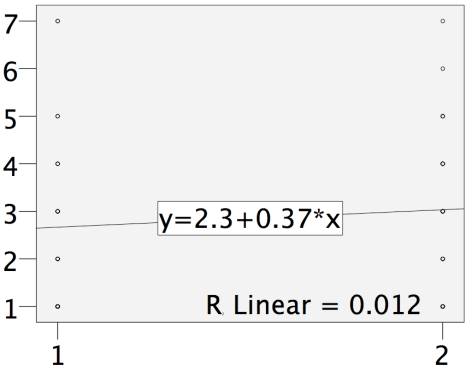
(a) Student 1



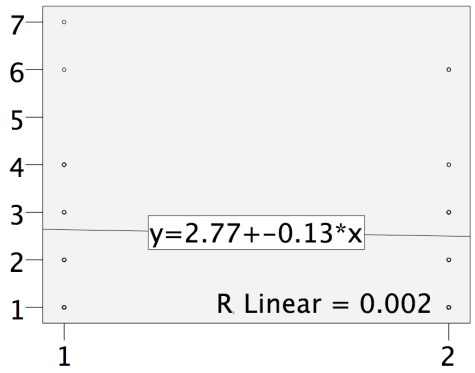
(b) Student 2



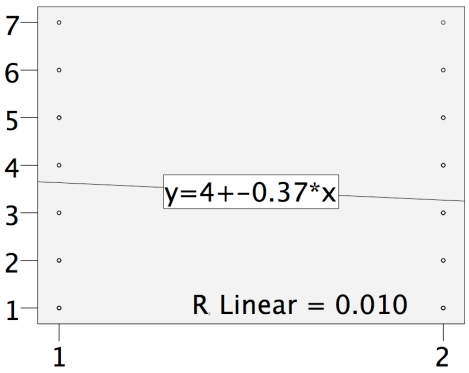
(c) Student 3



(d) Student 4



(e) Student 5



(f) Student 6

Fig. 3.5 Regression lines for the pronunciation of the students in the experimental group

In the case of the control group (Figure 3.4) only two students show a positive regression line: students 2 and 6 (+0.8 and +1.6). The rest of the students show a negative regression line. Student 1 shows a decline of -0.67; student 3, -0.73; student

4, -0.37; and student 5, the most gentle negative slope in this group, with -0.17. This student shows an R_2 of only 0.002, while student 6 shows the highest R_2: 0.209. Students 1, 2 and 3 show very similar R_2 (0.033, 0.051 and 0.040 respectively), while for student 4, it is only 0.009. None of these R_2 values are close enough to 1 to consider the line a good visual representation of the data.

In the experimental group (Figure 3.5) only two learners, again, show a positive regression line, students 2 and 4 (+0.27 and +0.37 respectively). The rest of the learners show a decline in the quality of their pronunciation. The highest decline is by student 3 (-1.37), which is the highest decline in all 12 learners. Student 1 shows a decline of -0.56 and student 6 of -0.37. Student 5 also has a decline, but a more gradual one (-0.13), the lowest value in all the learners. R_2 values in this group are lower than in the control one. Four of them are below 0.013, student 1 is 0.024 and student 3 has the highest with 0.125.

3.1.2.2.2 Within-judge and between-judge variability

This section will analyze the results extracted from the responses that the five judges reported regarding the pronunciation of words by each of the twelve students selected for this analysis (six from the control group and six from the experimental one). One of the first things to highlight from the data is the fluctuation. This fluctuation happens between judges but also within the same judge. Figure 3.6 presents an example of this fluctuation. It shows a graph representing the assessment by the different judges of the different words selected for one of the students in a certain test (pre or post). Each color bar represents one of the judges. The vertical axis shows the grade inside the Likert scale, which goes from 1 to 7. The horizontal axis shows the words being assessed.

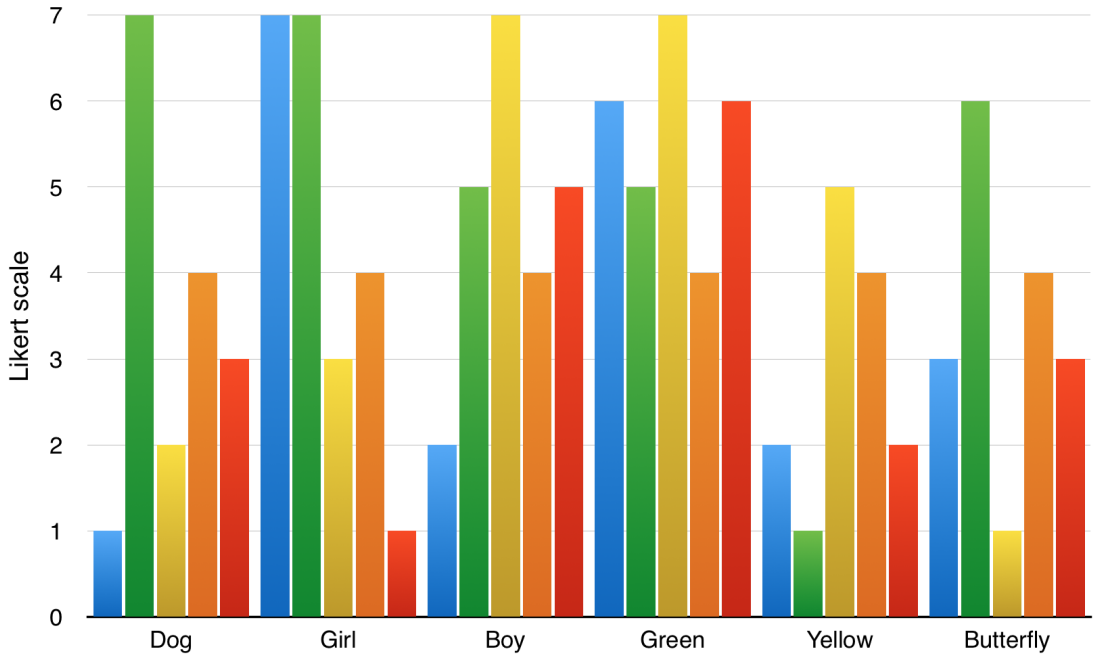


Fig. 3.6 Variability observed in the pronunciation assessment

The blue judge rated the pronunciation of the word 'dog' as completely non-native (1), whereas the green judge rated it as native-like (7). The other three judges ranged from 2 to 4. On the other hand, both the blue and green judges thought that the student had pronounced 'girl' in a native-like way (7), but the red judge thought she had pronounced it completely non-natively (1). The yellow and orange judges rated the pronunciation with 3 and 4 respectively. Then, in the word 'boy', the yellow judge considered the pronunciation worth 7, but the blue one only valued it 2 (while green and red rated it with a 5 and orange with a 4).

Figure 3.7 shows the mean of these grades (blue column) and the corresponding standard deviations. The word 'green' has the lowest standard deviation, around 1.0, and the word 'girl' has the highest standard deviation, above 2.5. Note that a deviation of 1.0 represents 14.3% over the 7 points in the Likert scale and a deviation of 2.5, represents 35.7%.

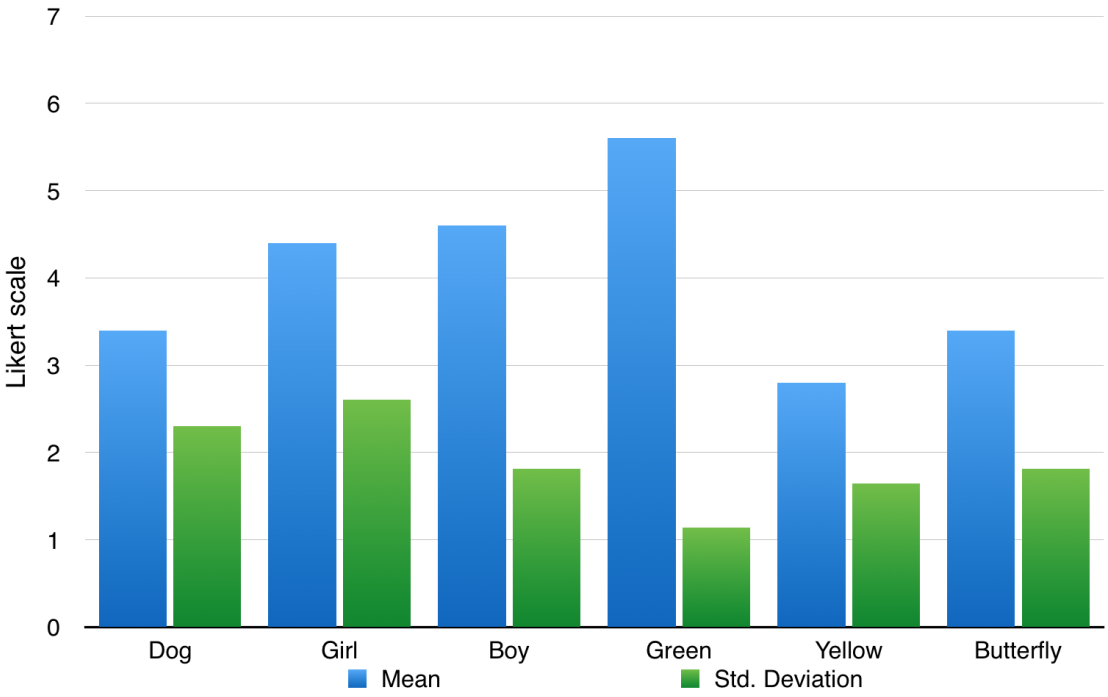


Fig. 3.7 Standard deviation in the grades of the judges

Table 3.9 is a complete list of all the standard deviations between the judges including the average standard deviations for each student and each test. It is divided in two subtables: Table 3.9a shows the results for each student in the pre-test, while Table 3.9b shows those results in the post-test. The last column in each of these subtables represent the average standard deviation within each learner.

The deviations are high considering that there were only five judges and that the Likert scale ranged from 1 to 7. The deviations go from 0.447 (student 4 in the control group, word 6 in the pre-test or student 5 in the experimental group, word 2 in the pre-test) up to 2.608 (student 4 of the experimental group, word 2 in the pre-test). In average, all the standard deviations of the different words of each student, both in the pre-test and in the post-test groups, range from 1.263 (post-test, control group, student 5) to 1.888 (pre-test, experimental group, student 4). The average standard deviation lowers from the pre-test to the post-test in most of the learners. There are

Table 3.9 Standard deviation between judges

(a) Pre-test deviations

Group	St.	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	**Average**
Cont.	1	1.732	1.342	1.517	1.581	1.817	2.280	**1.711**
	2	1.304	1.643	1.342	2.168	1.643	2.049	**1.692**
	3	2.302	1.643	1.000	1.643	2.280	1.304	**1.695**
	4	1.095	1.732	1.095	2.302	1.304	0.447	**1.329**
	5	1.414	2.280	1.304	1.304	0.894	1.000	**1.366**
	6	1.817	1.871	1.225	1.673	1.924	2.000	**1.752**
Exp.	1	2.121	1.414	1.817	1.304	1.643	1.517	**1.636**
	2	1.225	1.225	1.924	1.643	1.304	1.789	**1.518**
	3	1.949	1.643	1.483	1.924	0.707	1.924	**1.605**
	4	2.302	2.608	1.817	1.140	1.643	1.817	**1.888**
	5	2.074	0.447	1.140	1.517	1.304	2.345	**1.471**
	6	1.483	1.924	1.304	1.140	0.447	1.304	**1.267**

(b) Post-test deviations

Group	St.	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	**Average**
Cont.	1	2.588	1.924	1.140	1.000	1.581	2.000	**1.706**
	2	1.789	1.095	0.894	1.517	1.483	2.191	**1.495**
	3	2.000	2.121	1.000	0.837	1.342	1.643	**1.490**
	4	2.387	1.304	1.095	1.225	0.707	1.140	**1.310**
	5	1.095	1.924	1.789	0.837	1.095	0.837	**1.263**
	6	1.643	1.225	1.304	1.789	1.673	0.837	**1.412**
Exp.	1	1.414	1.304	2.191	1.000	1.140	1.949	**1.500**
	2	1.817	1.581	1.483	2.302	0.837	1.673	**1.616**
	3	2.387	2.168	1.140	1.581	1.140	0.548	**1.494**
	4	2.280	1.581	1.949	1.304	1.517	1.924	**1.759**
	5	1.304	2.168	1.643	0.707	0.894	1.517	**1.372**
	6	1.924	1.225	0.837	1.673	1.949	2.168	**1.629**

two exceptions to this, both in the experimental group: student 2 (1.518 in the pre-test against 1.616 in the post-test) and student 6 (1.267 in the pre-test against 1.629 in the post-test).

The difference in the standard deviations within subjects and in the same test can be very big. Student 5 in the experimental group has, in the pre-test, the lowest standard deviation, with 0.447 (Word 2). This same student also gets one of the highest ones, with 2.345 (word 6 in the pre-test). In this case, the average standard deviation was not too high, 1.471, but the fluctuation between the different words was. The same happens for other students and with other words, as is the case of student 4 in the control group, again in the pre-test, also with 0.447 (Word 6) and 2.302 (Word 2); or student 1 in the control group in the post-test, with a 1.000 (Word 4) and a 2.588 (Word 1). The student with the least fluctuation in the standard deviations is student 6 in the control group in the pre-test, ranging from 1.225 (Word 3) to 2.000 (Word 6).

3.1.3 Motivation tests

As mentioned in Section 2.5.4.1, the adaptation of the mini-AMTB showed some problems. Most were in the objective part of the test, but there were some unforeseen reactions in the subjective part too. Even so, this section will explore the problems found in the objective part of the test and will present the results extracted from the subjective part.

3.1.3.1 Objective questions

All the objective questions had incongruences in the experimental group and three of the questions showed incongruences in the control group. Table 3.10 shows the number of incongruent answers in each question and the percentage of incongruent answers in each group and in each question[4]. Note that the totals in these tables account for the total of students in each group that showed incongruences (they are not the addition of all the mistakes for all the questions).

[4]The experimental group had a total of twenty-one subjects that took both the pre-test and the post-test, whereas in the control group twenty-two subjects took both of those tests.

Table 3.10 Experimental school: incongruences in the objective questions

	Q 1	Q 2	Q 3a	Q 3b	Q 3c	**Total**
Experimental:	4 19.0%	5 23.8%	3 14.3%	3 14.3%	1 4.8%	**11** **52.4%**
Control:	6 27.3%	1 4.5%	0 0.0%	1 4.5%	0 0.0%	**7** **31.8%**

The percentages in the table show that more than half of the students in the experimental group presented some kind of incongruence (see page 46 for examples of these incongruences). The percentage in the control group drops down to 31.8%. Another thing to highlight is that the experimental group percentage of incongruences ranges from 4.8% in question 3c to a 23.8% in question 2. In the case of the control group, there are two questions with 0.0% incongruences (3a and 3c), but the first question shows 27.3%, which is the highest number in both groups.

As explained in Section 2.5.4.1, the incongruences presented in Table 3.10 required the mini-AMTB to be piloted again in a second school. Table 3.11 presents the incongruences found in this second school. The first column indicated which of the three groups the information refers to (i.e., A, B or C).

Table 3.11 Second school: incongruences in the objective questions

	Q 1	Q 2	Q 3a	Q 3b	Q 3c	**Total**
A:	3 34.8%	4 4.3%	0 60.9%	5 4.3%	0 0.0%	**17** **73.9%**
B:	6 13%	1 17.4%	0 0.0%	1 21.7%	0 0.0%	**10** **43.5%**
C:	1 4.2%	0 0.0%	1 4.2%	0 0.0%	2 8.3%	**7** **12.5%**

The results, when analyzed separately by group, are quite different from the results in the experimental school. In this case, group A had a higher level of incongruences than the groups in the experiment, reaching 73.9%. Group C, on the other hand, had

a lower percentage, having only 12.5%. Group B showed 43.5% incongruences, which would be a value somewhere between the two groups in the experiment. This table also shows that all the questions had incongruences in one group or another. In the case of question 1, they are present in all three groups. On the other hand, in the case of questions 3a and 3c, the incongruences only appeared in the third group which, paradoxically, is the group with a lowest number and percentage of incongruences. The table also shows that all the groups in this school have the lowest percentage range of 0.0%, but the highest range varies between groups. Group A reaches 60.9% in question 3a, while group B reaches only 21.7% in question 3b and group C merely 8.3% in question 3c. When comparing global numbers by joining all the students of the different groups (Table 3.12), both schools have very similar numbers[5].

Table 3.12 Comparison of number of students with incongruences between schools

	Q 1	Q 2	Q 3a	Q 3b	Q 3c	**Total**
Experimental school:	10	6	3	4	1	**18**
	23.3%	14%	7%	9.3%	2.3%	**41.9%**
Second school:	12	5	15	6	2	**30**
	17.1%	7.1%	21.4%	8.6%	2.9%	**42.9%**

All the questions in both schools show some degree of incongruence. In the case of the percentages, although the totals of the experimental school and the second school are very similar (41.9% and 42.9% respectively), there are big differences within some of the questions. For example, while in question 2 the former shows 14.0%, the latter shows only 7.1%. Or, while the groups in the second school answered question 3a with 21.4% incongruences, the ones in the experimental school shows only 7.0%.

[5]In the experimental school the test was taken by 43 students and in the second school by 70

3.1.3.2 Motivation questions

The purpose of using the motivation questions was to see how the motivation had changed in the experimental group in contrast with the control one over time. The mini-AMTB contained three questions that were designed to measure the motivation of the students. These questions were the only ones explored because they were the ones targeting the second hypothesis in this study.

3.1.3.2.1 Descriptive statistics for motivation

This section will explore the descriptive statistics for the motivation tests that were performed right before the experiment started and at the very end. Table 3.13 presents the descriptive statistics for the motivation questions in the mini-AMTB.

Table 3.13 Descriptive statistics for the motivation questions

(a) Control group

	N	Mean	Std. Deviation
Pre-test Mot 1	23	6.304	1.490
Pre-test Mot 2	23	6.217	1.475
Pre-test Mot 3	22	6.363	1.048
Post-test Mot 1	23	6.521	.897
Post-test Mot 2	23	6.565	.843
Post-test Mot 3	23	5.869	1.632
Valid N (listwise)	21		

(b) Experimental group

	N	Mean	Std. Deviation
Pre-test Mot 1	22	6.363	.953
Pre-test Mot 2	22	6.409	1.098
Pre-test Mot 3	22	6.590	.666
Post-test Mot 1	21	5.761	2.095
Post-test Mot 2	21	5.666	2.152
Post-test Mot 3	21	5.619	2.178
Valid N (listwise)	18		

Table 3.13a shows that, although both tests were taken by 23 students (N), question 3 in the pre-test was not answered by one of them (N = 22). Only 21 students in this control group took both tests, as can be seen at the end of the table—Valid N (listwise). Also, the first two questions have higher values in the post-test than in the pre-test (from 6.304 to 6.521 in the first question and from 6.217 to 6.656 in the second) but the third question shows a decrease in the post-test (from 6.363 to 5.869). Regarding the experimental group, Table 3.13b shows that 22 students (N) took the pre-test and 21 the post-test, but only 18 took both tests. In this case, the average motivation of the students in the post-test presents lower values than in the pre-test in all three questions (from 6.363 to 5.761 in the first question, from 6.409 to 5.666 in the second and from 6.590 to 5.619 in the third). Similar averages can be observed inside each set of three questions, except for the post-test in the control group, where question 3 shows a difference of more than 0.6 points (which in the 7-point Likert scale would correspond to more than 10%) compared with the other two questions..

As for the standard deviation, the last question of the post-test in the control group shows a higher value (1.632) than the rest of the questions in that group. The lowest values can be found in the first two questions of the post-test, with .897 and .843 respectively. This contrasts with the same questions for the same group in the pre-test, which had standard deviations of 1.490 and 1.475 respectively. Regarding the experimental group, the standard deviations in the pre-test range from .666 (Q3) to 1.098 (Q2). All the questions show higher values in the post-test, ranging from 2.094 (Q1) to 2.178 (Q3). Another thing that can be observed in the table above is that, although the experimental group shows higher averages and lower standard deviations than the control group in the pre-test, this shows the opposite trend in the post-test.

3.1.3.2.2 Mann Whitney U

The Mann Whitney U test illustrated in Table 3.14, does not show any significant differences between the groups in any of the questions, either in the pre-test or in the post-test, all showing significance values above .05.

Table 3.14 Mann Whitney U test comparing the motivation tests of both groups

(a) Pre-tests

	Pre Mot 1	Pre Mot 2	Pre Mot 3
Mann-Whitney U	226.000	243.000	229.500
Asymp. Sig. (2-tailed)	.463	.786	.722

(b) Post-tests

	Post Mot 1	Post Mot 2	Post Mot 3
Mann-Whitney U	206.000	193.000	238.500
Asymp. Sig. (2-tailed)	.312	.176	.938

3.1.3.2.3 Linear Regressions

Linear regressions were performed in order to gain a better insight into the descriptive statistics shown in Table 3.13 and to try to see the effect of time in each group. Table 3.15 shows the time coefficients (starting points) and their variation in time for both groups. Note that the time factor represented in these figures corresponds to a length of around three months (the whole length of the experiment), not like the coefficients presented in Section 3.1.1.3, which had a time lapse of around twenty days.

In the first two questions in the control group (Tables 3.15a and 3.15c), the time factor shows an improvement in the motivation. In the case of question 1, this is of 0.217 points. This improvement, though, has a standard error of 0.363, so it could become a decline of -0.146 or rise up to 0.580 points. Regarding question 2, the improvement in the motivation with reference to time is of 0.348, with a standard error of .354. Question 3, on the other hand, shows that with every lapse, the students

Table 3.15 Coefficient Time for the three motivation questions

(a) Question 1, control group

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	6.087	.574	.000
Time	.217	.363	.552

(b) Question 1, experimental group

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	6.965	.773	.000
Time	-.602	.493	.229

(c) Question 2, control group

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	5.870	.560	.000
Time	.348	.354	.332

(d) Question 2, experimental group

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	7.152	.813	.000
Time	-.742	.518	.159

(e) Question 3, control group

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	6.858	.654	.000
Time	-.494	.411	.236

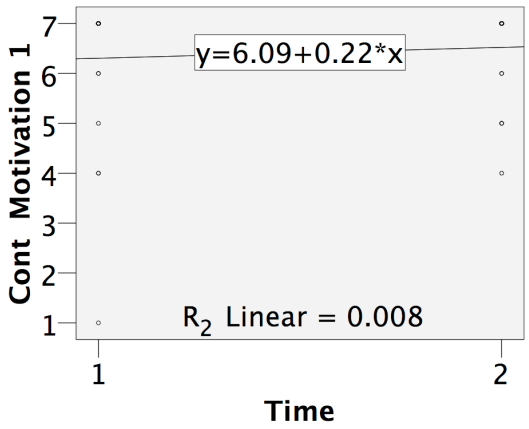
(f) Question 3, experimental group

Model	Unstandardized Coefficients β	Std. Error	Sig.
(Const.)	7.563	.764	.000
Time	-.972	.487	.052

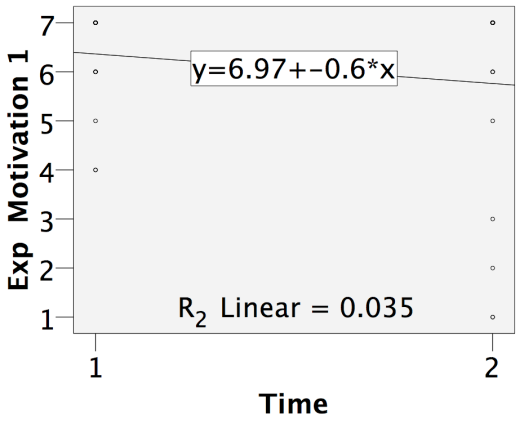
would lose an average of 0.494 points of motivation. Even if the standard error (0.411 points) was added to that number, the result would still be negative. The case of the experimental group is different. In this case, all three questions show a decline in time. The question with the highest decline would be Question 3 (Figure 3.15f), with a decline of 0.972. Also, although the standard errors of all three questions are bigger than those shown by the control group, they are not important enough to change the decline to an improvement of the motivation.

The data in table 3.15 can be better understood with the graphic representation of the regression lines. These regression lines can be seen in Figure 3.8, where the left graphs represent the control group and the right graphs represent the experimental one.

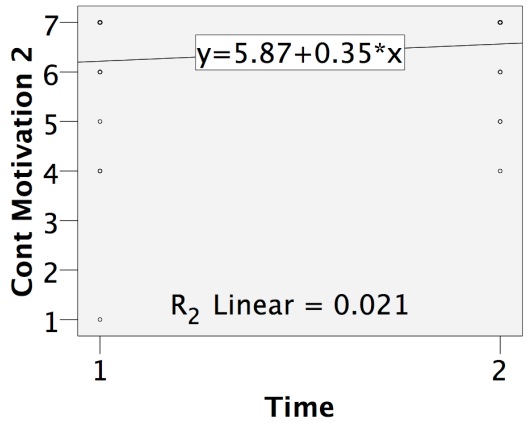
As expected by the data commented on in the previous sections, the first two questions in the control group present a slight rise (Figures 3.8a and 3.8c), while the



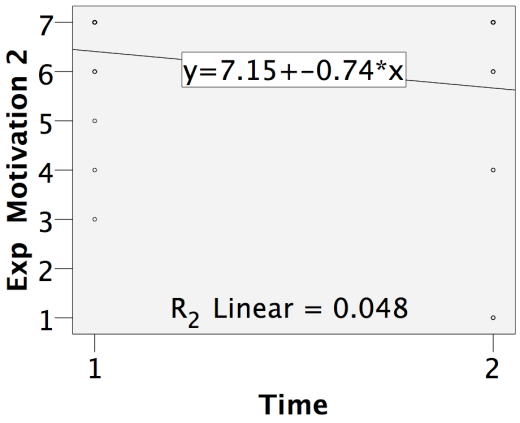
(a) Q1, control



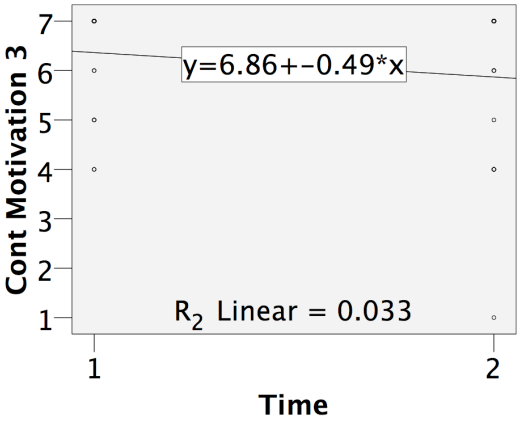
(b) Q1, experimental



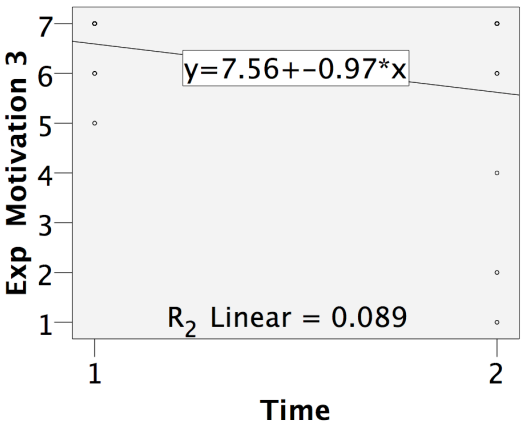
(c) Q2, control



(d) Q2, experimental



(e) Q3, control



(f) Q3, experimental

Fig. 3.8 Regression lines of the three motivation questions and the different groups

third question in the control group (Figure 3.8e) and all three in the experimental group (Figures 3.8b, 3.8d and 3.8f), present a decline. The decline in those questions of the experimental group is steeper, though not significantly, than the one shown in the third question in the control group: the negative slope in the control group is -0.49, while in the experimental group it is -0.6 in the first question, .0.74 in the second and -0.97 in the third. Also, the slope in the third question in the experimental group is steeper than the slope in the other two questions in this group.

3.2 Qualitative results

This section will explore the results of the qualitative part of the study. The purpose of using the information extracted from this exploration is to explain and expand on the results obtained in the quantitative part of the study and to gain better insight into what really happened in the classrooms.

3.2.1 Interviews with the teacher

These interviews, which were introduced in Chapter 2.1.2, will be analyzed in the following subsections. The most recurrent and important comments will be explored. A detailed list of these comments can be found in Annex D for the first interview and in Annex E for the comments in the second interview.

3.2.1.1 Analysis of the first interview

In the first interview, which lasted over thirty-eight minutes there were comments in each of the categories mentioned in Chapter 2.1.2 (command of the language of the students, motivation of the students, gestures, specific WBT techniques, general

classroom management and feelings of the teacher). A fast overview of the comments extracted from the interview shows:

- Fifteen positive comments regarding the command of the language of the students against a single negative one.
- Nineteen positive comments regarding the motivation of the students, but only a negative one.
- Five positive comments about gestures, one neutral and no negative ones.
- Seven positive comments regarding WBT techniques, and eight negative ones.
- Fourteen positive comments dealing with general classroom management but only a negative one.
- Only one negative comment regarding the feelings of the teacher but no positive one.

3.2.1.1.1 Comments about the language level

The general impression of the teacher after one month and a half was that the students were able to speak more and that they felt more comfortable speaking in English. As an example to prove this, she said that '*they always asked me the [sic] things in Catalan and now they speak more in English.*'. She also mentioned the case of two students that did not like English and refused to participate in class before the experiment started, but were participating then. She mentioned that she had the feeling that the students understood her better, maybe because of the gestures. She commented that she had the feeling that the experimental group had a better command of the language than the control group at that moment. On the other hand, she also said that another English teacher, one teaching the hour that was not part of the experiment, who was

mentioned in Section 2.3, thought that this difference was already there before starting the experiment.

The teacher commented that she could see that the students in the experimental group were now more prepared for the tests than the ones in the control group. She had the feeling that '*they remembered the vocabulary and the structures better than in the other method*' (the other method understood as the one applied in the control group). She also mentioned that she felt that the difference between the students with a higher English level and the ones with a lower one was getting smaller. She felt that the group was becoming more homogeneous.

3.2.1.1.2 Comments about the motivation of the students

One of the first things the teacher transmitted in the interview was that she felt that the children liked the new method. She said that she always forgot the scoreboard and the children reminded her to put it because '*they really love it.*'. On the other hand, and still referring to the scoreboard, she transmitted that '*she preferred to keep them motivated giving them nice comments every now and then, rather than using positive and negative rewards*'. The example of the two students that were mentioned in the previous point also refers to how she perceived that some students in the experimental group were increasing their motivation towards English. She said that these students now participated in the parts where she used *Teach! OK!*.

She talked about some children that did not bring the books to class and that had to use a notebook instead. She explained that in the experimental group they took the notebook out fast when needed while in the control group she had to insist several times. In the second part of this first interview, she mentioned how in the experimental group '*the students are more motivated*'. She illustrated this with the fact that all the students in the experimental group were submitting their homework while

a high number of students in the control group were not submitting it (in this case 9). She was asked if this also happened before the experiment and she replied that it did not. She said that, before the experiment, some students in the experimental group, although fewer than in the control one, were not submitting their homework.

At a certain point in the interview, she referred to the experimental group saying '*I think they are more motivated because, first, they speak more English, then, the class is more dynamic, [...] they lost the monotony [...] and they are like more active in the class.*'. And she said that she thought that they were paying more attention because they were more active. She also commented that in the experimental group speaking had become a game for the students. Also, she felt that no student was left behind in the speaking part, all of them were involved in it.

At a given point she commented that with WBT learners remembered content from one session to another. When she was asked why she thought this happened, she answered that '*because they are active in the class*'. She said that '*for them [sic] is like a game, this method*'. In the last part of the interview, she insisted several times on how the students were more active and participative. She said that they were now paying more attention.

3.2.1.1.3 Comments about the gestures

The comments related to the use of gestures were clustered in the last part of the interview. This contrasts with the comments regarding the level and the motivation of the students, which kept appearing from the beginning to the end. She mentioned that she found that gestures were very important and that, although she had always used gestures, she was using them even more with the experimental group. She gave the example of how she was starting to use gestures even with stories. She reckoned that although the students at the beginning were shocked about having to use gestures,

they soon started to like them. She thought that gestures were making the students be more active and participative.

3.2.1.1.4 Comments about specific WBT techniques

The first comment the teacher made regarding WBT techniques was that she was subconsciously trying to apply techniques as *Class! Yes!* with the control group. She thought this happened because she perceived this technique made things faster in the class. In fact, she realized she was using this technique a lot with the experimental group.

Then, she talked for a while about the *Scoreboard*. She mentioned that she did not like this technique. She felt that children should '*do things for themselves [...] and not because I'm going to give them a happy face or because they are going to receive a punishment [...]*'. Also, she commented that using the *Scoreboard* was stressful for her because she always had to have it in her mind, taking into account whether the students would have free minutes or homework at the end of the class.

She mentioned that she preferred other methods to keep the students motivated rather than the scoreboard, like giving them nice comments every now and then. She also commented that she did not like the fact that with the *Scoreboard* she was instructed not to use the names of the children when awarding them negative points. She said that it was usually the same person who was misbehaving every day. On the other hand, she saw that children understood the *Scoreboard* as a game or a contest. This made them really like the technique.

When she talked about *Teach! OK!*, she mentioned that it was encouraging some students to participate. She said that the fact that the partner would pressure them to share the information made them pay more attention and use English more. She

also said that by using this technique, the explanations were shorter and the students were able to repeat them to each other, thus, using the language.

She also referred to other little techniques inside WBT, like the command to deliver the books that was explained in Section 1.3.1: when the teacher says *Books*, all the students say *Books, books, books* at the same time they start fetching them and going back to their seats. She said that she did not like it because she thought the students enjoyed taking the responsibility of giving out the books on specific days[6]. In the second part of the interview, she talked about the rules (see page 14 for a list of the different rules and their explanation). She said that rule number 1 helped do things faster in the class. She also said that she found it difficult to locate the images for the rules on the Internet.

3.2.1.1.5 Comments about general classroom management

One of the first things she mentioned related to general classroom management was that she had noticed that in the classes with the experimental group she had more time to do things than with the control group. One change she saw between the groups is that in the experimental group, as they were mostly working in pairs, a lot of students were talking (using the language) at the same time, while in the control group there was only one student talking while the rest of the class listened. This meant that all the students in the experimental group were working at the same time, at least in the speaking part. She transmitted that she had the feeling that she had more control over the experimental group. When she was asked if, before the method, she had had the same feeling, she said that the experimental group '*was better but now [...] it was much more [sic] better.*'.

[6]It is common use in primary education to have some students assigned special chores inside the class. In this case, the teacher was referring to the chore of giving out the textbooks. When not using WBT, the teacher used two of the students to do this chore.

The teacher also mentioned that the students seemed to be more ready to do as the teacher said. She said that, while before the experiment they had to spend around ten minutes clearing up at the end of the class, now she only needed around three minutes. She was asked if after the experiment she thought she would be using the method, and she said that she thought she would. In fact, she commented that she was already using the method in other subjects with other groups (3rd graders) not involved in the experiment. In regard to this, she mentioned that the method worked even better in Science, where the students were using their mother tongue and could explain things in their own words and not just repeating when using *Teach! OK!*.

3.2.1.1.6 Comments about the feelings of the teacher

The only comment directly related to the feelings of the teacher was the one already mentioned in the previous section when talking about the *Scoreboard*. She said that using this tool was being stressful for her.

3.2.1.2 Analysis of the second interview

The second interview lasted around thirty-three minutes, close to five minutes less than the first one. There were also comments in each of the categories mentioned on page 34 (command of the language of the students, motivation of the students, gestures, specific WBT techniques, general classroom management and feelings of the teacher). A fast overview of the comments extracted from the interview shows:

- Four positive comments regarding the command of the language of the students and no negative ones.
- Nine positive comments regarding the motivation of the students and no negative ones.

- Two positive comments and a negative one about gestures.
- Nine positive and nine negative comments regarding WBT techniques.
- Eleven positive comments dealing with general classroom management and three negative ones.
- A positive comment and two negative ones regarding the feelings of the teacher.

3.2.1.2.1 Comments about the language level

The teacher mentioned that the children in the experimental group were producing more than before due to the fact that they were using techniques that allowed more people to be speaking at the same time. She commented that in the listening test (referring to the post-test), the students that did not like English before, had ended up performing very well.

Another change she detected was that the learners in the experimental group now realized they could understand some things in English. She thought that one of the reasons could be the gestures: She thought that '*they helped them* (the children) *memorize*' vocabulary.

3.2.1.2.2 Comments about the motivation of the students

Regarding the motivation, the teacher highlighted that the children were more involved in the English classes. When she was asked if she thought that the students that were unmotivated before the method were more motivated, at least, for the oral part, she said that '*yes, they are more motivated.*'.

When she was asked to compare that apparent motivation of the students in the experimental group with the motivation of the control group, she said that '*the attitude from the beginning* (of the experiment) *has been very different.*'. She was asked if this

had to do with the difference between the groups or because of the application of WBT, and she answered that she thought it was because of the method. On the other hand, she talked about the students in the experimental group being more active in class and liking the new method, but she said she did not think this was motivation.

She also highlighted that she perceived that the children liked the method. This made them remember the input they received better. One of the things from the method she thought children loved the most was the *Scoreboard*.

3.2.1.2.3 Comments about the gestures

Regarding the gestures, she reflected that, although she used a lot of gestures before the experiment, now she was using them much more. She thought that the use of gestures in the experimental group was helping the students memorize vocabulary.

One of the negative things she highlighted from the use of gestures was that sometimes the students would not use the *Mirror* technique when they were working in pairs. This technique, which was presented in Section 1.3.1, involved the students imitating the gestures that were being used by the teacher or by their peer. Thus, not responding to *Mirror* as she expected, the students were not using gestures as much as anticipated.

3.2.1.2.4 Comments about specific WBT techniques

The first comment the teacher made related to specific WBT techniques was about the use of the rules (see page 14 for a list of the different rules and their explanation). She felt that the rules were helping deal with the behavior of the children in the experimental group. She mentioned that she liked rule 4 (*Make smart choices*) because it allowed the children to think. She also liked rule 1 (*Follow directions quickly*). She said that she never used rule number 5 (*Keep your dear teacher happy*).

She also talked about *Teach! Ok!*, saying that, as the children had to explain to each other what the teacher just said, it was easier for them to realize what they had to do. This helped manage the class better. She confessed that sometimes she forgot about the order (student 1 or student 2). Also, she mentioned that she did not use *Switch*. She did not find it adequate to teach English. She would use it with longer explanations in subjects like science.

In this interview she confirmed the opinion she had transmitted in the first interview, where she said that she did not like the *Scoreboard*. She also said that she sometimes forgot about it. When she used it, she had the feeling she had to keep it in mind all the time and she could not relax. She mentioned that she thought that the *Scoreboard* was unfair. Punishing all the class because of just one student made the whole class turn against that student who always misbehaved. She also considered it unfair in the sense that the kids in the experimental group had to be punished with playtime when they were behaving much better than those students in other grades in the school and who were not being punished.

She said that she would be using WBT in both groups in the third term. She would not be using all of the techniques, though. She would be focusing on *Class! Yes!*, *Teach! OK!* and the first four rules. She reflected that she would be using it even if the results in this research project were inconclusive or even negative. She insisted on the fact that she would not be using the scoreboard.

3.2.1.2.5 Comments about general classroom management

One of the things that the teacher highlighted related to general classroom management is that, after carrying out the experiment, she had realized that it was possible to get the students to talk in English even if she had the whole group together. She commented that now she saw that there were alternatives to the teacher-to-student

student-to-teacher dynamic that she had been using before. She reflected that with techniques like the ones in WBT everyone could be practicing at the same time.

She highlighted the fact that she had more control over the experimental group than over the control one. She also said that in the former, with the same amount of time she had finished things faster than in the latter. She mentioned that they had ended having two extra sessions because they had been doing things faster. When she was reminded that at the beginning of the experiment she had mentioned the opposite[7], she said that '*[...] once they've realized what they have to do, they are faster [...]*'. She calculated that she saved around three or four minutes every class. She explained that this was because when she used WBT techniques like *Teach! OK!*, all the students understood what they were expected to do. She mentioned that, in contrast, in the control group she had some students that were '*disconnected*'.

A negative thing she highlighted about the method in terms of general classroom management was that she felt there was too much activity. There was always something to do, a command to follow, and the students had to do it fast. She did not know if this was due to the method or if it was influenced by the meetings she had had with the researcher who, she said, was insisting on keeping the students active. She thought there was no time for disconnecting, even between the classes, and she thought this was negative. She insisted on the idea that those disconnections are necessary for the kids to assimilate the knowledge. She highlighted two things in the method that had kept the once passive students active: the uncertainty of what was going to happen and the peer pressure.

[7]In one of the first meetings for the feedback about the application of the method and for revising the planning, the teacher had mentioned that using WBT made them go slower than the control group

3.2.1.2.6 Comments about the feelings of the teacher

One thing that the teacher did not like about WBT was that she realized the teacher had to be 100% active all the time. She found this stressful and difficult. Also, when she was asked whether this method could be extended to other situations (other schools, other teachers...), she said that she did not think so. She commented that it depended on the teacher. She thought it would only work if the teacher wanted to use it, not if the teacher was forced to use it.

Chapter 4

Discussion

This chapter will explore and discuss the results reported in Chapter 3. This exploration and discussion will shed some light on the two research hypotheses presented in Section 1.5. The results related to the language level (language and speaking tests) will be discussed first, followed by the results related to motivation.

4.1 The impact of WBT in the language level of the learners

This subsection will be divided into two further parts: one dealing with the general language tests and the other one dealing with the speaking tests. The results, both qualitative and quantitative, presented in Chapter 3 will be discussed in the light of the proposed hypotheses as well as with respect to the literature reviewed in Chapter 1.

4.1.1 WBT and the general language tests

Section 3.1.1 included a battery of statistical analyses to explore the differences between the control and the experimental group. The results from these tests will be discussed in this section.

4.1.1.1 WBT and the reading and writing tests

The most remarkable difference between the groups was found in the language tests through the Mann Whitney U test (see Section 3.1.1.2), which compared the exams of both groups. In this statistical test, there was a significant difference between the two groups in the reading and writing post-test, but no significant differences in any of the previous reading and writing tests. In this post-test, the control group had obtained an average grade of 65.291 points over 100, while the experimental group obtained 85.818. This meant that, at the end of the experiment, the experimental group was significantly better in reading and writing than the control group and that this difference had not been there before, as the lack of significance in the first three reading and writing tests confirm. In fact, in the pre-test, although the experimental group also had a better performance than the control group (an average of 76.913 against 74.652), this difference was smaller than the one obtained in the post-test (2.261 versus 20.527 points).

This difference was confirmed in the linear regression tests performed over these exams (see Section 3.1.1.3). Figures 3.1c and 3.1d showed a high contrast between both groups. While the control group had a negative slope with a β factor of -3.96, the experimental group had positive slope with a β factor of 1.29. This means that if the experiment had lasted 20 more days (approximate lapse between exams), the students in the control group would have lowered their grade around 3.96 points, while the students in the experimental group would have raised their marks around 1.29 points. It is important to point out that, although Tables 3.4c and 3.4d did not show significance in the time factor, the exploration of the regression line and the contrast of it with the Mann Whitney U results point at the fact that WBT played an important role in the improvement of the results of the experimental group.

This was corroborated and further explained by the qualitative interviews with the teacher reported in Section 3.2. Although she made a higher amount of positive comments regarding the command of the language of the students in the first interview than in the second (fifteen against four), she made only one negative comment regarding it and it was in the first interview. The teacher explained on several occasions that she had the feeling that the extensive use of gestures was helping the students in the experimental group follow the activities and learn the language better. She also mentioned in both interviews that the use of techniques like *Class! Yes!* or *Teach! OK!* was having a positive impact in her classes: the former because it was allowing her to do more things in the same amount of time since the students were more ready to do as she said (this seemed to be also influenced by the application of the WBT rules, especially number 1) and the latter because it was allowing the students more time to use the language (she contrasted the pair-distribution and interaction in WBT with the individual student distribution and the teacher-student-teacher interaction in the more traditional control group).

All these factors made her perceive that the learners in the experimental group were more prepared for the tests than the ones in the control group, even only six weeks into the experiment. In the first interview, she remarked that she thought the experimental group had a better command of the language but that another English teacher thought that this difference was there before the experiment. This has been proven to be partially true (see Table 3.1 for the descriptive statistics of the general language tests): the average grade in the pre-test of the experimental group was higher than the control group. Nevertheless, the differences between the groups were smaller in the pre-test than in the rest of the tests: in the pre-test the difference was 2.261 points; in the mid-tests they were 11.681 and 11.379; and in the post-test it went up to 20.527. This and the significance found in the Mann Whitney U test, together with

the linear regression, seems to prove that WBT did have an impact in those differences mentioned by the teachers.

The standard deviation in the reading and writing exams was higher in the experimental group than in the control group in the pre-test (9.909 points of difference), but in mid-test 1, this changed and the experimental group showed a lower value than the control group (7.983 points lower). In mid-test 2, both groups had similar results (with the experimental group 0.793 above the control one), but in the post-test, the experimental group had, once again, a lower standard deviation than the control group (6.820 points lower). Also, in the experimental group the pre-test showed the highest standard deviations while in the control group the pre-test showed the lowest standard deviations. Having a lower standard deviation in a class is an indication that the group level is more homogeneous, which means that the differences between the top tier students and the bottom tier students are smaller. A lower standard deviation with a higher average grade is always a good sign, and is precisely what can be observed in the experimental group but not in the control group.

In this case, the ideas presented in the previous paragraphs are backed by the opinions of the literature reported in Chapter 1. Gestures seem to have made input more comprehensible, a fact that Ellis (1985) mentioned as an essential factor for allowing language acquisition. The use of gestures in the experiment is in line with theories such as the TPR and, as explained in Section 1.2, authors like Gullberg & De Bot (2010), McCafferty (2002) or Gregersen (Olivares-Cuhat) believed that using gestures should improve the learners' comprehension and allow for a better creation of Zones of Proximal Development. This also links with the NLP theory of the perceptive chanels (VAK) mentioned in Section 1.1.4, which recommended using a variety of sensory learning styles to increase the group rapport. According to the comments of the teacher and the results in the reading and writing tests, this seems to have

played a role in the results obtained by the experimental group. Other authors like Lightbown & Spada (2006) or Richards & Rodgers (2001) have mentioned factors like interaction, conversation and language use or the use of pair-work as important factors in a Communicative Approach environment. These factors have been applied throughout the experiment, as the teacher explained, with the use of the *Teach! OK!* technique, and might have also contributed to the positive results mentioned before.

4.1.1.2 WBT and the listening tests

The statistics also showed differences in the listening exams. In this respect, in the Mann Whitney U test (see Table 3.1.1.2), there were no significant differences, but only tendencies. These tendencies can be found in the pre-test, mid-test 1 and mid-test2, with values approaching .05. In this case, it seems that there were some minimum differences between the group before the experiment, but these minimum differences were less important in the post-test. Those differences can be seen in the average grades of each group (see Table 3.1), where the experimental group always had results above the control group. These results, though, were not too big in the pre-test, with only 3.393 points of difference. In the other three exams, the difference was 13.182, 15.993 and 11.439 points respectively. Regarding the standard deviations, although the control group showed more homogeneity in the pre-test, it fast became less homogenous. In the mid-test 1, the mid-test 2 and the post-test, the experimental group showed deviation values below the control group. In fact, the deviation value in the listening post-test of the experimental group is the lowest in both types of language tests (reading and writing or listening) in all the different applications (pre-test, mid-tests and post-test), with only 12.861. Also, this post-test shows the maximum difference in homogeneity, with 15.294 points of difference between the group. Thus, although there is no real significance in the comparison of the two groups, the improvement in the experimental

group seems clear, especially when compared to the lower average grades and bigger and growing standard deviations of the control group.

This idea is supported by the results of the linear regression tests shown in Table 3.4, where the control group shows a significance of .046 in the Time factor. This significance implies that time had a real negative impact on the control group (the β factor was -5.166). There is no such significance in the experimental group, though.

Both groups got worse in their listening results (this might indicate that the difficulty of the exams was higher), as can be seen in Figures 3.1a and 3.1b, with both regression lines having negative slopes. The lowering of the grades, nevertheless, is higher in the control group, with a decline of 5.17 against the decline of 2.52 of the experimental group. This means that if the experiment had lasted twenty more days (approximate lapse between exams), the students in the control group would have lowered their listening grade approximately 5.17 points. Regarding the students in the experimental group, they would have lowered their listening grade around 2.52 points.

The quantitative results commented on in the previous lines had already been perceived by the teacher, who explained the reason for some of these changes. Some of the reasons were mentioned above, when referring to the reading and writing exams, and had to do with the use of gestures to improve the comprehension or with the use of certain WBT techniques to improve the speed in the response of the students or to increase the amount of use of the language. Another comment by the teacher that could explain the differences between the groups, especially in terms of lower standard deviations, was that the students in the experimental group who did not use to like English before the experiment, had started liking it and participating actively in the classes. In this sense, she also commented that the more passive students in the group were starting to be more active thanks to the uncertainty of what was going to happen in the class (the surprise factor linked either to the novelty of the new method

being applied or to the fun that the creators of WBT claim to be embedded within their teaching system) and thanks to the peer pressure (related to the extended use of pair-work). Regarding the average grades, one of the explanations she gave was that the learners in the experimental group were understanding more English and were starting to be more aware of their own progress. Also, this could have been affected by the fact that the students knew what they had to do in each learning activity thanks to the use of WBT techniques like *Teach! OK!*

The issues raised in the previous section regarding the reading and writing level improvement, and how this improvement is in accordance with previously published work in the field, also apply here. Gestures probably helped the learners in the experimental group have better listening grades and lower standard deviations; the combination of the perceptive channels (VAK) done by the WBT might have also contributed to these improvements; the amount of the interaction derived from the techniques embedded in WBT also seems to have had an impact on these changes. Section 1.4 explored how WBT fitted inside the twelve principles that Brown (2002) considered essential in any English teaching approach. The teacher's comments support some of the relations mentioned in that section and contribute to the understanding of the differences between the control and the experimental group in the listening exams.

The first principle mentioned by Brown was Automaticity. The teacher explained in several moments throughout the interviews that she felt that the students in the experimental group were able to repeat what she had just explained, which had benefitted both their understanding of what was expected from them and their time of use of the language. Principle number 6 talked about the Language Ego, and this was mentioned by the teacher when she referred to those students that used to be passive in the class and that had taken a step forward since the application of WBT and had started participating more. This comment could also be related to the eighth principle,

which talked about Risk Taking and about offering learners reasonable challenges. There was a moment in the first interview when the teacher explained that for children in the experimental group, speaking had become a game. The eleventh principle talked about Interlanguage and about how important it is to be aware of the interlanguage of the learners. The teacher explained that in the experimental group the main interaction was between students and not between teacher and student, and that this made her feel she had more control over the experimental group, probably because with WBT she had more time to get to know at what interlanguage her students were while the learners were interacting with each other and she could go around the class monitoring their productions.

4.1.2 WBT and the speaking tests

Section 3.1.2 explored the results obtained from the speaking tests. It analyzed the vocabulary the learners were able to use and the quality of their pronunciation in their productions. This section will extract conclusions from those results, from the comments of the teacher described in Section 3.2.1 and from the existing literature reviewed in Chapter 1.

4.1.2.1 WBT and vocabulary acquisition

Section 3.1.2.1 analyzed the results obtained in the speaking tests from the point of view of the vocabulary. It described the amount of passive and active language used both in the pre-test and the post-test by a selection of six learners from each group. Table 3.5 showed the descriptive statistics related to this analyisis of which several issues are worth highlighting. Both groups produced fewer passive words in the post-test than what they had produced in the pre-test. This would have been problematic had they not improved their results in the active use of words, but there

was an increase in the active vocabulary of both groups. Section 1.1.1 explored the process of vocabulary acquisition (Terrell, 1986) and stated that the only way to increase the active language of a learner was to access items that were already in their passive language. Thus, it is normal that, having transferred some of the words from the passive to the active vocabulary during the experiment, the amount of passive words lowered and the amount of active words increased.

In terms of active vocabulary, both groups showed an average gain of 6.5 words between the pre-test and the post-test. The difference is that the students in the control group had been able to use an average of 10.17 words in the pre-test, while the students in the experimental group were only able to use 7.67. So, the difference of 6.5 words in the control group would correspond to an improvement of roughly 69%, while in the experimental group it would mean an approximate 85% improvement. This difference is corroborated by the linear regression tests (see Table 3.6), where the experimental group shows significance in the time factor with a .016 value while the control group only shows a tendency with a .055 value. The β factors, which in this case correspond to the difference between the average of the pre-test and the post-test, are the same for both groups (6.500), but the improvement is more significant in the experimental group. This means that if the experiment had lasted three more months, the experimental group should had gotten even closer to the performance of the control group in terms of active vocabulary.

The teacher had confirmed this difference in the interviews by stating that she had the feeling that the students in the experimental group were understanding things better and were remembering the vocabulary and the structures better than the other group too. She insisted several times that the gestures applied in the WBT could have caused this difference. Also, the use of the *Teach! OK!* technique and the fact that they had to repeat what the teacher had said, might have helped the process of language

transfer and might have influenced this improvement. The teacher also mentioned, as was discussed in the previous section, that the differences between the students in the experimental group were getting smaller. This did not happen to the selection of 6 students in the speaking test regarding the active vocabulary. The standard deviation between the students in this group rose from 3.266 to 4.401 while in the control group it decreased from 5.565 to 4.761. So, although the teacher was referring to the whole group and only a small sample took this test, her opinion did not match the statistical results in this case.

The results mentioned in the previous paragraphs regarding the vocabulary acquisition are important because, if Terrell (1986) was accurate in his description of the process of language transfer, the active vocabulary of the students will be both relatively permanent and with an immediate access when needed because it will have been acquired. Thus, improving the process of language transfer will make the learners better users of the language, with better productive skills because they will be able to be more fluent when trying to use a specific word. This is related to the 12th principle stated by Brown (2002): Communicative Competence. And, as was presented before, the group that had a more significant impact in this sense was the experimental one.

4.1.2.2 WBT and the pronunciation

The results described in Section 3.1.2.2 show a negative effect on the pronunciation of the group that used WBT. This can be seen in Figure 3.3, where the control group shows a positive slope and the experimental group a negative one. Nevertheless, Table 3.7 shows that the time factor is not significant in either of the groups, so it is difficult to know if the control group would continue improving and the experimental one getting worse. Significance, though, is found when studying the regression of each individual student. Student 6 in the control group and student 3 in the experimental

one show values below .05 that confirm that, in their cases, the time between the pre-test and the post-test was crucial. In the case of the student in the control group, every unit of time (understood as the three months that the experiment lasted) should mean an improvement in pronunciation of 1.600 points (β factor), while for the student in the experimental group, it would mean losing 1.367 points. In the other ten students, however, there was no significance that could lead to a generalization for the rest of the group that did not participate in the speaking tests.

In fact, this generalization becomes even more difficult taking into account the big fluctuations found between judges, which question the validity of this part of the experiment. Figure 3.6 showed an example of the huge differences of opinion between judges and within each of them, and Table 3.9a showed this by specifying the standard deviation for each student and each word. The average standard deviation between judges ranged from 1,263 to 1,888 points. Considering that there were only five judges and that they were using a 1 to 7 Likert scale, these values are too big. Also, the results in this part clash against the results discussed in the previous section and are not supported by any of the comments given by the teacher in either of the interviews. So, taking into account the lack of significance when comparing the evolution of the groups, the isolated significance when analyzing the individual learners, the big fluctuation between the opinion of the judges and the lack of support for these results by the qualitative research in the study, it is safe to say that there is not enough proof to confirm whether WBT affects, positively or negatively, the process of acquisition of the pronunciation of the language.

4.2 The impact of WBT in the motivation of the learners

The quantitative part of this experiment shows no proof that WBT affects the motivation of the learners. This fact can be seen both in the Mann Whitney U test (see Table 3.14) and the linear regression test (see Table 3.15). Neither show any significance ($<.05$), even though the latter shows a strong tendency (.052) in question 3 of the experimental group. In this case, the β factor is -.972, which means that the learners would value this question almost 1 point less every three months. Even though there is no significance in any of these tables, they do show some differentiated trends between the groups. These trends can be observed in Figure 3.8, where the control group shows two motivation questions with a positive slope and only one question with a descent. In contrast, all three questions in the experimental group show a negative slope. This would mean that the students using WBT presented a slight deterioration in their levels of motivation while the ones in the control group presented an improvement, even if the differences between the groups were not significant. Regarding the standard deviations (Table 3.13), the control group showed a general improvement between the pre-test and the post-test. Questions 1 and 2 had lower standard deviations in the post-test, while question 3 had higher ones. In the experimental group, on the other hand, all three questions got higher standard deviations in the post-test. This shows how the control group got to be a more homogenous group in terms of motivation, while the experimental group became more heterogeneous. Nevertheless, as will be explained below, these results are not reliable.

The results that were presented in the previous paragraph can only be partially trusted. The adaptation of the mini-AMTB, as commented in Section 3.1.3, showed many incongruences between the answers of the students in the pre-test and the

post-test, even though it had been piloted before the experiment to check its validity. The test also showed incongruences when it was re-piloted in a second school to check if they were due to factors in the application in the experimental school or due to the test itself. Both schools, as shown in Table 3.12, had very similar percentages of incongruences. This fact renders the results of the adaptation of the mini-AMTB unreliable and should not be fully trusted until replicated in more experiments that can prove whether the test is reliable or should be discarded.

The previous quantitative results are opposite to the impression that can be extracted from the interviews with the teacher, where she highlighted on many occasions how she perceived higher motivation degrees in the experimental group than in the control group. In the first interview, she made nineteen positive comments regarding the motivation in the experimental group against only a negative one. In the second interview she made nine positive comments and no negative ones. She explained at several points during the interview that the children really liked WBT. She expressed that she felt that the students in the experimental group were improving their motivation and that she could see this by how active and how much English they were using in the class. She also brought up the example of how all the students in the experimental group had started submitting their homework, but how some students in the control group had continued failing to submit it. She used words like 'game', which children often associate with activities that can be motivating for them. She also commented that children liked the tools and the techniques in WBT, even the ones she did not really like, such as the *Scoreboard*. The only negative comment she had regarding motivation was about this tool. She did not like to use it because she thought there were better ways to keep the students motivated. Even so, the students asked her to use it. In the interview, she was specifically asked if the change in motivation had

anything to do with the use of WBT and she explained that she though it was because of it.

The opinion of the teacher could be biased by the fact that she was part of an experiment testing WBT and the results of the adaptation of the mini-AMTB are not fully reliable, so it is not easy to say whether WBT had an impact on the motivation of the students. Nevertheless, some of the techniques that Dörnyei (2005) recommended as motivational strategies and that were commented on in Section 1.1.1, point to the fact that WBT should help improve the motivation of the learners. The teacher commented on more than one occasion that the children in the experimental group seemed to be more comfortable and participative and that they felt like they were playing a game. This would have to do with what Dörnyei refers to as a 'pleasant and supportive classroom atmosphere'. This game-like situation is also related to 'presenting tasks in a motivating way' and to 'making learning stimulating'. Dörnyei also mentioned 'appropriate group norms' and the teacher highlighted the use of both the five classroom rules and the different techniques and tools in WBT. The teacher insisted on how the pair work and the peer pressure were helping some learners be more active and participative in the class, and this would be related to 'promoting cooperation among the learners'.

Also, the results of the WBT scientific studies presented in Section 1.3.2.2 support the opinion that the teacher transmitted regarding the improvement of the motivation of the students in the experimental group. None of these studies are fully reliable because of not having been published in peer reviewed journals, but they all seem to point at the fact that the motivation of the learners is higher when using WBT. So, even though the quantitative part of this research suggests a decrease in the motivation of the learners that used WBT, the results in the qualitative part are suggesting the opposite. The literature is more in line with the opinions shared by the teacher, but

as these opinions do not match the results obtained in the quantitative research and might be biased, it cannot be said with certainty that WBT has helped improve the motivation of the learners. Furthermore, if the improvement of the motivation had been fully confirmed by this research, this improvement could have partially been explained by the idea of Dr. Sohamy (see Section 1.3.2.1) that WBT could probably be effective at the beginning because 'the brain learns when things are surprising and interesting' but that learners would gradually lose interest.

UNIVERSITAT ROVIRA I VIRGILI
ENGLISH AS A FOREIGN LANGUAGE THROUGH WHOLE BRAIN TEACHING
IN PRIMARY SCHOOL
Edward Alvar Lockhart Domeño

Chapter 5

Conclusion

5.1 Concluding remarks

This chapter will contrast the research hypotheses with the actual results provided in the previous chapters and will explore if they have been confirmed or refuted. It will also provide an overview of what implications these results have for further research related to the use of WBT for teaching English as a foreign language in primary education.

5.1.1 Research Hypothesis 1

The first hypothesis in this research stated that 'The use of the WBT method can enhance the process of language acquisition, allowing the learners to improve their language skills more significantly with the same amount of exposure and, thus, to have better results in listening and reading comprehension and written and oral production tests.'. Section 4.1.1 discussed the results obtained in the different language tests and speaking tests. The results showed that WBT had had a positive impact on the reading and writing tests (see Section 4.1.1.1). This was reflected in the Mann Whitney U

test in a significant difference between both groups in the post-test and was further corroborated and further explained by the teacher in the different interviews.

The results of the experiment also point at the fact that WBT positively affected the results of the listening tests (see Section 4.1.1.1). In this case, the Mann Whitney U test that explored the differences between both groups did not show any significant differences as in the reading and writing tests, but the experimental group did show higher average grades and lower standard deviations that hinted that the group had improved both in terms of performance and homogeneity. These improvements were confirmed by the linear regression tests, that showed a significant negative tendency in the time factor in the control group but no significance in the experimental group. The qualitative part of the experiment also hinted at the improvement of the experimental group due to the use of WBT, with the teacher highlighting positive aspects of the method that had helped the learners improve their results.

The previous results are further confirmed by the improvement in the active vocabulary of the students in the experimental group (see Section 4.1.2.1). Although both groups showed a similar improvement in terms of the β factor, the time factor was only significant in the experimental group. In the case of the control group, it showed only a tendency in this time factor. Again, the teacher explained in the interviews that she perceived an improvement in the experimental group that she had not perceived in the control one and she thought it was because of WBT. This improvement and the two improvements mentioned in the previous paragraphs are supported by previous research and publications that were explored in Chapter 1 and discussed in Chapter 4. The use of gestures and of sensory learning styles, the higher amount of interaction through pair-work or some of the twelve principles that Brown (2002) considered essential for a good English teaching have been proven part of teaching English as a foreign language through WBT with primary school students.

The only result that diverts from the initial hypothesis is the one obtained in the pronunciation part (see Section 4.1.2.2). In this case, the results show a negative effect in the pronunciation of the students in the experimental group. In contrast, the results in the control group show an improvement between the pre-test and the post-test. Neither of the results show any significance, though. In fact, significance can only be found when checking the linear regressions of the individual students. A student in the control group shows a positive significant tendency through time while a student in the experimental group shows a negative one. These significant results were isolated, because none of the other ten learners showed significance in their linear regressions and because it was a test that was only taken by a reduced sample from each group. Also, as was explored in the previous chapters, the big fluctuation between and within judges reported in the big standard deviations hint at the fact that the results in this pronunciation test are not very reliable. This lack of reliability is supported by the fact that the teacher did not transmit any comments that could point in the direction of WBT being negative for the pronunciation of the students or for any other language skill whatsoever.

To sum up, there seems to be evidence of an improvement of the results of the students using WBT in terms of reading and writing, listening and active vocabulary. This evidence is further supported both by the qualitative part of the study with the interviews with the teacher and by the previous theories by other experts in the field of English teaching. There are some weak results pointing at a deterioration of the students in the experimental group in terms of their pronunciation in English, but these results are not backed up by what the teacher explained in the interviews. The research hypothesis 1 has, thus, been confirmed by this experiment: the use of the WBT method can enhance the process of language acquisition, allowing the learners to improve their language skills more significantly with the same amount of exposure and,

thus, to have better results in listening and reading comprehension and written and oral production tests.

5.1.2 Research Hypothesis 2

The second hypothesis in this research stated that 'The use of the WBT method can improve the motivation of the learners. This hypothesized improvement could lead to a lowering of the affective filter, which is considered a major factor in language acquisition'. Section 4.2 discussed the results obtained in the motivation tests and contrasted them both with the opinion of the teacher and the existing literature. From that discussion there are several ideas to highlight. There seems to be a non-significant diminishment in the motivation of the learners using WBT, as seen in the linear regression tests done from the results of the adaptation of the mini-AMTB, but this diminishment is questionable because of the problems of reliability shown by this questionnaire.

Also, the opinion of the teacher is in opposition to this decrease in the motivation of the learners in the experimental group. She commented, on several occasions throughout both interviews, that she believed that those students were more motivated and she specified that she thought it was because of the application of the method. She reasoned that she could see this through the implication of the students and through how active the students in the experimental group were in comparison with the students in the control group. Nevertheless, this opinion could be biased by the fact that she was aware that she was part of an experiment trying to test the effect of WBT in the teaching of English as a foreign language. Even so, many comments by the teacher referring to the techniques of WBT or to the motivation of the learners in the experimental group were in line with the techniques that Dörnyei (2005), one of the world's experts in motivation, considers that every English teacher should apply to

improve the motivation of the learners. This could suggest that the opinions of the teacher might not be completely biased. Nevertheless, Dr. Sohamy, a neuroscientist at Columbia University, expressed that the motivation of the techniques applied in WBT might not be durable and might be due to the surprise factor of the new method.

To conclude, the results for the motivation of the students cannot be considered definitive. Although the results of the mini-AMTB seem negative regarding the use of WBT, the test has been proven to be untrustworthy and, furthermore, the results were not even significant. Also, the teacher transmitted the feeling of WBT being directly responsible for an improvement in the motivation of those students in the experimental group, and her opinion is supported by Dörnyei's theories. All these factors seem to point at the fact that the use of the WBT method can improve the motivation of the learners. This hypothesized improvement could lead to a lowering of the affective filter, which is considered a major factor in language acquisition, but the results are also not definitive in terms of real improvement of the motivation and the sustained effect of this enhancement in time, because this apparent improvement might only be an illusion due just to the novelty of the new method.

5.2 Limitations and implications for further research

The present study has opened many doors to further research that should be explored to shed some more light on the results explored in the previous pages and sections. One of the tests that should be redesigned would be the speaking test. It could be redesigned in several ways. The experiment could be done with other groups of students with a better initial command of the language so their fluency could be tested. Also, the technological tools used in this part of the experiment could be more adequate: the recordings could be done in a language laboratory and the judges could be asked to use professional equipment for evaluating the productions. Also, a bigger sample

and a greater number of judges would probably minimize the standard deviations and could offer more reliable results.

Regarding the adaptation of the mini-AMTB, it should be tested whether students were too young for this kind of questionnaire or if the phrasing of the objective questions was problematic. The adaptation could also be compared to the original mini-AMTB to check if the latter would have the same problems in a similar application. For this purpose, both tests could be piloted with a wider range of ages that could go from 9 (current experiment) to 15 (experiment from which the original mini-AMTB was taken), in a situation of pre-test and post-test to see if the incongruences appeared in one, both or neither of the tools and to check if there was a critical age for those incongruences to happen.

The research was limited in terms of length of the experiment and size of the samples. All the results described before could be more reliable if a more longitudinal study and with more groups was carried oud. This could involve several schools with two groups per grade and a whole year or two of one group using WBT while the other attended classes the traditional way. This would involve children of more ages and more backgrounds. The samples to compare would be bigger, so the results would be more robust and would allow to check if the motivation perceived by the teacher was something only temporary as suggested by Dr. Sohamy. This way, also, the results of the study could be more generalizable than the results from the current experiment, which has a more limited scope. This would also be supported by having more than one teacher, which could allow to test if the results could be different from the results obtained here. On the other hand, and given the results of this research, this quasi-experimental proposal would raise potential ethical concerns about the use of what seems to be a better methodology with only one of the groups and not with the other and, thus, benefitting only one of the groups.

References

Armijo, L. (2009). A case study on the impact whole brain teaching has on student achievement at San Jacinto Elementary, unpublished manuscript. Retreived from https://apuroecasestudy.wikispaces.com/file/view/Whole+Brain+Teaching+case+study.docx (Last visited: June 2016).

Asher, J. J. (2014). Some myths about TPR. Retrieved from http://www.tpr-world.com/mm5/TPRarticles/myths.html (Last visited: June 2016).

Bandler, R., & Grinder, J. (1979). *Frogs Into Princes: Neuro Linguistic Programming*. Moab, Utah: Real People Press.

Beaumont, J. G. (2008). *Introduction to Neuropsychology*. 2nd ed., New York: Guilford Press.

Bernaus, M., Wilson, A., & Gardner, R. C. (2009). Teachers' motivation, classroom strategy use, students' motivation and second language achievement. *The Modern Language Journal*, *92*(3), pp. 387–401.

Biffle, C. (2007). Teaching challenging elementary school students. Retrieved from http://www.wholebrainteaching.com/index.php/Download-document/2-Teaching-Challenging-Students.html (Last visited: June 2016).

Biffle, C. (2009a). Power teachers training manual. Retrieved from http://www.powerteachers.net/index.php?/Download-document/10-Power-Teachers-Training-Manual.html (Last visited: June 2016).

Biffle, C. (2009b). Teaching challenging teens. Retrieved from http://www.wholebrainteaching.com/index.php?option=com_docman&task=doc_download&gid=8&Itemid=223 (Last visited: June 2016).

Brown, H. D. (2002). English language teaching in the "post-method" era: toward better diagnosis, treatment, and assessment. In J. C. Richards, & W. A. Renandya, eds., *Methodology in Language Teaching: an Anthology of Current Practice*, book section 1, (pp. 9–19), New York [etc.]: Cambridge University Press.

Carr, W., & Kemmis, S. (1986). *Becoming Critical: Education Knowledge and Action Research*. London: Falmer Press.

Churches, R., & Terry, R. (2007). *NLP for Teachers: How to Be a Highly Effective Teacher*. Bancyfelin: Crown House.

Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matterscademic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, *19*, pp. 198–205.

De Jager, T. (2012). Professional development of beginner teachers: an action research approach to mentoring, retrieved from http://upetd.up.ac.za/thesis/available/etd-08132012-123829/ (Last visited: June 2016).

Dörnyei, Z. (2005). *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Doughty, C., & Long, M. (2003). *The Handbook of Second Language Acquisition*. Malden, Massachusetts: Blackwell Publishing Inc.

Ellis, R. (1985). *Understanding Second Language Acquisition*. Oxford ; New York: Oxford University Press.

Ellis, R. (1997). *SLA Research and Language Teaching*. Oxford: Oxford University Press.

Escola Joan Ardèvol (2016). Projecte de llengües estrangeres. Retrieved from http://joanardevol.org/info/?page_id=36 (Last visited: June 2016).

Felder, R. M., & Henriques, E. R. (1995). Learning and teaching styles in foreign and second language acquisition. *Foreign Language Annals*, *28*, pp. 21–31.

Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.

Gardner, H. (1999). *Intelligence Reframed: Multiple Intelligences for the 21st Century*. New York: Basic books.

Gass, S. (2008). *Second Language Acqusition: An Introductory Course*. New York: Taylor & Francis.

Gill, S. (2000). Against dogma: a plea for moderation. *International Association of Teachers of English as a Foreign Language Issues*, *154*.

Gregersen, T., Olivares-Cuhat, G., & Storm, J. (2009). An examination of L1 and L2 gesture use: What role does proficiency play? *The Modern Language Journal*, *93*(2), pp. 195–208.

Gullberg, M., & De Bot, K. (2010). *Gestures in Language Development*. Amsterdam: John Benjamins Pub. Co.

Higgins, J. (2012). Teachers learn ways to keep students' attention, but are brain claims valid? Retrieved from http://www.ohio.com/news/local-news/teachers-learn-ways-to-keep-students-attention-but-are-brain-claims-valid-1.319731 (Last visited: June 2016).

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. 1st ed., Oxford: Pergamon Press.

Krashen, S. D. (1988). *Second Language Acquisition and Second Language Learning*. New York: Prentice Hall.

Krashen, S. D., & Terrell, T. D. (1983). *The Natural Approach: Language Acquisition in the Classroom*. 1st ed., Oxford: Pergamon Press.

Lightbown, P., & Spada, N. M. (2006). *How Languages are Learned*. 3rd ed., Oxford: Oxford University Press.

Lockhart, E. A. (2009). Power teaching, an effective methodology?, unpublished Master's thesis.

McCafferty, S. G. (2002). Gesture and creating zones of proximal development for second language learning. *The Modern Language Journal*, *86*(2), pp. 192–203.

Meddings, L., & Thornbury, S. (2003). What DOGME feels like. *Humanizing Language Teaching Magazine*, retrieved from http://www.hltmag.co.uk/nov03/sart1.htm (Last visited: June 2016).

Mohamed, S. (2012). *Big Surprise! 4*. Madrid: Oxford University Press.

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and Methods in Language Teaching*. Cambridge: Cambridge University Press.

Skehan, P. (1989). *Individual Differences in Second-Language Learning*. London: Hodder Arnold.

Szott, M., & Molitoris, M. (2010). "Class, Class...Are you listening?" ..."Yes, Yes!!" How can Whole Brain Teaching impact our classroom environment?, unpublished manuscript.

Taylor, I., & Taylor, M. M. (1990). *Psycholinguistics: Learning and Using Language*. London: Prentice-Hall International.

Terrell, T. D. (1986). Acquisition in the natural approach: The binding/access framework. *The Modern Language Journal*, *70*(3), pp. 213–227.

Thornbury, S. (2000). A dogma for EFL. *International Association of Teachers of English as a Foreign Language Issues*, *153*, p. 2, retrieved from https://esol.britishcouncil.org/sites/default/files/attachments/informational-page/AdogmaforEFL.pdf (Last visited: June 2016).

White, L. (1987). Against comprehensible input: the input hypothesis and the development of second-language competence. *Applied Linguistics*, *8*(2), pp. 95–110.

Wirani, F., Setiyadi, A. B., & Hasan, H. (2014). The implementation of "power teaching" approach in increasing students' speaking participation. *U-JET Unila Journal of English Teaching*, 3(4), retrieved from http://jurnal.fkip.unila.ac.id/index.php/123/article/viewFile/4501/2770, (Last visited: August 2015).

UNIVERSITAT ROVIRA I VIRGILI
ENGLISH AS A FOREIGN LANGUAGE THROUGH WHOLE BRAIN TEACHING
IN PRIMARY SCHOOL
Edward Alvar Lockhart Domeño

Appendix A

Email with the instructions to the judges

Dear '*name of the judge*':

First of all, I want to thank you for your help. I know we all have busy lives, so I really appreciate that you decided to devote 1 hour of your day to help me with my PhD.

The task is pretty simple once you know what you have to do. I have created a tutorial and a trial page so when you go to the real test you feel confident with what you have to do.

(Embedded video)

If you cannot see the video, please click on this link: https://youtu.be/b_FPDe6L40M

Once you have watched the video you can proceed to the Trial Test. You will be able to repeat it as many times as you want. You will find it in the following link: http://goo.gl/forms/8keRdqKhKa

Once you are familiar with the procedure you have to follow, you can proceed to do the complete test. You will find the complete test here: http://goo.gl/forms/u4WM27EdsC

Thank you once again for your help.

Appendix B

Original mini-AMTB

Qüestionari de Llengua Anglesa

1. Quantes llengües, a part de català i castellà, parles i entens prou be ? _____

2. Els teus pares parlen alguna llengua estrangera? ____Sí ____No

3. Quants anys has estudiat anglès a l'escola? _____

   Quantes setmanes en un país de parla anglesa? _____

   Quants anys en una acadèmia o escola d'idiomes ? _____

L'objectiu d'aquest qüestionari és saber què penses d'alguns temes relacionats amb l'aprenentatge de l'anglès. A continuació tens una sèrie de frases seguides d'una escala on t'has de situar segons el que tu creguis més convenient

1. La meva motivació per aprendre anglès per tal de comunicar-me amb gent de parla anglesa és:

   POCA ____:____:____:____:____:____:____ MOLTA

2. La meva actitud envers la gent de parla anglesa és:

   DESFAVORABLE ____:____:____:____:____:____:____ FAVORABLE

3. El meu interès per les llengües estrangeres és:

   POC ____:____:____:____:____:____:____ MOLT

4. El meu desig d'aprendre anglès és:

   POC ____:____:____:____:____:____:____ MOLT

5. La meva actitud envers l'aprenentatge de l'anglès és:

   FAVORABLE ____:____:____:____:____:____:____ DESFAVORABLE

6. La meva actitud envers el meu professor/a d'anglès és:

   FAVORABLE ____:____:____:____:____:____:____ DESFAVORABLE

7. La meva motivació per aprendre anglès per raons pràctiques (per exemple, aconseguir una bona feina) és:

   POCA ____:____:____:____:____:____:____ MOLTA

8. Em neguiteja parlar en anglès fora de classe:

   POC ____:____:____:____:____:____:____ MOLT

9. La meva actitud envers la classe d'anglès és:

   DESFAVORABLE ____:____:____:____:____:____:____ FAVORABLE

10. Em neguiteja parlar en anglès a classe:

    POC ____:____:____:____:____:____:____ MOLT

11. La meva motivació per aprendre anglès és:

    POCA ____:____:____:____:____:____:____ MOLTA

12. Els meus pares m'animen a estudiar angles

    POC ____:____:____:____:____:____:____ MOLT

Appendix C

Adapted mini-AMTB

Qüestionari de llengua anglesa

1. Quantes llengües, a part del català i el castellà, parles i entens bé? _____

2. El pare o la mare parlen alguna llengua que no sigui català o castellà?
   a. Si _____ b. No _____

3. Quants cursos has estudiat anglès? _____

   Has estat mai a un país on només parlessin anglès? _____

   Has fet mai classes d'extraescolar d'anglès o a alguna acadèmia? _____

Ara contestaràs unes preguntes sobre l'aprenentatge de l'anglès. Hauràs de fer una creu al lloc que expliqui millor com et sents tu. Per exemple, si ets súper fan del Barça i la frase diu:
"M'agrada el Barça:"

POC ____:____:____:____:____:____:__X_ MOLT

Marcaràs la creu el més a prop del "Molt" que puguis, com a l'exemple. Si, en canvi, no t'agrada el Barça, marcaràs la creu el més a prop del "Poc"

1. Vull aprendre anglès per poder parlar amb gent que només parla anglès:

   POC ____:____:____:____:____:____:____ MOLT

2. La gent que només parla anglès:

   NO M'AGRADA ____:____:____:____:____:____:____ M'AGRADA

3. M'agraden les llengües estrangeres:

   POC ____:____:____:____:____:____:____ MOLT

4. Tinc ganes d'aprendre anglès:

   POC ____:____:____:____:____:____:____ MOLT

5. M'agrada aprendre anglès:

   POC ____:____:____:____:____:____:____ MOLT

6. M'agrada aprendre anglès amb el meu professor / la meva professora:

   POC ____:____:____:____:____:____:____ MOLT

7. Vull aprendre anglès per utilitzar-ho (per veure dibuixos en anglès, per exemple):

   POC ____:____:____:____:____:____:____ MOLT

8. Em fa vergonya parlar anglès fora de la classe:

   POC ____:____:____:____:____:____:____ MOLT

9. Les classes d'anglès m'agraden:

   POC ____:____:____:____:____:____:____ MOLT

10. Em fa vergonya parlar anglès a la classe:

    POC ____:____:____:____:____:____:____ MOLT

11. M'ho passo bé aprenent anglès:

    POC ____:____:____:____:____:____:____ MOLT

12. Els meus pares em diuen que és important que aprengui anglès:

    MAI ____:____:____:____:____:____:____ SEMPRE

Nom i cognoms: _____ Curs: _____

Appendix D

Analysis of the first interview with the teacher

1st interview

Time	Questions	Time	Positive comments	Time	Negative comments	Time	Other comments
00:30	General impressions?	00:35	They speak more	02:14	I don't like the scoreboard		
		00:55	More comfortable speaking in English	02:28	It's better if they do the things for themselves and not because I'm going to give them a happy face or a punishment		
		01:02	They used to ask things in Catalan, they are now using more English	02:55	The scoreboard is stressful for me. It forces me to pay a lot of attention to the time.		
		01:20	Subconsciously I try to apply techniques like Class! Yes! with the experimental group, because it's faster	03:30	You don't need to give them points, you can reinforce them by giving them nice comments every now and then.		
		01:40	They like it	04:55	The fact that I cannot name someone when this person is misbehaving.		
		02:10	When I forget the scoreboard, they remind me because they like it.				
02:22	Why don't you like the scoreboard?	04:35	Kids are very excited about the scoreboard, which is like a contest.				
03:18	The use of prizes and punishments, is it part of the education when we are children?	05:20	There are two "gipsy" girls that did not like English at all and did not participate. Now in the listening they have good marks. Also in the speaking part.				
04:50	You don't like the scoreboard. Is there anything else you don't like?	05:58	They participate in the "Teach! OK!". Now they are involved, because the partner, if not, it does not work. It's a good thing.				

Time	Questions	Time	Positive comments	Time	Negative comments	Time	Other comments
06:32	Is that because of the peer pressure of "Teach! Ok!?						
						07:00	There was a very low level student sitting next to a high level one, who was not participating. When he started sitting with a student with a lower level, he started participating. You have to find the right partner.
				08:10	There are moments when you have to go and let the student who is misbehaving know. A group was playing cards, so I went and confiscated them. It's always the same person.		
		09:15	There are three kids without books. They have to work on a notebook. While in the control group I have to insist on them to take the notebook, in the experimental group they now take it out automatically.				
09:30	Does this have something to do with the motivation of the students?						
		09:35	Yes.				
09:40	What do you think about the motivation of the students?						
		09:43	I think they are more motivated because they speak more English, the class is more dynamic. They lose the monotony. They are more active in the class.				
10:38	So, they're paying more attention?						
		10:40	They are paying more attention because they are more active.				
		10:50	They really understand what you are explaining, because of the movements or whatever…				
11:15	Have you noticed any changes between the groups?						
		11:18	I noticed that the experimental group are better than the other one.				
				11:24	The teacher dealing with oral production transmitted that she does not notice any changes, because the experimental group was already better than the control group.		

Time	Questions	Time	Positive comments	Time	Negative comments	Time	Other comments
		11:50	I have noticed that in the classes I have time to do more things.				
11:55	Even though you are making them teach each other?						
12:00	At the beginning you told me that "I'm behind the other group".						
		12:05	No, but now I can do more things.				
		12:45	I can see that in the tests, they are more prepared than the others.				
		12:52	At the end, they reach more.				
13:00	If you had to recommend this method to other English teachers, what would be the main points that you would highlight?						
		13:18	They remember the vocabulary and the structures better than in the other method. For example, the colors, or to put the adjectives before the name.				
		13:53	Usually in English, there is one student talking. Here, there are two talking at the same time, so they produce more.				
		14:35	Sometimes you have the strong student pressing their partner to do the task.				
		15:03	All of them are working, at least at the speaking part.				
		15:15	The speaking for them is like a game.				
		15:57	They can memorize the things better.				
		16:15	I like "Class! Yes!" a lot. When you say that, they stop doing what they are doing and you can work faster.				
		16:40	I think the difference between the high level students and the low level ones is going to be smaller with this method because here they all speak. I don't know if we will be able to see this in the exams, but I think it will happen.				
		17:04	No student is out of the class, at least in the speaking part. All of them are involved. Active.				

Time	Questions	Time	Positive comments	Time	Negative comments	Time	Other comments
17:43	Let's now focus on you. How do you feel in each group and how did you feel in them before?						
		18:02	This Monday, in the control group, there was a moment when I wanted to do an activity and I didn't have the techniques of WBT and I thought "it's unfair that the others are doing it and these ones aren't".				
18:40	Well, but it could be unfair for the experimental group and not for the control one, if in the end the results with WBT are worse…						
		18:47	No, but I don't think so.				
		19:14	In the experimental group I feel I have more control over the class.				
19:26	And before the method, did you have the same control?						
		19:30	The experimental group was better, but now the control in the class it's much more better (sic). Especially for the "Class! Yes!"				
		19:52	I think the students in the experimental group are more motivated.				
		21:18	All the students in the experimental group are presenting their homework, while in the control group there are a lot who are not. (9)				
20:52	Before the method, in the experimental group, were there also more people doing the homework than in the control group?						
		21:10	No, no, there were some not presenting their homework in the experimental group. Now, every day they present it.				
22:40	Before the experiment, you already had some knowledge about classroom management, has your idea about how to manage a class changed in any way? Do you believe things you didn't believe before you started the research project?						
				23:28	For example: the technique for delivering the books. They really like to be the "encarregats" (students in charge).		

Time	Questions	Time	Positive comments	Time	Negative comments	Time	Other comments
						24:10	I have realized that when you say a thing, the students have to do it. When I say class, you (sic) have to listen to me.
		25:08	When we finish at 4:30 we need to spend less time to pick up everything.				
25:30	All those things are about the learners. Have YOU felt any changes in yourself?						
		25:46	Yes. When I say x, they have to do it, and with this method, it works.				
26:00	So, do you think that with this method you have earned more control of the situation?						
		26:14	Yes.				
		26:17	I have seen that before I had to finish the class 10 minutes before and now, sometimes with 3 minutes before it's enough.				
27:10	So, as a teacher do you think that with this method you have won "real" class time?						
		27:15	Yeah. Because they follow the directions quickly (rule 1).				
		28:14	With this method you realize that everybody remembers content from previous sessions.				
		28:20	With this method, all the students get the main ideas that you, as a teacher, are trying to transmit.				
28:50	Why do you think this happens?						
		28:50	Because they are active in the class.				
		29:02	The explanations are shorter, using simple sentences. And they have to reproduce them.				
		29:20	The kids that were not listening can be more aware of this. Their partners remind them, and they don't like this so they make a bigger effort.				
		29:42	For them it's like a game, this method (sic).				
29:50	What's your opinion about the gestures?						
		29:58	I think they are very important.				

Time	Questions	Time	Positive comments	Time	Negative comments	Time	Other comments
30:45	Do you think you have improved in the use of gestures? Are you using them more in one class than in the other?						
						31:05	I'm using them more in the experimental group
		31:15	Now I use gestures always, even in stories (I didn't use to use them then).				
31:25	Do you think these gestures have anything to do with everybody understanding more?						
		31:30	Yes.				
31:48	Do you think that gestures and how active they are have any relation?						
		31:58	Yes, they are more active.				
		32:08	They like it. They started asking why they were doing it, but now, they like it.				
		32:15	When I say student 1, but it had to be student 2, they remind me. So, they are really participative in the method.				
33:04	Some say that the methods work when they are first introduced, because they are new. Then, after some time, they go back to the same results. Do you think something similar will happen here?						
		33:27	No. I think that the results will improve. Kids are involved in the teaching.				
		33:50	They are paying more attention.				
		34:08	If it does not keep improving, it will maintain a higher level than before.				
34:20	After the Easter Holidays, what will you do: not use the method in any of the groups? Continue using the method in the experimental group only? Use it in both groups?						
		34:40	I think I will use the method in both groups.				
		35:00	I'm already using it with other grades in Science and Catalan.				
35:16	Is it easier to use it in Science or in English?						

Time	Questions	Time	Positive comments	Time	Negative comments	Time	Other comments
						35:22	For me it's better to use it in "Medi" (Sciences). In English they repeat exactly what you say, but in science, they use their own words instead.
		36:12	In English it's very good too.				
36:25	Do you think that the training you received was enough? And the support you've been receiving?						
		36:50	Yes.				
				36:55	Finding the rules on the internet was difficult. It would have been nice if you sent them to us, although I didn't ask you to...		

Appendix E

Analysis of the second interview with the teacher

2nd interview

Time	Questions	Time	Positive comments	Time	Negative comments	Time	Other comments
00:10	After 3 months working with WBT, what could you say about it?					00:18	I think that it has good things and bad things.
		00:25	I have to recognize that it's possible that the students speak in the class having the whole group.				
		00:45	Now I don't need to do Teacher-Student-Teacher, but having them by couples, so they speak more.				
		01:23	Everyone is practicing at the same time.				
		01:42	They produce more because before it was one student at a time. Now, everybody is speaking at the same time.				
2:00	Any more good things?						
		02:10	The kids are more involved in the English class.				
		02:20	In the test, the students that didn't like English before, did the listening test very well.				
						02:49	Not the writing.
						03:20	They don't even pick up their books. They have no interest in picking them up.
03:30	Do you think that the students that were unmotivated before the method are now more motivated, at least, for the oral part?						
		3:39	Yes, they are more motivated.				
04:20	Has this happened in the other group?						
		04:32	The attitude from the beginning was very different. They can, but they don't.				
05:12	Do you think that the newly motivated students from the experimental group would have had that change if they had gone to the control group?						
		05:14	No, I don't think so. It's because of the method.				
05:20	Would you call this "motivation"?						

Time	Questions	Time	Positive comments	Time	Negative comments	Time	Other comments
		05:24	I don't think it's motivation. I think that it's because they are more active in the class. They like that.				
		05:50	Now they realize they can understand some things in English.				
05:18	Any other good things?						
		06:25	With the experimental group you have more control of the class.				
		06:40	The rules help you with the behavior.				
		06:54	In the experimental group, with the same amount of time, we finished the things before. We had like 2 extra classes.				
07:20	At the beginning you transmitted that the method was slower than not having the method. So, that has changed…						
		07:38	Yes. Once they realize what they have to do, they are faster.				
		08:27	I think you save some time. It's like 3 or 4 minutes every class.				
08:40	When you present an activity through WBT (Class! Yes!, Teach! OK!), do you think you have to repeat less or not?						
		09:00	Everyone gets it.				
09:15	And does that save you time?						
		09:19	Yeah. In the other group I have to go to the ones who were disconnected.				
		09:40	With WBT, as they have to explain each other, they realize what they have to do.				
		09:55	In this method, everyone knows what they have to do.				
10:05	What can you tell me about gestures?						
		10:20	Before, I used gestures a lot, but not as many as with WBT.				
				10:30	What was quite difficult for them was the mirror technique. I didn't use it much, though. Maybe it was my fault, because I didn't put enough effort into it.		
11:20	Do those gestures help them in any way?						

Time	Questions	Time	Positive comments	Time	Negative comments	Time	Other comments
		11:29	Yes. They help them memorize vocabulary.			11:39	When they explain each other, they are just repeating, not paraphrasing. This is a neutral thing.
12:40	Can you now mention some of the bad things?			12:48	The scoreboard, being for the whole group, it stigmatizes always that very same kid.		
				14:45	For me, the scoreboard was very stressful. I want the kids to do things because they want to, not because they get a punishment.		
				15:15	Some times I forgot about the scoreboard.		
				15:40	It's not relaxing. I have to keep it in mind all the time.		
				16:00	It's unfair to punish them without playground when they have the 5th and 6th graders next door shouting and really misbehaving.		
16:20	Any more negative things about the method?						
				16:30	There's too much activity. There's always something to do and you have to do it fast. I don't know if this is because of the method or because of you (Researcher).		
				17:00	There is no time for disconnection, not even between classes.		
				18:20	I think it's better to give them some time to disconnect every now and them, so they can settle their ideas.		
18:40	Any more negative things to highlight?						
				18:52	Kids come to the school to learn contents, but also to learn things (sic). I remember that once I stopped a story to work on feelings and you told me I should not have stopped because they could understand the story. I like using the empathy in the stories.		
		21:15	Another good thing is that I think the kids like this method. They are involved.				
				21:23	One bad thing is that the teacher has to be 100% there always.		

Time	Questions	Time	Positive comments	Time	Negative comments	Time	Other comments
				21:40	One problem that I have is that I keep mixing student 1 and 2.		
		22:00	But they don't forget				
						22:30	Sometimes they don't wait for the "Teach!".
						23:10	I never used rule number 5.
		23:15	I liked rule number 4 a lot and rule number 1, "follow directions quickly".				
		24:10	Make smart choices is showing them to think, and I like that.				
24:40	Any other good or bad things?						
		25:00	I will start using the method in the control group in the third term.				
				25:35	Switch may be good for longer explanations, but not for the ones we give in English.		
		26:18	I will use "Class", "Teach" and the first four rules.				
				26:25	I won't use the scoreboard, even though the kids love it.		
		26:30	They love the scoreboard.				
						26:35	It was difficult for them to celebrate the sad faces.
27:26	We don't have results yet, but, if the results were to say that there is no improvement, would you apply the method or would you go back to the other method?						
		27:55	No, I would continue using it. Not all of it, but those 4 things.				
		28:20	The kids are really active.				
		28:30	The not knowing what's going to happen and the peer pressure have made those passive students be active.				
		29:25	The student teacher I had liked the method. I don't know if he will use it or not, but I have seen him using Class and Teach.				
30:05	Do you think this method could be extended to other situations (other schools, other teachers…)?						

Time	Questions	Time	Positive comments	Time	Negative comments	Time	Other comments
				30:20	No. It depends on the teacher. It depends if the teacher wants to do it.		
31:10	What kind of training should teachers that want to do it receive? Would a 30-hour course be enough? Would they need supervision?					31:30	Training course plus supervision. At least a couple of sessions. One at the beginning and the other one some weeks later.
						32:00	Maybe you don't need a 30-hour course. Maybe 20 of training and 10 of supervision.