Julia-Athena Spinthourakis, University of Patras
Maria Stathopoulou, Hellenic Open University
Vasilios Zorbas, Hellenic Open University

Dina Tsagari, Oslo Metropolitan University
Marina Tzakosta, University of Crete
Kosmas Vlachos, University of Athens
## Table of Contents of Volume 12, Issue 1, 2022

### Editorial

Material

<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactional Corrective Feedback in beginner level classrooms of Greek as a second language: Teachers’ Practices</td>
<td>Sophia Ioannou &amp; Dina Tsagari</td>
<td>7</td>
</tr>
<tr>
<td>Recasts, metalinguistic feedback, anxiety and grammatical development in EFL: A case study with young learners</td>
<td>Evangelia Paraskeva &amp; Eleni Agathopoulou</td>
<td>26</td>
</tr>
<tr>
<td>Digging deeper: Investigating the emotional impact of online learning on university students during COVID19 in Greece</td>
<td>Thomaï Alexiou &amp; Theologia Michalopoulou</td>
<td>44</td>
</tr>
<tr>
<td>Developing Strategic Readers in the EFL Class</td>
<td>Maria Avgeri</td>
<td>56</td>
</tr>
<tr>
<td>Improving lexical access and acquisition through reading the news: case studies of senior high school students in China</td>
<td>Yuandi Zhang &amp; James Milton</td>
<td>56</td>
</tr>
<tr>
<td>Investigating the Parameters Accounting for Language Proficiency: the case of an English language test used for Academic purposes</td>
<td>Dina Tsagari and Theodosia Demetriou</td>
<td>89</td>
</tr>
<tr>
<td>Exploring teachers’ stances regarding the development of digital literacy in the 6th grade textbook of the Greek Primary school</td>
<td>Stergios Roumeliotis &amp; Alexandra Anastasiadou</td>
<td>111</td>
</tr>
<tr>
<td>VR and Social Identity: The utilisation of Virtual Reality Technology for the study of Social Identity in children with a migrant background situated in Greece</td>
<td>Kyriaki Thoidou &amp; Vasilios Zorbas</td>
<td>130</td>
</tr>
</tbody>
</table>
The battle of language learning apps: a cross-platform overview
Athanasios Karasimos

Using poetry to foster Critical Thinking and Metacognition in a Primary School EFL context
Aspasia Deliligka & Christine Calfoglou

Developing soft skills in adult learners of English in online courses: The case of Skyeng online school and Skyes digital platform
Alexey Konobeiev, Violetta Usanova & Olesya Gilmudinova

All articles in this Journal are published under the Creative Commons License Deed. Attribution 3.0 Unported (CC BY 3.0)
In this twelfth (regular) issue of RPLTL, the greater domain of foreign language teaching and learning is stronger (we did not get ‘COVID-away’ by the pandemic!), more popular than ever and rich in research studies. Through thick and thin, even amidst the pandemic, excellent research and practice is being carried out in classrooms and we have this opportunity through the journal to promote interventions and good practices. This issue sheds light on many different aspects of language teaching and learning, with reference to EFL vocabulary, individual differences, emotional factors, digital literacy and IT.

This volume is a collection of selected papers on EFL learning and teaching and it is divided into 2 sections. The first section consists of five papers that focus on language learners and deal with issues like feedback, emotional impact of online learning, strategy instruction and vocabulary uptake.

More specifically, Sophia Ioannou and Dina Tsagari explore interactional corrective feedback in adult beginners’ classrooms of Greek as a Second Language. Their study investigates the frequency with which teachers of Greek as a Second Language implement corrective feedback in their teaching as well as the factors affecting the implementation of corrective techniques underlining their role in teacher education.

Evangelia Paraskeva and Eleni Agathopoulou examine the relationship between oral corrective feedback (OCF) and language learning anxiety in the acquisition of the regular past tense morpheme -ed and the subject-verb agreement morpheme –s in EFL. The findings indicate that implicit OCF (recasts) may be more beneficial to high-anxiety learners, while explicit OCF (metalinguistic) may be more helpful for low-anxiety learners.

Thomaï Alexiou and Theologia Michalopoulou ‘dig’ a little deeper and investigate the emotional impact online learning had on university students during the Covid-19 pandemic in Greece. The results show that although students feel confident with their technological skills and the support
they had received in online learning, they also report stress, anxiety, frustration and loneliness, which can lead to low motivation.

Maria Avgeri studies strategy instruction and investigates whether it improves learners’ reading comprehension and develops their metacognitive awareness. The research involves explicit strategy instruction among two homogeneous 16-member groups of experimental and control learners in a Greek State High School. It appears that reading strategy instruction results in developing strategic readers able to utilize the strategies needed to facilitate understanding and monitor comprehension.

Yuandi Zhang and James Milton report on an empirical study concerning the effects of extensive reading on vocabulary uptake, and speed of word access, among of Chinese high school students using English as a second language. Watching the news on TV and the reading subtitles could result in significant improvement in both orthographic vocabulary size and speed of lexical access. The results show an increase in reading speed which is assumed to indicate an improvement in the speed of lexical access, but no obvious growth in vocabulary size.

The second section includes six papers that focus on learning outcomes and teaching tools, International exams for Academic purposes, digital literacy, virtual reality, language learning platforms, and the development of soft skills even poetry.

Dina Tsagari and Theodosia Demetriou explore the (content) validity in high-stakes international English for Academic Purposes exams. Various measures of proficiency were taken into account regarding writing and speaking. Results showed that vocabulary mainly accounts for language proficiency and can be used as a predictor variable for the writing and the overall scores in the test. Fluency could also predict some of the variability in the speaking scores.

Stergios Roumeliotis and Alexandra Anastasiadou investigate the promotion of digital literacy in the textbook “English 6th Grade” which is used in primary schools in Greece and explore whether this book should be revised to a certain extent. The findings indicate that the present textbook fails to adequately promote digital literacy. Suggestions are provided on possible changes in the specific book.

Kyriaki Thoidou and Vasilios Zorbas probe into the use of virtual reality in the discovery of children's social identity through semi-structured interviews, social location maps and identity texts. The findings showed that VR was assistive in the vivid, in-depth detection of their social identity (including symbolic representations of essential aspects) and underscored the need for their inclusion in society at large.

Athanasios Karasimos provides a comprehensive evaluation of the popular and rapidly growing online language learning platforms, such as Duolingo, Rosetta Stone, Memrise, LingQ and Busuu. It delineates features of these platforms and underlines their strengths and weak points while addressing certain limitations and providing suggestions for further improvement of the respective learning platforms.
Aspasia Deliligka and Christine Calfoglou show that the teaching of poetry can have a positive effect not only on young learners’ critical thinking and use of metacognitive reading strategies but on their reading comprehension as well. The results reveal that there is indeed a statistically confirmed advantage in the learners’ critical thinking, use of metacognitive reading strategies as well as reading comprehension.

Last but not least, Alexey Konobeiev, Violetta Usanova and Olesya Gilmudtinova discuss the importance of soft skills for adult learners and provide examples of soft skills as well as online courses aimed at developing soft skills through Skyeng online school and Skyes digital platform for universities offered in Russia and several other countries.

I would like to sincerely thank all colleagues but also the MA students for their significant contributions to this volume. Special thanks go to the assistant editors, Athanasios Karasimos and Vasilis Zorbas and to all the reviewers of the papers who serve as the International Advisory Board of the Journal.

Although this the second issue of RTLPT that I am editing - as the former issue specialised on young learners was also edited by me -, this is a rather moving time as I’m doing it from the position of Editor in Chief. Nicos Sifakis who has been the founder but also the ‘heart and soul’ of this Journal, has honoured me with this demanding task of ‘carrying the torch’. I will try my best, but he is always going to be the one who made all of this possible. It was due to his inspiration, diligence, hard work and devotion that the RPLTL has reached this stage. The number of visitors and downloads of papers, the citation of many papers in other publications around the world demonstrate only part of the success of the RPLTL journal and Nicos is the reason for it all. I thank him for granting me the opportunity to carry his work and I do hope that I will be worthy of this task.

Thomaï Alexiou
Editor-in-Chief
Interactional Corrective Feedback in beginner level classrooms of Greek as a second language: Teachers’ Practices

Sophia IOANNOU & Dina TSAGARI

This study explores interactional corrective feedback in adult beginners’ classrooms of Greek as a Second Language. More specifically, the study aims to investigate the frequency with which teachers of Greek as a Second Language implement corrective feedback in their teaching, as well as the factors that teachers take into consideration when making decisions on the implementation of their corrective techniques. The sample consists of five beginners’ classes (67 adult learners) and five teachers of the Modern Greek Language Teaching Center of the University of Athens. Three three-hour classes from each teacher were observed. Instances of learners’ errors and corrective feedback in oral interaction were transcribed, quantified, and statistically analyzed. The analysis focused on the frequency and distribution of oral corrective feedback types following learners’ errors. Results showed that teachers were inclined to correct a significant number of learners’ errors, while the communicative value of the error seemed to be a highly significant factor that affected the implementation of their teaching practices. As far as the type of corrective feedback teachers favoured, the findings indicated an overwhelming tendency for teachers to use recasts in response to learners’ errors. The paper concludes with a discussion of the centrality of the role of feedback and the importance of teacher education and training in the area.

Key words: Corrective feedback, Greek as a Second Language, teachers’ practices, corrective practices

1. Introduction

The role of corrective feedback (CF) in second/foreign language (L2) classrooms has been at the heart of second language research and pedagogy for the past three decades. Chaudron (1988) defines CF as “any teacher behaviour that attempts to inform the learner of the fact of error” (Chaudron, 1988, p. 150). A considerable number of studies have focused on the provision and efficacy of oral CF both in classroom
and laboratory settings (Mackey, 2020). However, most of the studies mainly focused on English as a second or foreign language. The provision of oral CF in the instructional setting of Greek as an L2 is still uncharted territory. This observational study aims at capturing teachers’ corrective practices during oral interaction in Greek as a second language in beginners’ classrooms.

Interactional feedback is a form of CF “that is generated in response to both linguistically erroneous and communicatively inappropriate utterances that learners produce during conversational interaction” (Nassaji, 2015, p.45). It has been proven that the interactional feedback that takes place during meaningful communication where learners favour form for meaning leads to L2 development and therefore assists L2 acquisition (e.g. Li, 2010; Mackey, 2006; Mackey & Goo, 2007; Nassaji & Kartchava, 2017; Russell & Spada, 2006). While there is strong empirical evidence in favour of the usefulness of CF, research now focuses on the factors that affect the effectiveness of oral CF. These factors are related to error and feedback characteristics, learner-related factors (e.g. proficiency, developmental readiness, individual differences), interlocutor factors (e.g. teacher’s background and experience), task characteristics, instructional context, and the organization of interaction in the classroom.

This study investigates the frequency of CF in beginner level classrooms of Greek as a second language. It also examines the factors that seem to affect teacher’s decisions to implement corrective techniques (e.g. factors that relate to learners’ errors, the context of the interaction and learners’ general performance according to their level). The study also investigates the characteristics of the feedback provided to the learners (i.e., teachers’ corrective types and the timing of CF).

2. A brief review of empirical studies

A considerable number of descriptive studies attempted to identify if and how CF occurs in oral interaction in different instructional settings, what patterns it takes and how students respond to it.

In their seminal article on oral CF in 1997, Lyster and Ranta identified six types of interactional feedback in French immersion classrooms. These CF types included recasts, metalinguistic feedback, explicit correction, elicitation, repetition, and clarification requests. Among the six CF types, recasts were the most frequent type (55%), followed by elicitation (14%), clarification requests (11%), metalinguistic feedback (8%), explicit correction (7%) and repetition (7%). Recasts, the most frequently used CF type, led to a lesser amount of learners’ uptake\(^1\) (31%) and repair (18%), while elicitation (46%), metalinguistic feedback (45%), explicit correction (36%), repetition (31%), and clarification requests (28%) led to more successful uptake. Recasts have been found to be the most frequently used CF type in different instructional settings and have been associated with low rates of successful uptake (Kamyia, 2014; Lyster, 1998; Panova & Lyster, 2002; Roothoft, 2014; Sheen, 2004).

Unlike in the previously mentioned studies, Ellis et al. (2001) found that recasts were the most frequent CF type in an adult ESL context in New Zealand and led to high levels of uptake (71.6%) and repair (76.3%). Differences were attributed to the instructional context. More specifically, the researchers assumed that learners enrolled in ESL programs (unlike learners enrolled in immersion programs) tend to focus on linguistic forms during oral interaction. The instructional setting and language structure typology proved to be important variables that affect both CF provision and efficacy. In a comparative analysis of CF in

---

1 The term *uptake* refers to the immediate learner response following CF (Chaudron, 1977). It is employed by researchers as a measure of feedback success.
French immersion classrooms in Canada and Japanese immersion classrooms in the USA, Lyster and Mori (2006) found that the differences between English and Japanese predisposed the learners of Japanese and their teachers toward a more form-focused orientation compared to the learners of French. It was also assumed that the teachers’ beliefs and behaviour in Japanese immersion classrooms, as shaped by their professional training and cultural background, led the recasts to be more beneficial compared to their effectiveness in the French immersion classroom.

It has been suggested that the effectiveness of recasts in L2 classrooms can be mediated through specific formal characteristics of recasts and the type of error in students’ utterances. Loewen and Philp (2006) found that recasts were the most frequent CF in their data (49.03%) and that teachers used them equally to treat morphosyntactic, lexical and phonological errors. They also found that recast characteristics associated with successful uptake and accuracy in a posttest were stress and declarative intonation. Regarding the target of CF, Mackey et al. (2000) found that, although recasts were mostly provided to morphosyntactic errors, learners did not perceive them as CF moves. Carpenter et al. (2006) found that morphosyntactic recasts were less accurately recognized than phonological or lexical recasts. Brown’s (2016) meta-analysis on the linguistic foci of CF, reports that grammar errors received the greatest proportion of CF, followed by lexical errors and phonological errors. Moderator analysis of feedback type and linguistic foci revealed that contextual factors, such as proficiency, teacher experience and instructional context, might influence CF across teaching context.

While there has been a vast amount of research on CF in the field of English or French as L2, there is an absence of focus on CF in less commonly spoken (and taught as L2) languages such as Greek. This study aims to contribute to the literature by identifying teacher’ practices in one more instructional context.

3. Research questions

To investigate teacher CF practices, three research questions were formed as follows:
1. What is the frequency of oral CF at beginner level classrooms?
2. Which factors affect teachers’ decisions regarding their corrective practices?
3. How is CF implemented at beginner level classrooms?

4. Research Methodology

An observational study was designed to answer the above research questions, as, according to Gass and Mackey (2007, p.165), observations “allow researchers to gather detailed data on the events, interactions, and patterns of language use within particular foreign and second language classrooms settings”.

4.1. Participants and Instructional Setting

The study took place at the Modern Greek Language Teaching Centre² of the National and Kapodistrian University of Athens where courses are divided into six levels (A1, A2, B1, B2, C1, C2) following the Common European Framework of Reference for Languages (Council of Europe, 2001). The courses offered are based on a communicative curriculum. The participants of this study attended an intensive beginners’ course in Greek that lasted for 8 months where students take Greek classes every day for three hours. Beginners start this course having only limited or no knowledge of Greek and they are expected to achieve

² Modern Greek Language Teaching Center website: https://en.greekcourses.uoa.gr, accessed 9/5/2021
an A2 level at the end of the course. The course also welcomes students who have basic communication skills but never participated in conventional Greek language classes.

The sample consisted of five beginners’ classes and five instructors. There were 12-15 students in each class. Students were 67 adult learners of Greek as an L2, 37 women, and 30 men whose age ranged from 18 to 42 years old, with different origins and first language (L1) backgrounds. Graph 1 presents students’ origin.

![Graph 1: Countries of students’ origin](image)

Most of the students enrolled in the course to improve their communicative competence. Their primary motivation was to communicate successfully in everyday life situations in Greece and speak with members of their Greek family and friends.

The five teachers (T) were all native speakers of Greek, and they had all completed the MA program in teaching Greek as a second or foreign language offered by the University of Athens. T2 and T5 had completed their MA studies in 2015 and 2017, while the rest had graduated from 1995 onwards. At the time of the research, T3 and T4 had a Ph.D. in teaching Greek as a second language, and T2 was a Ph.D. candidate researching aspects of Greek language teaching and learning. Among the five instructors, four were females, and one (T5) was male. Their teaching experience varied from 1 to 23 years. T1 had 15 years of experience, T2, had 4 years, T3 has been teaching Greek for 12 years, T4 had 23 years of experience while T5 had 1 year of experience. Their ages ranged from 27 to 48.

### 4.2. Data collection

The observations took place from the 5\(^{th}\) until the 7\(^{th}\) month (i.e., February until April) of the courses attended during the academic years 2016-2017 and 2017-2018. Permission by the institution and the teachers was requested to observe 4-5 classes that included speaking activities. Students were also asked to participate in the current research and signed a consent form. Before the data collection started, one or two pilot lessons per teacher were observed. During the pilot observations, all five teachers introduced their students and students were encouraged to introduce themselves and interact with the researcher (first author). This helped to establish good rapport with the students during data collection and to
increase the ecological validity of the study. During the observations, a non-participatory approach was employed where the researcher sat at the back of the class recording on an observation sheet (See Appendix I) students’ mistakes and teachers’ reactions, CF types, as well as the timing of feedback while at the same time using a digital voice recorder to record oral interaction. While the audio recorder captured only verbal interaction, detailed notes containing non-verbal interaction descriptions (e.g. facial expressions, gestures), students’ names and speech excerpts were kept during the observations to supplement the recordings. The observer recorded three 3-hour lessons from each classroom (15 lessons in total).

4.3. Coding

After the observations were completed, students’ errors and CF patterns during the interactional activities were delineated and transcribed using the NVivo10 software. Seven hours and thirty-five minutes (455 minutes) of oral interaction were transcribed, coded, and analyzed. A coding scheme was developed (see Appendix II, table 1) based on the error treatment sequence presented in the seminal article of Lyster and Ranta (1997) combining the coding schemes developed by Ellis and his colleagues (Ellis et al., 2001; Loewen, 2004) in order to examine the characteristics of learners’ errors and focus-on-form episodes. Based on the pilot classroom observations, the researchers added some characteristics that were assumed to be factors related to the provision of interactional CF.

Error classification

Errors were classified into three categories: morphosyntactic, lexical and phonological. Instances of L1 or sometimes of an intermediate language, usually English, were included, following the coding scheme of Lyster and Ranta (1997). We do not consider that these instances are necessarily errors per se. Since learners of Greek as an L2 often use English when they cannot express themselves in L2, we were interested in investigating teachers’ reactions when we thought that an ill-formed utterance indicated a gap in their interlanguage development. These instances usually include the choice of an English word to replace a linguistic item that was taught and, consequently, students were expected to have at least partial knowledge.

Source

This characteristic describes what could potentially cause the implementation of CF. A code-related episode results from the inaccurate use of linguistic forms that do not cause communication problems, while a meaning-related episode causes problems in communication.

Prior knowledge

---

3 According to Long (2015, p. 317), focus on form is one of the methodological principles of task-based language teaching that involves a reactive (triggered by a communication problem) use of pedagogic procedures to draw learners’ attention to linguistic problems in contexts, as they arise during communication. He claims that focus on form can create opportunities for the learners to synchronize the code features with their internal syllabus, developmental stage and processing ability.
This characteristic was included to examine if prior knowledge of a specific linguistic element is a significant factor that affects the implementation of corrective practices. We consider ‘previous knowledge’ all the linguistic elements that students were taught and thus expected to may have at least partial knowledge of, according to the syllabus and their instructors, as opposed to the elements that students had not been taught explicitly by the time the research was conducted. This category was included to explore teachers’ reactions to errors that seemed to appear constantly in students’ utterances during the pilot observation. Such errors were the use of wrong cases by the students (e.g. nominative instead of accusative), the inappropriate use of the definite/indefinite article and verb-subject or adjective noun agreement. These phenomena are taught explicitly and are emphasized in the Greek as a second language classroom. Students are repeatedly exposed to them in beginners’ course books (see Σιμόπουλος κ.ά, 2010), where explicit teaching takes place through grammar tables and grammar activities, as well as in oral interaction, when they participate in speaking activities that include these structures.

Relation to the lesson’s target

Errors were also classified according to their relevance to the linguistic target (topic-related vocabulary, grammar) of the lesson. This characteristic was included because we wanted to investigate whether teachers’ decisions to implement their corrective techniques were affected by the current focus of the lesson.

Context of oral interaction

The context of interaction was classified based on the observations. Meaningful oral interaction took place in the context of free discussions, role-playing games, and an oral presentation.

Student’s performance

Students’ general performance was measured through a placement test that consisted of five parts: reading, listening, speaking, use of language, and writing. Students were divided into high-performance and low-performance within the beginners’ A1 level.

Teacher’s response to the error

This characteristic examines whether the teacher decided to implement CF or not.

CF types

Seven verbal CF types were identified in previous observational studies (Nassaji 2015).

1. **Explicit correction** refers to the teacher’s utterances that both rephrase the erroneous utterance into a correct form and clearly indicate to the learner that his/her utterance is incorrect (Nassaji, 2015: 52).

2. **Recasts** consist of the teacher’s reformulation of all or part of a learner’s utterance minus the error (Lyster & Ranta, 1997: 46)

3. **Metalinguistic feedback** is the feedback that includes metalinguistic information in combination with the correction.
4. **Metalinguistic cue** refers to metalinguistic information provided to the learner in the form of hints, comments about language, rules, or questions about the grammaticality of the learner’s utterance without providing the correct form.

5. **Clarification request** refers to the feedback that occurs when the teacher does not/pretends not to fully understand the learner’s utterance and asks for clarification (e.g., “Pardon me?”, “I am sorry?” “excuse me?” etc.).

6. **Repetition** refers to feedback that repeats all or part of the learner’s erroneous utterances with a rising intonation, without providing the correct form.

7. **Elicitation** is the feedback that attempts more overtly to push the learners to provide the correct form. This can happen when teachers elicit completion of their own utterances by pausing and waiting for the learner to fill in the blank, by posing questions (e.g. “how do we say this in Greek?”), or asking the learner to reproduce the utterance.

**Type of CF response**

CF moves are divided into two categories based on the way that the corrections is made: input-providing strategies and output-prompting strategies (Lyster 2002). Input-providing strategies (also known as reformulations), refer to the teacher’s utterance that repeats what the learner has said correctly. Input-providing strategies include recasts and explicit correction. On the other hand, output-prompting strategies (also known as prompts) “push” the students to be more accurate in their output. According to Lyster (2004), prompts include the following CF types: clarification requests, repetitions, metalinguistic feedback, and elicitation.

**Timing of feedback**

Interactional feedback could be immediate as it happens in reaction to an error, or it can take place in the form of delayed feedback. In the latter case, the teacher may listen carefully to the students while they speak, record their erroneous utterances, and then address the errors interactionally after the activity (Nassaji, 2007).

**4.4. Data analysis**

After the coding of the data, it was subjected to statistical analysis. In order to examine the factors that affect teachers’ decisions to implement their corrective practices, the characteristics of the error (e.g., error classification, source, prior knowledge, relation to the lesson), the context of the interaction, and students’ performance were correlated with the teachers’ responses. Furthermore, the variables of each category (e.g., teachers’ responses to the error, CF type, type of CF response, and the timing of feedback) were compared in order to discover teachers’ preferences regarding the CF types. Statistics were calculated using the Statistical Package for the Social Sciences (SPSS) 21.0. The values of the variables are presented in frequencies (F) and percentages (%). A chi-square test or Fisher’s exact test was performed in order to investigate the association between the variables. A chi-square goodness of fit was performed in order to examine the correlation between the categories of each variable. All the tests are two-sided and p-value <0.05 was set to indicate statistical significance.

**5. Findings**

**5.1. The frequency of oral CF in beginners’ classrooms**
Oral interaction in beginners’ classrooms consisted of eight role-plays, free discussions on similar topics such as holidays and a presentation on students’ dream houses. The exact content of the lessons is presented in Appendix II (see Table 3).

Table 1 shows the frequency of students’ errors (F=990) and the occurrence of teachers’ CF in each classroom and in total (74.8%). Results revealed that teachers provided feedback to a statistically significant number of errors ($\chi^2 = 245.0, df=1, p<0.0001$).

<table>
<thead>
<tr>
<th>Errors</th>
<th>CF</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency (F)</td>
<td>Frequency (F)</td>
<td>Percentage (%)</td>
</tr>
<tr>
<td>T1</td>
<td>209</td>
<td>183</td>
</tr>
<tr>
<td>T2</td>
<td>186</td>
<td>124</td>
</tr>
<tr>
<td>T3</td>
<td>137</td>
<td>95</td>
</tr>
<tr>
<td>T4</td>
<td>329</td>
<td>256</td>
</tr>
<tr>
<td>T5</td>
<td>129</td>
<td>83</td>
</tr>
<tr>
<td>Total</td>
<td>990</td>
<td>741</td>
</tr>
</tbody>
</table>

Table 1: Frequency of errors and CF

5.2. The distribution of oral CF practices

This section reports the descriptive statistics of the characteristics of students’ errors and feedback episodes as well as the relationship between them and teachers’ responses.

<table>
<thead>
<tr>
<th>Errors</th>
<th>CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency (F)</td>
<td>Percentage (%)</td>
</tr>
<tr>
<td>Morphosyntax</td>
<td>675</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>145</td>
</tr>
<tr>
<td>Phonology</td>
<td>71</td>
</tr>
<tr>
<td>Use of L1</td>
<td>99</td>
</tr>
</tbody>
</table>

Table 2: Distribution of CF to students’ errors

Table 3 shows the occurrence of CF according to the source of the error. We can see that most students’ erroneous utterances were code-related (77.3%), while 22.7% of the errors caused communication problems. Teachers corrected 87.6% of the errors that caused communication problems and 71.2% of the code-related erroneous utterances. Fisher’s exact tests revealed statistically significant differences between the two variables (Fisher’s exact test, p<0.001).
Table 3: Source of the error and occurrence of CF

Table 4 illustrates the occurrence of CF according to students’ prior knowledge of the erroneous form. It seems that teachers provided feedback more frequently when students’ erroneous utterances included a previously taught form (75.7%) compared to new forms (67.1%). Statistical analysis showed marginal significance between the two categories (Fisher’s exact test, p = 0.080).

Table 4: Prior knowledge of the erroneous form and CF

Table 5 presents the occurrence of CF according to the relevance of the type of the error to the lesson target. Most students’ errors were irrelevant to the lesson target (59.2%), while teachers’ reaction did not differ. No statistical significance was found between the two variables (Fisher’s exact test, p = 0.503).

Table 5: Lesson target and occurrence of CF

Table 6 shows the distribution of feedback according to the context of interaction. Most of student’s mistakes were found during role-plays (59.5%) while teachers’ reactions did not seem to differ as they provided CF to 73.9% of the errors that occurred in free discussions, 75.2% to the errors that occurred during role-plays and 75.5% of the errors in the context of presentations (χ² = 0.299, df=2, p = 0.905).

Table 6: The context of oral interaction and CF
Table 7 shows the distribution of CF according to students’ performance. Most students’ mistakes were made by low-performance students. Teachers provided feedback to 74.3% of low-performance students to 75.5% of high-performance students. No statistical significance was found between the two variables (Fisher’s exact test, p= 0.709).

<table>
<thead>
<tr>
<th>Errors</th>
<th>CF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency (F)</td>
</tr>
<tr>
<td>Low-performance students</td>
<td>589</td>
</tr>
<tr>
<td>High-performance students</td>
<td>401</td>
</tr>
</tbody>
</table>

Table 7: Students’ performance and CF.

5.3. Teachers’ corrective practices

This section reports the descriptive statistics of the teachers’ corrective practices. Table 8 presents the types of CF responses. 81% of teachers’ responses were input-provided, while 19% was output-prompted. Chi-square analysis revealed statistical significance between the two variables ($\chi^2 = 284.32$, df=1, p< 0.0001).

<table>
<thead>
<tr>
<th>CF</th>
<th>Frequency (F)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input-provided</td>
<td>600</td>
<td>81</td>
</tr>
<tr>
<td>Output-prompted</td>
<td>141</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 8: Types of CF responses

Table 9 shows the oral CF strategies used during oral interaction. In total, there were 923 instances of provision of CF by the teachers. The most frequent type of CF used in beginners’ classroom were recasts (66.5%), followed by metalinguistic cues (11.5%) while only small percentages of the rest of the strategies were found.

<table>
<thead>
<tr>
<th>CF Strategies</th>
<th>Frequency (F)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recasts</td>
<td>614</td>
<td>66.5</td>
</tr>
<tr>
<td>Metalinguistic cue</td>
<td>106</td>
<td>11.5</td>
</tr>
<tr>
<td>Metalinguistic feedback</td>
<td>52</td>
<td>5.6</td>
</tr>
<tr>
<td>Elicitation</td>
<td>61</td>
<td>6.6</td>
</tr>
<tr>
<td>Repetition</td>
<td>53</td>
<td>5.8</td>
</tr>
<tr>
<td>Explicit feedback</td>
<td>26</td>
<td>2.8</td>
</tr>
<tr>
<td>Clarification requests</td>
<td>11</td>
<td>1.2</td>
</tr>
<tr>
<td>Total</td>
<td>923</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 9: CF strategies

Table 3 (see in Appendix II) and graph 2 provide a breakdown by teacher. Before we proceeded to the breakdown, the categories of ‘metalinguistic cue’ and ‘metalinguistic feedback’ were merged for practical reasons, since both provide feedback with metalinguistic information. We observe teachers’
overwhelming preference for recasts. Especially T1, T3 and T4 used a significant number of recasts (73.3%, 84.11%, 75.4%, respectively). The second commonly used strategy was metalinguistic feedback used by four teachers except T2. T2’s second more frequent strategy was repetition. We can see that while T1, T3 and T4 used only recasts and metalinguistic feedback, T2 and T5 used a range of strategies, including repetition and elicitation.

Graph 2: CF strategies implemented by each teacher

Tables 10 and 11 illustrate the timing of CF. It seems that all five teachers preferred immediate feedback (91.8%) while only a small percentage of feedback was delayed (8.2%). The difference between the two variables is statistically significant ($\chi^2 = 517.08$, df=1, $p<0.0001$).
To summarize and discuss the results, we will revisit our research questions. The first research question sought to explore the frequency of oral CF in beginner level classrooms of Greek as an L2. We found that teachers provided feedback to 74.8% of students’ erroneous utterances. It seems that the number of CF moves in our study is significantly higher compared to the amount of feedback occurred in descriptive studies in immersion classrooms (Lyster, 1998; Lyster & Mori, 2006), ESL classrooms (Panova & Lyster, 2002) and EFL classrooms (Roothoft, 2014). The range of CF in these studies is 20.43% (Kamiya, 2014) to 61% (Lyster & Mori, 2006). Findings can be attributed to the instructional setting in combination with the typology of Modern Greek, the language teaching methodologies used by the teachers of Greek as an L2, as well as the teachers’ background. Unlike the young learners of immersion classrooms who were engaged in content-based instruction, the participants of this study were adult learners, highly motivated to improve their communicative competence in order to interact with Greeks in Greece. Thus, learners in this study were more motivated and more cognitively able to attend to form than the participants in Lyster’s studies (Lyster, 1998; Lyster & Ranta, 1997). These findings are in line with the studies of Ellis et al. (2001) and Lyster and Mori (2006).

Based on the significant amount of CF provided by the teachers, we can infer that CF has a central role in Greek language instruction. The centrality of the role of CF could be attributed to Greek language typology in combination with the language teaching methodologies used by the teachers. Although teachers did not draw on a structural syllabus in any apparent way during the observations, it was found that during the instruction, teachers combined form-focused instruction with more analytic teaching strategies (e.g. drills, use of metalanguage). CF seems to constitute an interface that brings together the communicative syllabus with analytical teaching strategies. Hence, we can assume that interactional CF is congruent with teachers’ approaches to language teaching. Another factor that is possibly related to CF frequency is teachers’ educational and cultural background. The teachers were all native speakers of Greek and had graduated from various instructional environments in Greece where they had received traditional, explicit grammar instruction with a major emphasis on accuracy. We can infer that the teachers’ background predisposes them towards a more form-focused orientation (Lyster and Mori, 2006).

### Table 11: The timing of CF by each teacher

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Immediate CF</th>
<th>Delayed CF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency (F)</td>
<td>Percentage (%)</td>
</tr>
<tr>
<td>T1</td>
<td>183</td>
<td>100</td>
</tr>
<tr>
<td>T2</td>
<td>69</td>
<td>57</td>
</tr>
<tr>
<td>T3</td>
<td>95</td>
<td>100</td>
</tr>
<tr>
<td>T4</td>
<td>251</td>
<td>98</td>
</tr>
<tr>
<td>T5</td>
<td>82</td>
<td>98.8</td>
</tr>
</tbody>
</table>

4 Greek is an inflectional language and consequently demanding for learners coming from L1 backgrounds with a different typology.

5 For more information about Grammar Instruction in Greek as an L2 see Ιακώβου & Μαγγανά (2014).
rather than a meaning-focused orientation that allows errors if they do not cause communication problems (Richards, 2006).

The second research question investigated the factors that affect teachers’ decisions to implement their corrective strategies. Results revealed that one of the main factors that influenced teachers’ provision of CF characteristics set by the researchers was the source of the problem in students’ erroneous utterance. It seemed that teachers prioritized mistakes that caused communication problems. This factor was expected to be significant given the communicative focus of the lessons. In addition, it appears that teachers provided feedback to an essential number of accuracy errors. When looking at students’ errors, one can observe that the real number of students’ morphosyntactic errors (F=745) is significantly higher compared to lexical (F=145) and phonological errors (F=71). Moreover, teachers provided CF 495 times on morphosyntactic errors and 114 times on vocabulary errors. The frequency of teachers’ CF on morphosyntactic errors can be associated with the fact that Greek grammatical structures include both grammatical and semantic information. For example, since the subject in Greek is denoted in the ending of verbs, the erroneous use of a verb ending can cause problems in communication. Thus, although grammatical errors are seemingly associated with problems in learners’ accuracy, morphosyntactic errors can also cause communication breakdowns that lead to teachers’ use of CF.

Regarding the type of the students’ error, the use of intermediate language triggered significantly more CF moves presumably for the teachers to sustain interaction in Greek and to provide their students with the equivalent vocabulary in Greek. The fact that teachers provided equal percentages of feedback to morphosyntactic and lexical errors possibly shows teachers’ intention to focus both on students’ fluency and accuracy. As for the phonological errors, based on teachers’ CF frequency, we can assume that teachers did not consider them to be important, especially when they do not impede communication. This fact raises questions regarding students’ pronunciation development, as constant errors without any teacher feedback may lead to fossilization while CF scholars claim that pronunciation-focused CF can be a crucial component of pronunciation development (Saito, 2021).

The third research question investigated how CF was implemented in beginner level classrooms. Results demonstrated teachers’ preference for recasts. It seems that teachers prefer recasts, as they serve to provide CF in an unobtrusive way and without interrupting the flow of communication. The results are in line with the findings of most descriptive studies mentioned in Section 2. Furthermore, research on teachers’ beliefs (Kamiya, 2014) has shown that teachers tend to use recasts, as they are afraid that if they use more direct strategies (e.g., elicitation) their students would feel bad and they may be discouraged to participate in oral interaction. On the other hand, researchers (Ellis et al., 2001) suggest that teachers use a variety of CF strategies, as their effectiveness can differ according to other variables (e.g., proficiency level, instructional settings). Moreover, experimental studies showed that output-prompting strategies lead to better learning outcomes. More specifically, output-prompting strategies were found to be particularly more effective for beginners or low-performance students as well as when correcting morphosyntactic errors (Nassaji, 2015). Thus, the results of the current study reveal a gap between teachers’ practices and the findings as well as suggestions made on the basis of previous research focusing on oral CF.

Results also revealed that some teachers (e.g., T2 and T5) used a range of CF strategies while other teachers used mostly recasts. This probably stems from teachers’ professional training. It seems that teachers’ professional education and training is a factor that might have an important influence on how teachers implement (or not) CF in the language classroom. Specifically, T2 and T5 graduated from the same MA program relatively recently, as opposed to the rest of the
teachers, who graduated more than ten or twenty years ago. T2 and T5 participated in the course “Practice in classroom – teaching technics”. During this course, teachers were explicitly introduced to CF techniques, observed Greek language lessons as non-participant observers, and they had to fill out an observation sheet (Ιακώβου, 2015) that included CF strategies. After the lessons, students reviewed, reflected on the observations, and discussed aspects of oral interaction with their tutor. For the rest of the teachers, we can assume that they might have been unaware of the range of techniques they had at their disposal or underestimated the effectiveness of the rest of CF strategies.

As for the timing of CF, it seems that four teachers provided immediate feedback while only one provided both immediate and delayed feedback. The instructor’s choice probably stems from her attempt to distinguish the goal of the oral interaction (accuracy vs. fluency). Some teachers’ guides (e.g. Scrivener, 2011) suggest this practical technique for the teachers to balance students’ accuracy and fluency during the instruction. Whilst there is not enough evidence to prove which form of CF is more appropriate or when, researchers so far ensure that immediate feedback is associated with learning (Nassaji & Kartchava, 2017).

7. Conclusion

To the best knowledge of the researchers, this study constitutes the first attempt to capture interactional CF in adult classrooms of Greek as a second language. The frequency of oral CF in Greek language classrooms indicates its prominent role in second language teaching in one more instructional setting. In this instructional context, the provision of CF is linked to the approaches to teaching grammar, teachers’ educational background and professional training, and the course's orientation. Findings underline the importance of teacher trainings and the need to link the gap between second language research and pedagogy. Teacher development and education programmes offered on an undergraduate, postgraduate and in-service level for student and experienced language teachers need to prioritise the role of feedback and guide and train teachers to offer constructive and facilitative types of feedback that lead to successful language learning. Hopefully, the current study design and findings can be used as an impetus for follow-up studies in the use of CF and other types of feedback in the teaching of more languages used as L2 and other contexts.

References


Ιακώβου, Μ. (2015). Παραδόσεις μαθήματος Πρακτική Άσκηση στην Τάξη, Τεχνικές Διδασκαλίας.


### Appendix I

**Observation scheme:**

<table>
<thead>
<tr>
<th>Time&lt;sup&gt;6&lt;/sup&gt;</th>
<th>Students’ erroneous utterances</th>
<th>Type of error</th>
<th>Corrective feedback</th>
<th>Uptake</th>
<th>Comments (e.g. paralinguistic and extralinguistic cues)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

<sup>6</sup> On the digital voice recorder.
## Appendix II

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Definition</th>
<th>Categories</th>
</tr>
</thead>
</table>
| General error classification                 | Error classification according to the language aspect                      | 1. Morphosyntax  
2. Vocabulary  
3. Pronunciation  
4. L1 or intermediate language |
| Source                                        | What causes the feedback episode                                          | 1. Code: inaccurate use of language forms without apparent communication problems  
2. Message: problem in understanding the meaning |
| Prior knowledge                              | If this form was previously taught                                        | 1. Previously taught form  
2. New form |
| Relation to the lesson’s target              | If the error is related to the lesson’s target structure(s)               | 1. Relevant to the lesson’s target  
2. Irrelevant to the lesson’s target |
| Context of oral interaction                  | What was the context in which the error was made                           | 1. Free conversation  
2. Role-plays  
3. Presentations |
| Student’s performance                        | Student’s performance based on a proficiency test                          | 1. High-performance students  
2. Low-performance students |
| Teacher’s response to the error              | Teacher’s choice to intervene and provide feedback                         | 1. Provision of oral CF  
2. Topic continuation |
| CF strategies                                | What oral CF strategies were implemented                                  | 1. Explicit correction  
2. Recasts  
3. Metalinguistic feedback  
4. Metalinguistic cue  
5. Clarification request  
6. Repetition  
7. Elicitation |
| Type of CF response                          | How CF is provided                                                          | 1. Input-provided  
2. Output-prompted |
| Timing of feedback                           | When teacher’s response occurs                                              | 1. Immediate feedback  
2. Delayed feedback |

Table 1: Characteristics of students’ errors and focus-on-form episodes
<table>
<thead>
<tr>
<th>Instructor</th>
<th>Content of oral interaction</th>
<th>Duration</th>
</tr>
</thead>
</table>
| T1         | • Discussion “Christmas Holidays”  
|            | • Role-play “Arrange to go out”  
|            | • Discussion “Summer holidays”  
|            | • Role-play “At the police station: I had my bag stolen!” | 95 minutes |
| T2         | • Presentation “My dream house”  
|            | • Discussion “The story of my life”  
|            | • Role-play “The story of my life. Interviews from imaginary characters” | 120 minutes |
| T3         | • Discussion “Holidays”  
|            | • Discussion “Family” | 80 minutes |
| T4         | • Role-play “Going to the shrink”  
|            | • Role-play “Going to the fortune-teller”  
|            | • Role-play “Interviews of imaginary characters” | 110 minutes |
| T5         | • Discussion “Holidays”  
|            | • Role-play “Arranging holidays”  
|            | • Role-play “Going to the pharmacy” | 50 minutes |

Table 2: The content of oral interaction
Table 3: Distribution of feedback types

<table>
<thead>
<tr>
<th>Teacher</th>
<th>CF</th>
<th>Recasts</th>
<th>Metalinguistic feedback</th>
<th>Elicitation</th>
<th>Repetition</th>
<th>Explicit feedback</th>
<th>Clarification request</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>221</td>
<td>162 (73.30%)</td>
<td>37 (16.74%)</td>
<td>16 (7.24%)</td>
<td>2 (0.90%)</td>
<td>2 (0.90%)</td>
<td>2 (0.90%)</td>
</tr>
<tr>
<td>T2</td>
<td>173</td>
<td>68 (39.30%)</td>
<td>33 (19.07%)</td>
<td>13 (7.51%)</td>
<td>39 (22.54%)</td>
<td>5 (2.90%)</td>
<td>15 (8.67%)</td>
</tr>
<tr>
<td>T3</td>
<td>107</td>
<td>90 (84.11%)</td>
<td>11 (10.27%)</td>
<td>1 (0.93%)</td>
<td>-</td>
<td>2 (1.87%)</td>
<td>3 (2.80%)</td>
</tr>
<tr>
<td>T4</td>
<td>305</td>
<td>230 (75.40%)</td>
<td>49 (16.06%)</td>
<td>20 (0.66%)</td>
<td>-</td>
<td>1 (0.33%)</td>
<td>5 (1.64%)</td>
</tr>
<tr>
<td>T5</td>
<td>117</td>
<td>64 (54.70%)</td>
<td>28 (23.93%)</td>
<td>11 (9.4%)</td>
<td>12 (10.26%)</td>
<td>1 (0.85%)</td>
<td>1 (0.85%)</td>
</tr>
</tbody>
</table>

Sophia Ioannou ([ioannou.k.sophia@ucy.ac.cy](mailto:ioannou.k.sophia@ucy.ac.cy)), PhD, is a Special Scientist at the School of Modern Greek, University of Cyprus, where she teaches Greek as a Second language. She holds a MA and a Ph.D. in Linguistics from the University of Athens. Her doctoral dissertation is on interactional Corrective Feedback in Greek as a Second language beginners’ classrooms. Her research interests focus on second language acquisition, language assessment, interactional feedback, form-focused instruction and task-based teaching and learning.

Dina Tsagari ([dintsa@oslomet.no](mailto:dintsa@oslomet.no)), PhD, is Professor, Department of Primary and Secondary Teacher Education, Oslo Metropolitan University, Norway. She has also worked for the University of Cyprus, Greek Open University and Polytechnic University of Hong Kong. Her research interests include language testing and assessment, materials design and evaluation, differentiated instruction, multilingualism, distance education and learning difficulties. She is the editor and author of numerous books, journal papers, book chapters, project reports etc. She coordinates research groups, e.g. CBLA SIG – EALTA, EnA OsloMet and is involved in EU-funded and other research projects (e.g. NORHED, KrIT, DINGLE, TRIBES, ENRICH, TALE, DysTEFL, PALM, etc)
Recasts, metalinguistic feedback, anxiety and grammatical development in EFL: A case study with young learners

Evangelia PARASKEVA & Eleni AGATHOPOULOU

This case study investigated the relationship between oral corrective feedback (OCF) and language learning anxiety in the acquisition of the regular past tense morpheme -ed and the subject-verb agreement morpheme –s in EFL. The participants were two Greek L1 eleven-year-old female twins at an A2 level of English language proficiency. A language anxiety test and interview data showed that one of the twins was a low-anxiety learner while the other was a high-anxiety learner. The teaching interventions lasted for two weeks, providing corrective feedback on the target structures during 8 oral story-retelling tasks. Results from pre-tests, immediate post-tests and delayed post-tests (which contained an untimed grammaticality judgment test and an oral imitation test) showed that the low-anxiety learner benefited from metalinguistic feedback whereas the high-anxiety learner benefited from recasts in the acquisition of subject-verb agreement. On the other hand, there was no statistically significant improvement in the acquisition of the past tense morpheme –ed. Our findings comply with previous studies which indicate that implicit OCF (recasts) may be more beneficial to high-anxiety learners, while explicit OCF (metalinguistic) may be more helpful for low-anxiety learners.

Key words: Oral Corrective Feedback, language anxiety, young EFL learners, recasts, metalinguistic feedback, subject-verb agreement, simple past tense

1. Introduction

Although Oral Corrective Feedback (OCF) on learners’ errors in non-native language (L2) teaching has always been a customary teaching practice, research in OCF essentially started in the 1990s (for a recent overview, see Agathopoulou, 2020). Based on recordings from L2 French immersion classes in Canada, Lyster and Ranta (1997) classified OCF into six types, which were verified, enriched and further classified by following studies (e.g. Sheen & Ellis, 2011). The effect of OCF on L2 development investigated through a pre-test, post-test and delayed post-test methodology has shown that whether OCF may boost L2 development or whether particular OCF types are more beneficial than others may depend on multiple factors, among which are the target grammatical language forms and
the learners’ L2 developmental readiness (Ammar, 2008), the type of OCF (e.g. implicit/explicit or prompts/reformulations), and language learning anxiety (for an overview, see Ellis, 2017).

The current research is a case study of two twin EFL learners who, as we will show, differed from each other only with respect to levels of language learning anxiety. We explored the role of recasts and metalinguistic feedback in the development of the Subject-Verb agreement morpheme –s and the past tense morpheme –ed, as well as the relationship between OCF and language anxiety.

In the next sections, first we briefly review research in recasts and metalinguistic feedback, as well as in language learning anxiety and then we present our method and results. The final section includes a discussion of our results, the limitations of our study, suggestions for future research and the pedagogical implications of our findings.

2. Recasts and metalinguistic OCF

Recasts, which consist in the correct reformulation of the learners’ utterances, seem to be the most frequent OCF type across various language teaching contexts (Sheen, 2004). However, recasts also seem to elicit fewer reactions (uptake) from learners than other types of OCF, such as metalinguistic feedback or elicitation (Lyster & Ranta, 1997; Milla & Mayo 2014; Panova & Lyster, 2002). Consider the following examples from Agathopoulou et al.’s (2015) classroom data which involved primary school learners of EFL in Greece, simplified for our purposes here. (T=teacher, L=learner).

(1)  
T: I wish …  
L: have dog.  
T: I wish I could have a dog, ha?  
L: Yeah.

(2)  
L: Expensiver.  
T: Mmm. Yes. Good. I agree with this, we have two things and we compare them, but do you remember what I told you about this long word you had? Difficult. Expensive is as long as this one.  
L: More expensive.

In (1) the teacher tries to elicit a sentence containing the structure of the second conditional, the learner offers an ungrammatical answer and the teacher recasts the whole target sentence adding “could”. The uptake on the part of the learner consists solely of “Yeah”, which is ambiguous as to whether s/he affirms that s/he understood the grammatical correction or that s/he agrees to that s/he wishes she could have a dog. In (2), the teacher offers metalinguistic feedback on an error of the comparative form of long adjectives (“expensiver”), without providing the correct form and the learner’s uptake contains the correct form (“more expensive”).

Based on observations that learners respond less to recasts than other types of OCF, it has been assumed that recasts may not be salient enough for learners to notice their errors, especially in content-based instruction (Lyster, 1998). It has also been observed that when teachers use recasts, they often do not give learners the time and opportunity for uptake (Sheen, 2004). However, although a learner to whom a recast is addressed may not react to it, other learners in the same classroom respond by privately repeating the correct form (Ohta, 2000, reported in Lightbown &
Recasts can be made more noticeable either by emphasizing the wrong form (e.g. “I wish I COULD have a dog”) or by reformulating only the part of the learner’s utterance with the target form (e.g. “could have”) (cf. (1)). These types of recasts have been termed “didactic recasts” and are distinguished as more explicit OCF than conversational and communicatively-oriented recasts which are considered implicit OCF (Ellis & Sheen, 2006; Llinares & Lyster, 2014; Sheen, 2006).

Unlike recasts, which are considered implicit OCF, metalinguistic comments/clues belong to the explicit category of OCF and they generally elicit uptake (and self-repair) more than recasts (Lyster & Ranta, 1997). The way teachers choose to formulate metalinguistic feedback may vary according to the instructional context (e.g. communicative vs. structure-based language teaching), the learners’ age and their knowledge of metalanguage. For instance, in example (2), the context is a class of young learners in a primary school where EFL is taught mostly through a meaning-based approach and without explicit grammar teaching. So the teacher does not employ metalinguistic terms such as “comparative form” but talks about forms that are compared and long words instead. On the other hand, in studies where the participants were adults learning English through a grammar-based approach, metalinguistic feedback contained comments such as “You need past tense” (Ellis, Loewen & Erlam, 2006, p. 353) or “You should use the definite article ‘the’…” (Sheen, 2010, p. 233).

The brief literature review that follows is limited to three studies which investigated the effects of recasts versus metalinguistic feedback or other types of prompts on the development of the structures investigated in the current study. Ellis et al. (2006) compared the effectiveness of recasts versus metalinguistic feedback on the formation of the regular simple past tense with low intermediate adult EFL learners. In their teaching interventions (where OCF was provided), they employed picture story retelling and oral text reconstruction tasks. The learners’ progress was monitored through pre-tests, immediate post-tests and delayed post-tests involving oral imitation and untimed grammaticality judgments. The results showed superiority of metalinguistic feedback in most respects. Yang and Lyster (2010) too used pre-tests, immediate post-tests and delayed post-tests to examine the effect of recasts and prompts on Chinese learners’ acquisition of the regular and irregular past tense. The teaching interventions included picture-based oral narratives, a question-and-answer oral activity with cards and a dictogloss activity. Their results showed that prompts, which included metalinguistic feedback, were more effective than recasts in the regular past tense, while recasts were effective too regarding the use of the irregular past tense. Last, Guo and Yang (2018) focused on the treatment of third person –s with Chinese intermediate EFL college learners. What was particularly interesting in their teaching intervention were the two oral tasks: one with questions and answers and a picture description task. Results from pre-, immediate post- and delayed post-tests showed that the learners who had received OCF in the form of prompts had a general advantage over those who had received recasts.

3. Language Learning Anxiety and OCF

Anxiety is an individual factor considered to have an important role in successful language learning (Hardacre & Güvendir, 2020; Horwitz, 2001; Sheen, 2008; Szyszka, 2017). The symptoms of anxiety may vary from simple nervousness (Resnik & Dewaele, 2020), sweat and difficulty in concentration to apprehension, trouble and even fear and dread (Gkonou, 2013; Horwitz, Horwitz & Cope, 1986). The causes of language anxiety may be linked to external or internal factors (Szyszka, 2011). External factors may concern the classroom context, teacher-student relationships (Gkonou, 2014; Szyszka,
2017) as well as the learners’ understanding of the classroom input difficulty and fear of negative evaluation (Gkonou, 2013). Some internal factors of language anxiety may be the learners’ beliefs about themselves (Szyszka, 2017), their low self-esteem (Piechurska-Kuciel, 2011) and the learners’ competitiveness and perfectionism towards their peers (Gkonou, 2014; Gregersen & Horwitz, 2002).

Using the Foreign Language Anxiety Scale (FLCAS) (Horwitz et al., 1986), studies have shown that anxiety is debilitative (Horwitz, 2001) and potentially leads to poor linguistic success (Sheen, 2008). High-anxiety learners seem to be those whose experiences are shaped negatively from anxiety and whose general L2 academic progress and development are interrupted and impeded (Ellis, 2001; Gkonou, 2014; Horwitz et al., 1986; Piechurska-Kuciel, 2011; Szyszka, 2011; Liu & Chen, 2013). Let it be noted, however, that anxiety may affect various language skills / aspects of language learning differentially (for a recent review, see Deyuan, 2018).

To our knowledge, there is very little research regarding the relationship between OCF and language learning anxiety. Dekeyser’s (1993) study with 35 Dutch senior high school learners of French showed that there is a clear interaction between OCF and anxiety, since low-anxiety learners performed better on written grammar tests than high-anxiety learners after intensive OCF treatment sessions. Sheen (2008) found that after receiving recasts on errors regarding English articles, low-anxiety learners outperformed their high-anxiety peers. Last, Rassaei (2015) explored the effects of oral recasts and oral metalinguistic feedback in the acquisition of L2 English articles by EFL learners in Iran. Results from a pre-test/post-test design which included three tests each, indicated that high-anxiety learners benefited most from recasts while low-anxiety learners improved from both CF types but mostly from metalinguistic feedback.

4. The current study

4.1 The Research Question

Our aim was to answer one basic research question: What is the interaction between recasts (implicit OCF), metalinguistic feedback (explicit OCF) and language learning anxiety in the development of Subject-Verb agreement and regular past tense in young EFL learners?

Assuming that one of our participants would be a low-anxiety learner while the other one would be a high-anxiety learner and based on previous research findings, we hypothesized that metalinguistic feedback, which is a more salient/explicit type of OCF, would be more beneficial than recasts for the low-anxiety learner while the high-anxiety learner would benefit more from recasts, which are implicit and less anxiety-inducing than metalinguistic feedback.

4.2 Methodology

Before proceeding with the current research, the participants and their mother were informed about the aims of the study and were ensured that the data would be anonymous. Also, a written consent was obtained from the twins’ mother.

Participants

The participants (called M and E here for purposes of anonymity) were two 11-year-old female twin sisters learning EFL in Greece. At the time of the study, they were in the 6th grade of primary school,
and their English instruction was mainly form-focused with frequent tests and exams targeting grammar and vocabulary. Each participant had had exactly the same amount of previous exposure to English instruction, which was approximately 600 hours.

Observational data indicated that M and E differed from each other in terms of language learning anxiety. For instance, during EFL classes M always seemed relaxed and confident whereas E was anxious about finishing her assignments and seemed less confident than her sister regarding her ability to learn new things. For these reasons, we decided to investigate their language learning anxiety by means described later here (see *Instruments*).

**Target Structures and CF types**

Our target structures were the suffix –ed for regular past formation and the suffix –s for Subject-Verb (S-V) agreement. We selected these structures because although they are very frequent in English (Guo & Yang, 2018) and are introduced early in EFL teaching, they are difficult to acquire (Ellis, 2006; Krashen, 1982). Other criteria for the selection of the particular structures were that they are both grammatical morphemes of the same language proficiency level (elementary to lower intermediate) (Erlam, 2006) and are both considered more amenable to explicit rather than implicit corrective feedback (Ellis, 2006). Additionally, we decided to include these two grammatical structures rather than one of them in order to check possible effects in this domain, given that the S-V agreement morpheme –s is acquired later than the past tense –ed (Goldschneider & Dekeyser, 2001).

We employed two CF types: (a) metalinguistic feedback, which involved comments on errors using metalanguage and (b) recasts, presented as correct reformulations of only the part of the learner’s utterance which contained the error, in the declarative form (Ellis et. al., 2006; Guo & Yang, 2018). As already mentioned, metalinguistic feedback may be more anxiety-inducing due to its explicit nature (Rassaei, 2015), while recasts, which are rather implicit CF, are considered nontargeting for the learners and may thus cause less anxiety (Sheen, 2008; Rassaei, 2015).

**Instruments**

**A. The Anxiety Questionnaire**

This questionnaire was adapted in Greek (see Appendix A) from two other questionnaires mentioned below. The questionnaire consisted of two parts to be completed. Part A included eight questions aimed at obtaining biographical data (age, gender, years of learning the target language, knowledge of other non-native languages). All of these eight questions, as well as the guidelines for the completion of the questionnaire (“Message to the Student”) were adapted from Gkonou and Oxford’s (2016) questionnaire. Part B focused on language learning anxiety and consisted of 36 closed type questions rated on a five-point Likert scale (1=totally disagree, 5=totally agree). In Part B questions 1-33 were adapted from the questionnaire in Horwitz et al. (1986). Some examples are “I never feel quite sure of myself when I am speaking in my foreign language class.”, “I am usually at ease during tests in my language class.”, “I don’t understand why some people get so upset over foreign language classes.” and “It embarrasses me to volunteer answers in my language class.”. The last three questions (34-36) were adapted from scenarios in Gkonou and Oxford’s (2016) questionnaire. For example, Scenario 2 in Gkonou and Oxford was “You make a mistake during a classroom oral activity. Your teacher corrects you in front of the class” and it was followed by 9 questions regarding the emotions this incident may cause. The first question asked whether in a situation like the above, the learners experience negative or positive emotions. In our questionnaire
the relevant adapted item was “I feel uncomfortable when the foreign language teacher corrects an error I made in front of the whole class”.

B. The interviews

Semi-structured interviews were conducted both with the two learners and their mother. Most of the prepared questions were adapted in Greek from the scenarios in Gkonou and Oxford (2016). The scenarios employed during the interviews with each of the two girls were: (a) Imagine that you make an error during an oral activity. Your teacher corrects you in front of the class. How would you feel? Have you ever encountered a similar situation?, (b) The teacher asks you a question in class. You do not understand the question and, therefore, you ask the teacher to repeat it. After the repetition, you still don’t get the question. How would you feel? Would you ask the teacher to explain the question again? (c) You go to your language class unprepared because something happened and you didn’t have time to complete your homework. How would you feel?, (d) A student in your language class always tries to outperform you. This student achieves a score of 100% on a language test, but you achieve only 75%. How would you feel? and (e) Your family is planning a vacation to a location where people speak the language you are studying. How would you feel?

In the interview with the girls’ mother, after the first question “How would you describe your daughters’ feelings and anxiety about learning English?”, she was asked how she thought each of her daughters would feel in the scenarios (a)-(e) above.

C. The Oxford Quick Placement Test (OQPT 2001)

The OQPT is a standardized test including a total of 60 multiple-choice items to be completed in 30 minutes. Scores between 0-15, 16-23 and 24-30 correspond to “Breakthrough”, “Elementary” and “Lower Intermediate” respectively, while higher scores correspond to “Intermediate”, “Advanced” and “Very Advanced” (for more details, see https://www.cambridgeenglish.org/Images/23127-research-notes-12.pdf).

D. The untimed Grammaticality Judgment Test (GJT)

The GJT was created and used as a pre-, post- and delayed post-test to measure the participants’ explicit knowledge of the target structures (S-V agreement in simple present tense, and regular past tense). The GJT contained 40 sentences (23 grammatical and 17 ungrammatical) adapted from various English textbooks and the British National Corpus (2007) and was digitized in a young learner friendly design. The participants were required to judge the grammaticality of each sentence by clicking one of the three faces below the sentence, standing for “correct/I don’t know/incorrect” (Figure 1). Prior to the actual test, the participants had had practice with 10 sentences targeting other structures. Correct/incorrect responses were coded as 1 and 0 respectively, while “I don’t know” responses were not counted in.
Figure 1. Example of an ungrammatical sentence from the GJT

We watch a scary film yesterday.

E. The Elicited Imitation Test (EIT)

The EIT was based on Marinis and Armon-Lotem (2014) and it was used as a pre-, post- and delayed post-test to measure the participants’ implicit knowledge of the target structures. It consisted of 30 items extracted from English textbooks as well as from Marinis and Armon-Lotem (2014) and Erlam (2006). All the sentences were grammatical, recorded by native-like speakers of English and incorporated in a Power Point Presentation illustrating a treasure hunt scene, where the goal was for the main character (a monkey) to reach a final destination (the top of a banana tree with a bunch of bananas) (Figure 2). The monkey moved on the pink circles and the participants had to repeat the sentences they heard each time the monkey stepped on a pink circle as well as possible. (Examples from the EIT: “The horse kicked the man in the back.” and “My cat climbs on the furniture all the time.”)

Figure 2. The final location in the elicited imitation test

The immediate and the delayed post-tests contained the same items as those in the pre-tests (both the GJT and the EIT). The immediate post-tests were administered after each intervention week (see Figure 3, Research Design) while the delayed post-tests were administered four weeks after each post-test session.

The Teaching Interventions

The two participants were treated individually. They each undertook eight sessions that lasted for two weeks. These sessions involved the same story re-telling tasks based on Simon’s Cat short comic videos (Todfield, 2017). During each session, a short video was first presented to the participants and afterwards they were asked to study a brief written account of the same story for a few minutes. Then the story was removed and they were provided with a list of verbs to use in retelling the story
and thus to create as many obligatory contexts as possible for the target structures. The context for using the simple past tense or the third person –s was also created by providing the participants with some opening words for their narration, for example ‘Yesterday, Simon’s cat...’ or ‘Every evening, Simon...’ (cf. Ellis, 2006).

During the first week of interventions, the focus was only on S-V agreement and M received metalinguistic feedback whereas E’s errors were treated with recasts. During the second week, the focus was on regular past tense (–ed) and CF was applied reversely to the two participants. All the tasks were recorded on an audio tape recorder and then transcribed. Figure 3 illustrates the summary of the research design and (3)-(6) are examples of corrective feedback episodes (CFE) from the teaching interventions, slightly adapted from our recordings. OCF is in italics and T stands for teacher.

(3) Target structure: -s agreement, OCF: metalinguistic  
M: In the story, Simon’s cleans the kitchen. Eeeem.... Simon’s mop the floor...  
T: *What kind of ending should you use here?*  
M: Mops...the floor!

(4) Target structure: -s agreement, OCF: recast  
E: Every afternoon Simon sit the...  
T: *sits*  
E: sits the sofa

(5) Target structure: past -ed, OCF: recast  
M: Yesterday, the Simon’s cat go...walk in the living room.  
T: *walked in the living room*  
M: walked in the living room.

(6) Target structure: past -ed, OCF: metalinguistic  
E: Last night Simon decorated the Christmas tree. Emm the Christmas tree. Then, the..the looked and the remember...he remember  
T: *This happened last night, so what ending should you put?*  
E: remembered.
5. Results

5.1 The Anxiety Questionnaire

To measure the participants’ anxiety level, the results of the piloting group from the anxiety questionnaire were first taken into account and these results were afterwards compared with the participants’ scores. Results from the pilot group yielded scores between 1.22 and 4.67. The learners who scored more than the mean score (2.68) plus the standard deviation (1.37) (M=4.05) were classified as high-anxiety learners and those who scored lower than the mean score minus the standard deviation (M=1.31) were characterized as low-anxiety ones. An independent samples T-test showed a significant difference between the two groups (F = 4.842, Sig. = .050) (t (9) = -21.454, p < .001). Turning to the participants of the main study, M’s anxiety
score was 1.30, while E’s score was 4.53, which categorizes them as a low-anxiety and a high-anxiety learner, respectively.

### 5.2 The Interviews

The different anxiety profiles of the two learners were further verified in the interviews. Next, first we expose the main points made by the learners and then by the learners’ mother.

E’s responses to the scenarios (for the scenarios, see 4. Methodology):

(a) She does not want to commit errors when she speaks in the classroom, otherwise she feels ashamed and thus she does not speak in the classroom unless she is sure about her answer. (b) If she didn’t understand a teacher’s question, she would feel really anxious. She would not be able to ask the teacher to repeat the question because she would freeze in fear of being scolded by the teacher. In one occasion when she was unable to answer a teacher’s question, she started to cry. (c) She would never dare go to the language class unprepared because she hates doing so and she wants to be perfect in school. If she had to do something like this, she would definitely lose her sleep the night before and she would ask her mother to accompany her to school in order to explain everything to the teacher. (d) Although she always tries so hard and studies a lot before a test, during the test she forgets everything she studied at home. (e) She would not travel abroad on her own because she would be too nervous if she had to communicate in English; she would worry that she would not be understood and thus she would prefer to travel with her mother and depend on the latter’s ability to communicate in English. Alternatively, she would depend on her sister who doesn’t have a problem to speak to people in English.

M’s responses to the scenarios:

(a) She said she always makes errors but she thinks it’s OK since “we learn from our errors”. If the teacher corrected her, she would feel relaxed because she thinks the teacher does not correct errors in order to punish learners but because she wants them to become better. (b) If she didn’t understand a teacher’s question, despite asking the teacher to repeat the question, she would repeat that she didn’t understand and would ask the teacher to explain it again, in different words. (c) If, for a good reason, she had to go to class unprepared, she would feel relaxed; she would explain to the teacher why she had not done her homework and promise that she would not do it again. (d) She would not mind if a classmate scored much higher than her on a test and if the classmate was a friend, she would be happy about it. In case the particular classmate was not a friend and was very competitive towards her, she would try to be better in the future. (e) She would be very happy to travel abroad and have the chance to communicate in English because she thinks she speaks English very well and if she did not know some words, she would look them up on the Internet.

The mother’s responses in the interview:

The mother said that both girls like learning English and find it useful for their future; M is always relaxed, doesn’t care much about making errors, unlike E who feels stressed about it. Also, if, for example, M does not manage to finish her homework on the previous day because she is tired, she will wake up earlier the following morning before she goes to school to finish everything. On the other hand, E would be very anxious about whether she would manage to finish her homework on time. The mother added that unlike with M, E’s teachers say that she is highly anxious also during the lessons and if E doesn’t understand something she won’t ask about it because she is ashamed to ask
a question in front of the whole class. Also, unlike M, E suffers from anxiety regarding learning English as well as learning German. In the mother’s opinion one reason for E’s anxiety may be because that she competes with her sister, M, who is slightly better than E at school. The mother also verified that M would not really care if a classmate who competed with her, scored higher than her on a test but, on the other hand, such a situation would cause anxiety to E because “E is always concerned about comparing herself to the other students”. Last, the mother said that on a trip abroad E would certainly feel insecure to communicate in English, very much unlike what M would do.

5.2 The Quick Oxford Placement Test

The two participants’ scores in the QOPT were very similar; M scored 20/60 (33.3%) and E scored 21/60 (35%). These scores placed both participants on the A2 (“elementary”) level of English proficiency, according to the CEFR (Common European Framework of Reference for Languages).

5.3 The Untimed GJT

Table 2 presents the participants’ scores on the GJT for the two structures. There were no significant differences between the participants’ pre-tests for the third person –s (χ² (32, 1) = 3.463, p = .063) or for the past simple –ed (χ² (29, 1) = .032, p = .858). A comparison between results in the pre- and the immediate post-test revealed a statistically significant improvement for both M’s (χ² (34, 1) = 7.889, p = .005) and E’s scores (χ² (32, 1) = 5.236, p = .022), for –s agreement, but no significant differences for past –ed (M: χ² (30, 1) = .475, p = .491; E: χ² (29) = .318, p = .573). Also, there were no statistically significant differences between scores in the immediate and the delayed post-tests, either for –s (M: (χ²(32, 1) = .064, p = .790); E: (χ²(28, 1) = .571, p = .571)) or for –ed (M: (χ²(31, 1) = .605, p = .450), E: (χ²(28, 1) = .450)).

The above results indicate that M significantly improved from metalinguistic feedback and E from recasts in errors regarding S-V agreement –s and that this improvement was retained four weeks later. On the other hand, M did not significantly improve from recasts and neither did E from metalinguistic feedback on past tense –ed in the short or long term.

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Immediate Post-test</th>
<th>Delayed Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-s</td>
<td>-ed</td>
<td>-s</td>
</tr>
<tr>
<td><strong>M</strong></td>
<td>18.8 (3/16)</td>
<td>50 (7/14)</td>
<td>66.7 (12/18)</td>
</tr>
<tr>
<td><strong>E</strong></td>
<td>50 (8/16)</td>
<td>46.7 (7/15)</td>
<td>87.5 (14/16)</td>
</tr>
</tbody>
</table>

Table 2: Results from the Untimed GJT in percentages (–s agreement and –ed past tense)

5.4 The EIT

Table 3 illustrates the participants’ overall performance in the EIT. No significant differences were found between the participants’ scores in the pre-test (S-V –s: (χ² (30, 1) = .159, p = .690); past –ed: (χ² (30, 1) = .144, p = .705). The increase in scores between the pre- to immediate post-test for –s was statistically significant both for M (χ² (30, 1) = 8.571, p = .003) and for E (χ² (30, 1) = 4.821, p = .028). However, the apparent improvement in –ed was not significant for neither of the participants (M: χ²
(30, 1) = 3.394, p = .065; E: $x^2(30, 1) = 2.143, p = .143$). Last, the decrease in scores from the immediate to the delayed post-test did not yield any statistically significant differences either for M ($-s$: $x^2(30, 1) = .186, p = .666$; $-ed$: $x^2(30, 1) = .159, p = .690$) or for E ($-s$: $x^2(30, 1) = .833, p = .361$; $-ed$: $x^2(30, 1) = .536, p = .464$).

The above results are similar to those from the GJT, as they too suggest that M significantly benefitted from metalinguistic feedback ($in -s$) but not from recasts ($in -ed$), while E significantly benefitted from recasts ($in -s$) but not from metalinguistic feedback ($in -ed$). It may be worth pointing out that E improved even more in $-s$ in the delayed post-test, although, as already mentioned, this improvement was not statistically significant.

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Immediate Post-test</th>
<th>Delayed Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$-s$</td>
<td>$-ed$</td>
<td>$-s$</td>
</tr>
<tr>
<td>M</td>
<td>26.7</td>
<td>40(6/15)</td>
<td>80(12/15)</td>
</tr>
<tr>
<td>E</td>
<td>33.3</td>
<td>33.3(5/15)</td>
<td>73.3(11/15)</td>
</tr>
</tbody>
</table>

Table 3: Results from the EIT in percentages ($-s$ agreement and $-ed$ past tense)

Last, although the participants had generally higher scores in the EITs than in the GJTs, no significant task effect was observed between the two tasks (p > .275 regarding all the tests, and both of the target structures). In other words, there were no statistically significant differences regarding the development of explicit versus implicit knowledge of the target structures as measured by the GJT and the EIT, respectively.

6. Discussion and Conclusions

Our findings showed that one of our twin learners had a high anxiety profile while the other one was a low-anxiety learner. The high-anxiety learner mastered the structure treated with recasts (S-V agreement $-s$), but not the structure to which she had received metalinguistic feedback (past tense $-ed$) while the reverse was found for the low-anxiety sister. So they both progressed significantly in mastering the S-V agreement $-s$ but neither did so in the past tense $-ed$. As we have shown, the two learners had similar scores in the target structures in the pre-test, they were of the same English proficiency level, they had the same amount of exposure to English and they were of exactly the same age. Thus, we may plausibly assume that our results yielded a correlation between the level of language anxiety and the type of OCF.

Further support for our assumption may be sought in previous research (Ellis et al., 2006; Guo & Yang, 2018; Yang & Lyster, 2010) where generally metalinguistic OCF proved more beneficial than recasts, while in our study this effect holds for the low-anxiety learner but not for the high-anxiety one. Also, the S-V agreement $-s$ is acquired later than the past tense $-ed$ in L2 English (Goldschneider & DeKeyser, 2001) but both of our learners progressed only in the S-V agreement $-s$, which is the most difficult structure.

Although, as mentioned, the recasts employed here were more explicit than conversational and communicatively-oriented recasts (see examples (4) and (5)), they were still more implicit than metalinguistic feedback. Thus, we may cautiously suggest that
implicit OCF, even in the form of these more ‘didactic’ recasts (Ellis & Sheen, 2011), is more suitable than explicit OCF for high-anxiety learners since this type of implicit feedback interrupts the flow of narratives much less than metalinguistic feedback.

The benefits from OCF in our study may also be due to the fact that it was offered individually. Unlike in other studies where OCF was used in classrooms with twenty or more learners and often there was no uptake, especially when recasts were employed (e.g. Lyster & Ranta, 1997; Milla & Mayo, 2014; Agathopoulou et al., 2015), in our study individual OCF resulted in a 100% uptake for the low-anxiety learner and 92% for the high-anxiety one. This result may point to the fact that in private tutoring, OCF becomes more effective for learners. Thus, we think that the effect of other important individual factors relevant to the effect of OCF (such as aspects of working memory) may be irrelevant in our study. Moreover, as previously discussed, a lack of uptake has been linked to the fact that teachers do not give learners the opportunity to react to OCF (Sheen, 2004). On the other hand, in our individual treatments the learners were offered enough waiting time to respond to OCF.

Regarding the pedagogical implications of our study, language teachers should be aware that decisions about how to correct errors may depend on individual learner factors, among which is anxiety. Having an awareness of the role of language anxiety is important since this factor may also negatively affect “willingness to communicate” (Lightbown & Spada, 2013, p. 86) and, consequently, it may affect progress in L2 development. Teachers should acknowledge that anxiety is something their learners should not be ashamed of (Resnik & Dewaele, 2020) and that they should try to reduce it in the EFL classroom (Gkonou, 2014) by creating a friendly environment (Gkonou & Miller, 2017). To this end, teachers may employ affective language teaching strategies which involve joyful activities, for example, puns, songs and humor (Psaltou-Joycey, 2020, p. 181).

As a final note, we are aware that since our study involved only two participants, our results could hardly be generalisable, which is the main limitation of the current research. More data would probably yield different results, for instance, with respect to the development of the explicit/implicit knowledge of the target structures, which we attempted to investigate by means of the two different tasks in our tests, the GJT and the EIT. However, we would like to point out that the twin sisters in our study, who differed only with respect to their language anxiety profile, have offered a rare and rather ideal case of participants to use in exploring the link between implicit/explicit OCF and anxiety.

Acknowledgments

We thank Christina Gkonou for sending us the questionnaire devised by Gkonou and Oxford (2016). We also thank the two anonymous reviewers of our paper.

References


APPENDIX A

The Anxiety Questionnaire

Μέρος Α: Κάποιες πληροφορίες για σένα

A. Φύλο: ☐ Αγόρι ☐ Κορίτσι
B. Πόσοι χρόνοι είσαι; __________________________
C. Ποια είναι η μητρική σου γλώσσα; __________________________
D. Ποια ξένη γλώσσα μαθαίνεις αυτή την περίοδο; (Γράψε την κύρια γλώσσα που μαθαίνεις, αν μαθαίνεις παραπάνω από μια) __________________________

Ε. Για πόσο καιρό μαθαίνεις αυτή τη γλώσσα;
α. 1 – 3 χρόνια
β. περισσότερο από 3 χρόνια

ΣΤ. Γιατί μαθαίνεις αυτή τη γλώσσα; (Κύκλωσε όλους τους λόγους που σου ταιριάζουν)
α. Με ενδιαφέρει η γλώσσα
β. Είναι υποχρεωτικό να την μάθω
γ. Θέλω να μάθω περισσότερα για τον πολιτισμό αυτής της γλώσσας
d. Έχω φίλους που μιλάνε αυτή τη γλώσσα
ε. Την χρειάζομαι για την μελλοντική μου καριέρα
ς. Την χρειάζομαι για να μπορώ να ταξιδεύω
ζ. Θέλω να σπουδάσω στο εξωτερικό
η. Άλλο: __________________________

Ζ. Πόσες άλλες γλώσσες μιλάς; ___
Η. Πόσο σημαντικό είναι για σένα να μαθαίνεις ξένες γλώσσες;
α. Πολύ σημαντικό
β. Λίγο σημαντικό
γ. Καθόλου σημαντικό

Μέρος Β: Ερωτήσεις

Βάλε ✓ στην στήλη που ταιριάζει καλύτερα στην απάντηση σου.

<table>
<thead>
<tr>
<th>Ποτέ δεν νιώθω σίγουρος/η για τον εαυτό μου όταν μιλάω μέσα στην τάξη της ξένης γλώσσας.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δεν ανησυχώ για το αν κάνω λάθη στην ξένη γλώσσα.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Τρέμω όταν ξέρω πως είναι η σειρά μου να μιλήσω μέσα στην τάξη της ξένης γλώσσας.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Με τρομάζει όταν δεν καταλαβαίνω τι λέει ο/η δάσκαλος/α μου στην ξένη γλώσσα.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Δεν θα με πείραξε καθόλου να κάνω περισσότερα μαθήματα ξένης γλώσσας.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
6. Κατά την διάρκεια του μαθήματος της ξένης γλώσσας, κάποιες φορές σκέφτομαι άλλα πράγματα που δεν έχουν καμία σχέση με το μάθημα.

7. Σκέφτομαι συνέχεια ότι οι άλλοι μαθητές είναι καλύτεροι στις ξένες γλώσσες από εμένα.

8. Συνήθως νιώθω άνετα όταν γράφω τεστ στην ξένη γλώσσα.

9. Με πιάνει πανικός όταν πρέπει να μιλήσω στην ξένη γλώσσα χωρίς να έχω προετοιμαστεί.

10. Ανησυχώ για το ποιες θα είναι οι συνέπειες της αποτυχίας μου στο μάθημα της ξένης γλώσσας.

11. Δεν καταλαβαίνω γιατί μερικοί άνθρωποι ταράζονται τόσο πολύ σχετικά με την εκμάθηση των ξένων γλωσσών.

12. Στην τάξη των ξένων γλωσσών, μπορεί να αγχωθώ τόσο πολύ ώστε να εξελθώ πράγματα που ξέρω.

13. Το να σηκώω το χέρι μου για να πω μια απάντηση μέσα στην τάξη της ξένης γλώσσας με φέρνει σε αμηχανία.

14. Δεν θα αγχωνόμουν να μιλήσω την ξένη γλώσσα με φυσικούς της ομιλητές.

15. Αναστατώνομαι όταν δεν καταλαβαίνω τι μου διορθώνει ο/η δάσκαλός/α μου.

16. Ακόμα και αν έχω προετοιμαστεί πολύ καλά για το μάθημα, πάλι νιώθω άγχος για αυτό.

17. Συχνά νιώθω ότι δεν θέλω να πάω στο μάθημα της ξένης γλώσσας.

18. Νιώθω αυτοπεποίθηση όταν μιλάω μέσα στην τάξη της ξένης γλώσσας.

19. Φοβάμαι ότι ο/η δάσκαλός/α μου θα μου διορθώσει οποιοδήποτε λάθος και αν κάνω.

20. Νιώθω την καρδιά μου να χτυπάει γρήγορα όταν είναι η σειρά μου να μιλήσω μέσα στην τάξη της ξένης γλώσσας.

21. Όσο περισσότερο διαβάζω για ένα τεστ, τόσο περισσότερο μπερδεύομαι.

22. Δεν νιώθω πίεση για να προετοιμαστώ πολύ καλά για το μάθημα της ξένης γλώσσας.

23. Πάντα νιώθω ότι οι άλλοι μαθητές μιλάνε καλύτερα την ξένη γλώσσα από εμένα.

24. Νιώθω πολύ άβολος και νευρικός στο μάθημα της ξένης γλώσσας από ότι στα υπόλοιπα μαθήματα.

25. Νιώθω άγχος οποτε οι άλλοι μαθητές μιλάνε καλύτερα την ξένη γλώσσα με χωρίς να έχω προετοιμαστεί.
| 28. | Όταν είμαι στο δρόμο για το μάθημα της ξένης γλώσσας, νιώθω σίγουρος/ή και χαλαρός/ή. |
| 29. | Αγχώνομαι όταν δεν καταλαβαίνω όλες τις λέξεις που λέει ο/η δάσκαλος/α μου. |
| 30. | Νιώθω πυγμένος/ή από τον αριθμό των κανόνων που έχει κανείς να μάθει για να μιλήσει στην ξένη γλώσσα. |
| 31. | Φοβάμαι ότι οι άλλοι μαθητές θα γελάσουν μαζί μου όταν μιλάω στην ξένη γλώσσα. |
| 32. | Πιθανότατα θα ένιωθα άνετα να είμαι μαζί με ανθρώπους που έχουν την ξένη γλώσσα ως μητρική τους. |
| 33. | Νιώθω άγχος όταν ο/η δάσκαλος/α της ξένης γλώσσας μου κάνει ερωτήσεις που δεν έχω προετοιμάσει από πριν. |
| 34. | Νιώθω άβολα όταν ο/η δάσκαλος/α της ξένης γλώσσας μου διορθώνει ένα λάθος που έκανα μπροστά σε όλη την τάξη. |
| 35. | Νιώθω άβολα όταν κάποιος συμμαθητής μου έκανε λάθος που έκανα μπροστά σε όλη την τάξη. |
| 36. | Πιστεύω πως όσο σκληρά και αν προσπαθήσω, πότε δεν θα καταφέρω να τα πάω καλά στην ξένη γλώσσα. |

Πως σου φάνηκε η εμπειρία σου; Έχεις κάτι άλλο να προσθέσεις; Γράψε κάποια σχόλια:

________________________________________________________________________________
________________________________________________________________________________
________________________________________________________________________________

Σε ευχαριστώ πολύ για την συμμετοχή σου στο ερωτηματολόγιο!

**Endnote:** This article is based on Evangelia Paraskeva’s (2020) MA thesis, supervised by Eleni Agathopoulou at the Aristotle University of Thessaloniki.

Evangelia Paraskeva ([evapar13@gmail.com](mailto:evapar13@gmail.com)) is an EFL teacher holding an MA in Foreign Language Learning and Teaching from Aristotle University of Thessaloniki. She has been teaching English as a Foreign Language both privately and in Greek private language institutes for 7 years. Her main research interests concern CF, anxiety and classroom technology.

Eleni Agathopoulou ([agatho@enl.auth.gr](mailto:agatho@enl.auth.gr)) is Associate Professor at the School of English Language and Literature, Aristotle University of Thessaloniki. She holds a BA and a PhD from Aristotle University and an M.A. in Applied Linguistics from the University of Reading, U.K. She has published and presented papers on L2 morphology and syntax, learning strategies, focus-on-form instruction and English for Academic Purposes.
Digging deeper: Investigating the emotional impact of online learning on university students during COVID19 in Greece

Thomaï ALEXIOU & Theologia MICHALOPOULOU

The paper investigates the emotional impact online learning had on university students during the Covid-19 pandemic. For this purpose, a quantitative research was conducted in which 83 university students answered a questionnaire. The results indicated that, although the students feel confident with their technological skills and the support they have received in online learning and trust their professors, they are not satisfied with this kind of instruction. Dissatisfaction is associated with long hours spent attending online classes. Students who had not been used to face-to-face instruction prior to the pandemic are more overwhelmed by their professors’ expectations in online learning as opposed to their peers who had some familiarization with the latter. A connection was found between stress and deteriorated academic performance, while the participants reported struggling to balance their personal and university life and feeling frustrated that they cannot see their classmates due to turned-off cameras. This led to feelings of loneliness, which was associated with low motivation. Anxiety, exam- and workload-related stress, isolation, and compromised mental health were among the recurring issues. Nevertheless, bonding with the professors and enjoying the flexibility of online learning were highlighted, while the study gave voice to socially anxious students who embraced online learning.

Key words: online learning, COVID19, university students, emotional impact

1. Introduction

The rapid spread of the Covid-19 virus around the globe triggered a series of lockdowns and curfews that resulted in drastic changes in the everyday lives of citizens everywhere. University students were among the groups of the population whose life was significantly affected by the Covid-19 restrictions (Baloran, 2020; Arora et al., 2020).

In compliance with the national regulations that were enforced by governments worldwide, universities shifted to online classes in what has been characterized as “Hobson's choice”, and students were called to adjust to the new learning methods under remarkably difficult and unprecedented circumstances (Arora et al., 2020, p.1).
2. Theoretical background

The younger generation presents a level of familiarization with technology which renders its representatives “hyperconnected and facile with computers and the internet” (Boysen et al., 2016, p.103). Contrary to their university professors belonging to older generations that may be struggling to keep up with the constant changes in technology, university students’ clear advantage is that the majority of them belong to ‘Generation Z’ thus being by definition digital natives in the sense that they were introduced to technology and the use of the Internet from the beginning of their lives (Boysen et al., 2016; Prensky, 2001).

Nevertheless, online learning is something which entails challenges that university students – despite their familiarization with technology – may not be able to deal with. Being familiar with social media platforms and applications to ease one’s everyday life does not necessarily entail that one is also in position to attend online classes in higher education on a daily basis and cope with their demands. Particularly in the context in which the shift to online learning was made, university students found themselves having “hardly any time for preparation and coping with new context” (Arora et al., 2020, p.2). Research has shown that this shift, characterized by its abrupt and forced nature, led to negative effects concerning university students’ socialization as well (Hoss et al., 2021).

When online instruction is conducted in the context of a pandemic caused by COVID19, the emotional impact that this can have on university students can be particularly severe (Baloran, 2020). The pandemic has brought along “not only the risk of death from infection but also unbearable psychological pressure” (Cao et al., 2020, p.1). Research has shown that “[a]nxiety and isolation can be part of online learning”, while communication restrictions which are “inherent in online courses can reduce immediacy”, exacerbating feelings of loneliness (Reilly et al., 2012, p.100-4). Indeed, it has been found that university students’ exam-related anxiety levels have increased in the context of the coronavirus pandemic (Arora et al., 2020). Moreover, excessive use of technology in general “can negatively impact the physical, mental, emotional, and social health of youth” leading at times to addiction (Halupa, 2016, p.1).

It has been argued that “[e]ven after the COVID-19 pandemic is over, its repercussions likely will linger for years” (Kaplan-Rakowski, 2021, p.134), considering that medical scientists have indicated similarities between the psychological impact of the pandemic and that of post-traumatic stress disorder (Masiero et al., 2020; Tang et al., 2020). The extent of the validity of this hypothesis remains to be seen. In that context, the present paper aims to address the research gap pertaining to online learning in tertiary education in Greece during the pandemic. More specifically, it intends to investigate the emotional impact that online education has had on university students in the context of the Covid-19 pandemic.

3. The study

3.1. Research questions

The present study is constructed upon three basic research questions, namely:

1. What perceptions do university students hold towards online learning during the Covid-19 pandemic?
2. What kind of emotional impact does attending university classes online during lockdown have on university students?
3. What positive and negative aspects does online learning during the pandemic have according to university students?

To this end, a questionnaire was designed and distributed online to university students.

3.2. Participants’ profile

The questionnaire was answered by 83 students, 89% of whom were female, 8% were male, and the remaining 2% preferred not to state their gender. Almost half of the respondents, namely 49%, belonged to the age group of 18-22, followed by 22% who were aged 23-30, 16% who were older than 40 and 13% belonged to the 31-40 age group. The majority (65%) were Bachelor’s students, followed by 34% who were studying for their Master’s degree. Finally, 1% were in the process of obtaining their Ph.D.

In response to the question about the city and country in which their university is located, 62% of the students reported that they study at Aristotle University of Thessaloniki, followed by 37% who studied at the Hellenic Open University (online distance education), while the rest did not provide a clear response.

3.3. Instruments and data analysis

The questionnaires used for the purposes of this research were written in Google forms. Each questionnaire is comprised of five sections. The first section is comprised of profile-building questions in the form of multiple-choice. The second included multiple-choice questions related to aspects of online learning. In the third section, students were asked to provide answers to a series of statements presented in Likert scales regarding their perceptions towards online learning. Section four included more Likert scales and aimed at determining the emotional impact that learning online during the national lockdowns had on university students. Finally, the fifth section comprised a series of open questions allowing the respondents to share their views on online learning. Once the answers collection process was completed, data analysis was conducted using the Statistical Package for the Social Sciences (SPSS). The findings of the analysis will be presented in the next section.

4. Findings and discussion

The research showed that most students reported spending up to 5 hours per week doing university-related tasks, including preparing activities for attending classes and participating in student projects. Class attendance and material studying took up more of the respondents’ weekly time, as shown in the following chart:

Chart 1: Time spent online on university-related activities by university students
4.1. Research question 1: What perceptions do university students hold towards online learning during the Covid-19 pandemic?

Addressing the first research question related to students’ perceptions towards online university classes during the pandemic, the respondents were asked to state their level of agreement with a series of statements. Generally, students expressed a positive attitude towards online learning with more than half of them expressing agreement and strong agreement with issues such as their confidence with their technological skills, and the support they have received during their online learning experience, as can be seen in the following charts:

![Chart 2: University students’ confidence with their online learning skills](chart2.jpg)

Chart 2: University students’ confidence with their online learning skills

![Chart 3: University students’ perception of their ability to cope with online classes](chart3.jpg)

Chart 3: University students’ perception of their ability to cope with online classes

On the other hand, 30% expressed strong agreement that they would avoid online classes if they had the choice, and 35% of the students strongly agreed that they feel that having online classes is less satisfying than attending face-to-face classes. A chi-square test was performed to determine any association between the students’ reported hours spending attending online classes with their satisfaction levels. A statistically significant correlation emerged ($\chi^2(24, N=83) = 43.62$, $p = 0.008$), proving that the more hours university students spend attending online classes the less satisfied they probably feel with their learning experience.
Interestingly, on many occasions, the respondents avoided taking a clear position adopting a neutral point. That was the case with statements such as whether they have lost their interest since they started having classes online (30%), whether they feel bored in online classes (37%), if it is easier to be assessed online during exams (39%), if the online assessment is not a fair method of assessment (40%), if they use more strategies when learning online (35%), and if they have to make a lot of changes to adapt to online learning (33%).

4.2. Research question 2: What kind of emotional impact does attending university classes online during lockdown have on university students?

The second research question focused on the emotional impact of studying online during the lockdown. Students emphasized that their workload has increased significantly since they started studying online (35% agree, and 31% strongly agree). 52% of university students feel overwhelmed with the amount of time it takes them to study for their online classes and 49% feel bored when they study online.

An independent-sample t-test was conducted in order to compare the responses given by the students studying at the Aristotle University of Thessaloniki (AUTH) (M=3.60, SD=1.25) on the issue of whether they feel overwhelmed by their professors’ expectations with those given by the students studying at the Hellenic Open University (HOU) (M=2.53, SD=1.17). The latter has been offering almost exclusively online courses since its establishment. There was significant difference between the two groups (t-test=3.86, df=64.40, p=0.000<0.05). It appears that AUTH students who had not been used to online learning prior to the pandemic were more overwhelmed compared to their classmates at the HOU who had been exposed to online classes from the beginning of their studies.

41% of the respondents reported that they would feel better if their classmates had their cameras on while 34% disagreed, and 47% expressed a feeling that their relationship with their classmates has deteriorated since they started studying online although 29% disagreed with the latter statement. The turned-off cameras could be intensifying the feeling of loneliness that students experience while learning online. In order to determine whether there is an association between students’ reported feelings of loneliness and their motivation during online classes, a chi-square test was performed. A statistically significant correlation emerged (χ²(16, N=83) = 49.86, p = 0.000). This finding could indicate that the lonelier students feel the more their motivation to attend online classes is affected.

Another chi-square test was conducted to investigate any association between students’ reported feelings of stress to participate in online classes and their views on whether their performance has deteriorated. A significant correlation was found (χ²(16, N=83) = 50.09, p = 0.000), proving that the stress that students experience when they need to participate in online classes can be connected to their perceived academic performance deterioration. At the same time, the following chart shows university students’ struggle to balance their personal and university life amidst the pandemic:
An interesting finding in students’ responses was that they have not lost their trust in their professors, reporting confidence towards them during online learning (65%). Similarly, even though they claimed above that their relationships with their classmates have deteriorated, 43% reported that they have received useful advice from their classmates in the process of adapting to online learning, and 46% that they feel confident with the way they cooperate with them during online communication. Interestingly though, 37% held a neutral position on the first matter, and 31% were neutral on the second.

4.3. Research question 3: What positive and negative aspects does online learning during the pandemic have according to university students?

In response to the third research question, the respondents were first asked to share a pleasant surprise in the process of learning online. The variety of ICT tools and the skills that they were given the chance to develop were among the most popular answers given by the students. Many of them also addressed the communication and rapport that they managed to establish with both their classmates and professors thanks to online classes as a pleasant surprise, as highlighted in the following comments:

“The cooperation with classmates”
“How helpful the tutors are”
“Communication with the professor”
“Teacher-Student interaction”
“The support of a specific teacher”
“Meeting and working with colleagues I had never worked with before”
“A professor expressing themselves that they are grateful to our contribution and that the students really mattered to them and that they missed them”
“An unexpected bonding in one of the toughest courses in my school with both my classmates and the professor”
“Attitude of the teachers (extremely helpful / understanding)”

Some students also expressed how helpful online classes have been in boosting their confidence and alleviating their social anxiety:

“I do not feel lonely at all. In fact, it has facilitated the alleviation of my social anxiety and I get to participate much more, since I do not feel exposed.”
“I am naturally extremely shy and introverted and online learning has boosted my confidence a LOT. Also, I do enjoy staying at home a lot and spending time by myself, so I do not really mind it.”

A pleasant surprise to the respondents was also the convenience and flexibility that learning online entails, freeing them from time wasted (“Time gained from having to commute”, “I am never late and I can watch the whole lesson from the start”). It enables them to eliminate distances (“I can participate in contact session from another country”, “I could study from Halkidiki”), providing them with a more convenient learning environment and easier access to it (“I didn’t have to use the bus to go to the lectures”, “I can study at the comfort of my home”), and giving them chances to multi-task while attending lessons (“I could watch the lecture while at the same time eating”). Finally, the students were pleased with how online classes brought solutions to some of their practical problems. For instance, one of them referred to their chance to prove their capabilities thanks to online exams:

“My handwriting is very bad, and…during exam periods…instructors have given me a lower grade because they were not able to make out my lettering. That is not a problem now since we use keyboards…and now my grades are representative of my studying for the exams.”

On the other hand, among the respondents’ worst experiences with online learning were those related to connection and other technical issues which triggered feelings of frustration, exacerbated by the fact that they were at times forced to miss classes due to those issues. Technical challenges have been significantly associated with student dissatisfaction with online learning (Fawaz & Samaha, 2020). This is a point that was raised by 35% of the students in the present research.

Another serious issue that students had to deal with was stress and anxiety, with a lot of them mentioning that they had been experiencing headaches. Tertiary education students face several stress factors as they find themselves in a “transitional life-stage” navigating new environments and adjusting to adulthood (Othman et al., 2019, p.1). Their susceptibility to such stress factors is exacerbated particularly when unprecedented crises occur, such as a global health crisis. Research has concluded that online learning during the pandemic has “given rise to depression and anxiety disorders among undergraduate university students” with “depression, anxiety, and stress” being significantly correlated with student satisfaction (Fawaz & Samaha, 2020, p.52). A study conducted in China, a country that attracted a lot of negative attention especially at the onset of the pandemic, found that university students’ pandemic-related anxiety was associated with factors such as “their place of permanent residence, source of parental income, whether living with parents and whether a relative or an acquaintance was infected with COVID-19” (Cao et al., 2020, p.3).

In the present research, anxiety was closely connected to the online exams and the exhausting workload which, according to the participants, increased during online learning, though perhaps the presence of more stress factors was overseen or underplayed by the respondents. Attending online classes has been particularly stressful for some shy students who were having a hard time participating with their cameras on (“Anxiety, when I have to talk or turn on my camera even if I don’t want to”, “Being forced to have my camera on.” “The stress I have when I have to participate”). For some students, the increased workload was a cause of anxiety and frustration (“[S]ome professors saw online classes as an opportunity to assign many more projects than they would do under normal circumstances.”, “Truly overwhelming extra work”).

Oral presentations also became a common technique of evaluation utilized by many professors in an attempt to form a clearer opinion on their students’ progress. However, these presentations turned into a stress factor for several students, with one of them highlighting that “many professors chose to add presentations…which piled up to a lot more than what I could handle as a socially awkward person”.
Despite their familiarization with technology and their excitement to be exposed to new tools, as seen at the beginning of the questionnaire, transitioning from face-to-face to online exams caused significant distress on students who at times felt at a loss of control due to the possibility of unbalanced factors such as technical issues jeopardizing their performance and hard work. It has even been found that university students’ exam-related anxiety is greater than the anxiety caused by the pandemic and the constant threats to their physical health (Arora et al., 2020). It appears that online exams could seem more stressful to university students than a deadly pandemic. The following comment given by one of the participants attests to this finding:

“[O]nline exams are so stressful because we are not given enough time to think or even read the whole question … and we have to think of the connection or if the laptop will make it until the end of the exam because we overuse it or of any noise in the house as well or anything else that can happen during exams because we are at home.”

University students’ mental health appears to be compromised by the pressure caused by the lack of adequate time during online exams. A respondent described their experience having a panic attack during their online exams, as well as the aftermath of their poor performance:

“It was during the exams of the winter semester, … the instructor gave us so little time to answer so many questions. I panicked, I got extremely anxious … I wish that instructor knew about my panic attack and gave me a second chance.”

Another student elaborately addressed their poor mental health conditions highlighting the lack of support that they received from their professor.

“Once a professor realised that my mental state was wrecked and saw the visible exhaustion and burnout on my face and performance. But in a personal meeting that professor still condemned my work as not enough and advised me to try harder and try harder to put priorities in my life…I was barely hanging to maintain some kind of health and not pass out sick.”

Social isolation and distancing, along with the insecurities and fears induced by the pandemic, sometimes make it hard for people to share their hardships or empathise with those who get the courage to be open about their mental and emotional struggles. However, “prioritizing the wellness of students…may be a desirable modification that educators should address…during times of…mental stress that accompany the pandemic” (Kaplan-Rakowski, 2021, p.134).

In a similar context, lack of interpersonal relations was also discussed by students as being part of their bad online learning experience as one of the respondents emphatically stated “[w]henever the internet connection dropped, I couldn’t ask what we were doing because I didn’t know who my classmates were”. It would seem that breakout rooms may have eliminated potential communication barriers and could have given students a chance to get to know each other and feel part of the community. However, their answers indicate the opposite, as many respondents highlighted their struggle to collaborate with their classmates in breakout rooms, due to the latter’s reluctance to participate actively: “the majority of them cannot communicate and co-operate when needed. They simply chose not to talk at all, which for me is at least disrespectful”.

Finally, the most pessimistic comments brought forward the feelings of boredom (“online classes are boring anyway, so there was nothing that could make online learning worst”), isolation (“[I was] feeling lonely and isolated because of cameras being turned off”), and loss of interest (“[I have lost my interest in classes”]. Being in a state of quarantine for excessive periods has indeed been found to
impact individuals’ mental health, with the aforementioned feelings being among the most reported (Brooks et al., 2020).

The participants’ responses to the question regarding what the greatest challenge learning online had been, were quite similar to those discussed previously. The issues of coping with online exams, communication issues, and the increased workload, along with the struggle to “[stay] motivated and not [feel] lonely” were brought up many times. The respondents also addressed some practical challenges such as having to remember the different links to their classes or “having to look for the zoom link each time” and to study online material instead of “physical copies of books”, or health-related issues connected to the sedentary lifestyle they were forced to adopt ("I…faced some problems with my neck and back”, “too much screentime/ back pain”, “eyes hurting”).

Maintaining concentration despite the various distractions was a major challenge for several respondents, as shown by the following comments:

“When you are on your computer you are always distracted and tempted to just surf on the internet and look up random things.”
“I get easily distracted because I am at home.”
“[A]ttending classes when there’s so much going on at home”
“[T]oo many things at home that distract you”

Indeed, other studies confirm the students’ views emphasising that the social isolation caused by the quarantine is closely connected to “procrastination and feelings of worthlessness”, while it has been found that anxiety is correlated to feelings of low self-efficacy (Fawaz & Samaha, 2020, p.53; Arora et al., 2020).

4.4. Other interesting findings

Overall, the various challenges that university participants needed to face during online learning could be summarized in the following comment expressed by one of them: “Combining a ‘real’ life with an online student presence”. Indeed, attempting to balance those two aspects of their lives “while keeping a good mental and physical health (sic)” has affected students’ motivation, stress levels, and physical well-being.

In response to the question “what do you feel you are missing when you are learning online?”, interaction and socialization were the most frequent answers given by the respondents demonstrating the negative socio-emotional impact of online learning. One of them commented that they are missing “[t]he university experience”, something which for some respondents takes the form of “hanging out in the library and borrowing books”, or “the big rooms, the hustle and bustle of the university, the library, meeting people”. Another form of the “university experience” was given by a respondent who admitted having lost their motivation as a result of online learning: “I was going to university classes so I could meet my friends and fellow students after the class. Since I’ve lost this motivation, I really skip classes”. These findings are in accordance with another study that concluded that “exclusive online instruction and learning methods have rendered the students dissatisfied with their learning experience” (Fawaz & Samaha, 2020, p.55).

Despite the challenges and unpleasant experiences, more than half of the respondents were in favour of being taught with the use of a blended curriculum in the future, combining the freedom and flexibility brought by online learning, along with the chances for communication and development of interpersonal relations characterizing in-class learning. This finding is in contrast with research conducted among Filipino university students, the majority of whom were opposed to the
introduction of a blended curriculum in their academic institutions (Baloran, 2020), an opinion shared by Polish university students as well (Rizun & Strzelecki, 2020).

In the final section of the questionnaire, the students were offered the opportunity to share any thoughts on the emotional aspect of online learning they desired. The responses were admirably powerful and honest, varying from total disregard towards online learning to great enthusiasm for the experience. For some students, online learning during the pandemic is “[a]wful, unsupportive, lacking, stressful”, and “tiring, mind-numbing and not very helpful”. It “is rather impersonal and makes you feel detached”, “lonely… disconnected and very angry”. The most dramatic of the respondents mentioned that “I have lost every inch of motivation” or that “[online learning] has wrecked me emotionally”.

For others, online university classes were a positive experience, helping them feel safer, as shown in the following excerpts:

“Being a person with social anxiety, online learning offers some kind of relief. But, it is also extremely stressful in a different kind of way.”
“Online learning is an ideal and less stressful way of learning for busy people.”
“It has helped me so much! People tend to neglect us anxious students a lot and they do not seem to understand how draining human interactions can be.”
“I feel more safe at home rather than at campus. Certain colleagues of mine have bullied me in the past, so I do not even want to imagine what might happen should we return back to normal.”

These are the findings of this study which by no means can be representative of the entire Greek university population. The results, however, address main issues raised regarding online learning and indicate the emotional and socio-emotional impact that online learning in combination with the pandemic COVID19 has had on university students.

5. Conclusion

Looking at online education during the Covid-19 pandemic through the “pandemic vs. education” dilemma (Rizun & Strzelecki, 2020, p.1) is a rather nihilistic approach that fails to do justice to the potential that online learning entails and the opportunities it has offered to university students. Undoubtedly, the pandemic has been a challenge to “the entire educational system”, causing a tremendous impact on students’ mental health and raising obstacles to their struggle for emotional and mental well-being (Arora et al., 2020, p.12). Nevertheless, “in hindsight, [it] has opened opportunities for universities to revamp their content delivery” and re-direct their attention towards the “emerging cloud-based technologies” (Arora et al., 2020, p.12-13).

The present research has aimed to investigate the emotional impact that learning online during the Covid-19 pandemic has had on university students. The findings highlighted that online learning has aggravated students’ stress and frustration as a result of the proliferation of online exams and increasing workload. Students’ insecurities were closely connected to the potential issues caused by technical malfunctions, while distance communication caused feelings of isolation. The study, however, gave voice to a group of the student population for whom online learning has been a relief from social anxiety, while for many others it has been a great opportunity for exposure to new tools and learning environments.

The small number of participants is the main limitation of the research. The authors also recognise that the emotional impact that the respondents attributed to online learning should not be located only in the teaching mode but could be the result of a number of factors – personal, social, financial,
health-related, and others – affecting their emotional state. Similarly, the connections reported by the respondents between online classes and their emotional impact may not always be causal, but rather temporal.

Finding ways to overcome the obstacles and highlight the methods which can render online learning a positive experience for university students should be the aim of future research. In this way, the pandemic vs. education dilemma will be part of the past giving way to a “pandemic and (online) education” future.

References


**Thomaï Alexiou** (thalexiou@enl.auth.gr) is an Associate Professor at the Department of Theoretical and Applied Linguistics, School of English, Aristotle University of Thessaloniki, Greece. Her expertise is in early foreign language learning, methodology of teaching languages, vocabulary and material development for young and very young learners. She has authored/co-authored books for children learning English as a foreign language as well as online resources in the areas of CLIL (CLIL-Prime), learner strategies (‘Thalis project’), dyslexia (DysTEFL2) and EFL for young learners (PEAP). She has co-edited and co-authored ‘Magic Book’ 1&2. As of 2021, she’s the Academic Leader of the ‘EAN’
project, a funded innovative project for the introduction of EFL in state pre-primary schools in Greece.

Theologia Michalopoulou (thelmamich@gmail.com) is a graduate of the School of English, Aristotle University of Thessaloniki. She holds a BA in English Language and Literature and an MA in Language, Literature, and Digital Media in Education. She is currently an MA student in the Department of Balkan, Slavic, and Oriental Studies, University of Macedonia, working on her thesis on human rights education in the EFL class using movies. She has been working as an EFL teacher in Secondary Schools, as well as online language schools based in the UK, Spain, and China. Her research interests include online learning, human rights education, and ICT tools.
The present experimental study forms part of the research investigating whether strategy instruction improves learners’ reading comprehension and develops their metacognitive awareness (Grabe & Stoller, 2011; Zhang et al., 2014). More specifically, research involving explicit strategy instruction among two homogeneous 16-member groups of experimental and control learners was set up in a Greek State High School. The intervention followed the gradual release of responsibility model involving prediction, previewing, skimming, graphic organizers, visualization, making connections and summarization strategies. The experimental group engaged in all of the stages of the intervention, gradually becoming independent users of the strategies taught. The control group engaged only in the first stage. The findings indicate a clear advantage for the experimental group whose ability to comprehend a text and construct meaning improved significantly following the effective implementation of the strategies taught (Bouchard, 2005; Harrison & Vallin, 2018; Harvey & Goudvis, 2014; Weir et al., 2009). At the same time, their autonomy was enhanced, since they developed a deeper understanding of how to manipulate the reading process (Flavell et al., 2002). Therefore, it appears that reading strategy instruction results in developing strategic readers able to utilize the strategies needed to facilitate understanding and monitor comprehension (Clarke et al., 2014).

**Key words:** reading comprehension, metacognitive awareness, reading strategies, reading strategy instruction

---

**1. Introduction**

Reading is an essential skill in the foreign language classroom. The ultimate goal of reading instruction is to enable learners to construct meaning while reading a text.

Successful reading comprehension is generally thought to entail drawing relevant information from the written text effectively through the employment of reading strategies and the reader’s background knowledge (Rativa et al., 2012; Rios & Valcarcel, 2005). The written text does not carry
meaning by itself. Readers engage actively in the reading process, interacting with the text, which provides directions as to how they should retrieve or construct meaning using their prior knowledge (Brown, 2001). To comprehend a text, readers should not only sample it, they should also try to reflect on the reading process by looking at it from the outside (Flavell et al., 2002).

In an attempt to add to the work done on metacognitive strategy development and reading competence, the present article presents an action research conducted in an English as a Foreign Language (EFL) class at a Greek state school, focusing on cognitive and metacognitive strategy instruction. After presenting the relevant literature and setting the theoretical framework for the employment of reading strategies in reading strategy instruction, I proceed to the methodology employed and the presentation and discussion of the research findings.

2. Theoretical Framework

2.1 Reading

2.1.1 The cognitive view and schema theory

The cognitive theory of reading or top-down processing emerged in the 1960s and had a great influence on the EFL field (Pardede, 2010). According to this method, reading involves the connection of information in a text with the reader’s background knowledge (Celikoz et al., 2016). In this sense, reading is an active, cognitive process on the part of the learner whose background knowledge plays a key role in constructing meaning through sampling a text, making hypotheses, confirming or rejecting them, making new hypotheses, and so forth (Asnita, 2013).

Schema theory is closely related to top-down processing. Its fundamental principle is that written text provides directions for readers as to how they should retrieve or construct meaning from their own previously acquired knowledge. New knowledge is built upon past knowledge of objects, situations, events and procedures of how to retrieve, organize, and interpret information (Carrell & Eisterhold, 1988). Therefore, comprehension results from old knowledge interacting with new knowledge as new information is added to the reader’s system through restructuring his/her existing schemata (Brown, 2001) namely, a “reader’s existing knowledge” ... “packed into units” (Fahriany, 2014, p.17).

Schema theorists distinguish between formal schemata (i.e., the way a text is structured according to its genre) and content schemata (i.e., knowledge about the subject matter of a text) (Carrell & Eisterhold, 1983). Employing background knowledge of both schemata through specific strategies enables readers to predict events and infer meaning from a broader context (Landry, 2002).

According to schema theory, reading entails a constant interplay on two levels (Anderson et al., 1977; Brown, 2001; Carrell & Eisterhold, 1988; Yang, 2013). The first level involves the text and the reader. Regarding the text, in order for L2 reading instruction to be effective, teachers should expose learners to meaningful texts, relevant to their needs, interests and cultural background in order to help them activate pre-existing schemata or build new ones (Nuttall, 2000). The reader is the one responsible for decoding the message and his/her prior knowledge is of paramount importance since comprehension occurs when new information from the text interacts with the reader’s old knowledge (Asnita, 2013). To that end, reading instruction should follow the pedagogical pre-, while-, and post-reading stage framework to make sure that learners acquire the relevant schemata for text comprehension, to enhance the interaction between reader and text and to give learners the chance to evaluate their adequacy of interpretation and develop their metacognitive awareness (Grabe & Stoller, 2011; Hedgcock & Ferris, 2009).
The second level involves an interaction between bottom-up and top-down processing (Carrell & Eisterhold, 1983). In bottom-up processing, learners rely on specific data from the text while in top-down processing, they rely on higher-level schemata to make predictions and seek to confirm them through searching on a more specific level (Yang, 2013). As suggested in the literature (Birch, 2015; Grabe & Stoller, 2011), reading comprehension results from the interaction between surface structure and deep structure processing, validating both methods of understanding as well as the strategies that learners use to gain meaning from the text.

2.1.2 The metacognitive view

Since studies have demonstrated that successful reading comprehension depends on a directed cognitive effort (Cakici, 2017), research in the field probes into the control readers exercise while trying to comprehend a text (i.e., metacognition) (Flavell et al., 2002; Grabe & Stoller, 2011). Employing metalinguistic processes helps foreign language learners become aware of the way the new linguistic system works to convey meaning, so that they can solve comprehension problems (Clarke et al., 2014).

Metacognitive knowledge consists of knowledge about cognition and regulation of cognition, allowing the reader to plan, regulate and monitor the reading process. Knowledge about cognition entails recognizing patterns of structure, and organization. Monitoring of cognition involves recognizing information problems in a text or the lack of ability to achieve comprehension (Meniado, 2016; Williams & Burden, 2008).

2.2 Reading comprehension strategies

Reading comprehension strategies involve readers in making deliberate decisions before, during, and after reading in order to make sense of their reading (Brown, 2007; Oxford, 1990). Research has indicated that providing learners with explicit comprehension strategy instruction facilitates reading comprehension and enhances their autonomy (Ediger, 2001; Gooden, 2012).

Comprehension strategy instruction involves explicit teaching of reading comprehension strategies through modeling, scaffolding, guided practice, and independent use of strategies aiming at increasing learners’ understanding of a text as well as monitoring and repairing their own comprehension (Block et al., 2002; Van Keer, 2004). Urquhart and Weir (2013) emphasize that the distinguishing feature of strategies is that they are intentional. Their intentional nature makes them teachable. Gradually, these conscious processes become unconscious and automatic and turn into skills (Urquhart & Weir, 2013).

Harvey and Goudvis (2014) argue that teaching reading strategies should aim at enabling learners to move from the tacit level of understanding, in which they are not aware of how they think when they read, to a higher level of understanding how to manipulate the reading process. In other words, tacit readers should become strategic readers that is, readers able to utilize the reading strategies needed to facilitate understanding and to monitor comprehension. Iwai (2016) points out that comprehension strategy instruction is most effective during actual engagement with the text. As suggested in the second language acquisition literature (Brown, 2007), reading strategies are learning strategies since reading is a receptive skill. The learning strategies most frequently employed in reading and the ones employed for the purposes of the present study are cognitive and metacognitive strategies.
Cognitive strategies are mental processes requiring specific actions and goal-oriented cognitive steps employed by learners in comprehending a text (Williams & Burden, 2008). Cognitive strategies relate to specific contexts and learning tasks and involve direct manipulation of the learning material (Brown, 2007).

The cognitive strategies discussed in this article are: a) Prediction. Bouchard (2005) suggests that prediction is a cognitive strategy, which can improve reading comprehension. When learners make predictions, which confirm or disconfirm hypotheses while reading, they set a purpose for reading, and interact more with the text, increasing their interest and improving comprehension (Oczkus, 2003). b) Skimming, involving rapid reading aiming at quickly identifying the main ideas from a text (Harmer, 2007). Obtaining an overview of the text helps learners read in a more focused and efficient way (Weir et al., 2009). c) Using graphic organizers namely, a schematic representation of information in a text using key vocabulary terms whose purpose is to activate students’ prior knowledge and relate it to new material to make it more familiar and meaningful (Alvermann, 1981; Jiang & Grabe, 2007).

During reading, readers employ metacognitive strategies purposefully and intentionally in order to plan for learning, monitor comprehension or production and self-evaluate what has been learnt (Zhang et al., 2014). Since the ultimate goal of strategy instruction is to develop learners’ ability to use strategies whenever necessary, metacognitive strategy instruction ensures that learners will be able to use these strategies well into adulthood (Meniado, 2016).

The metacognitive strategies employed for the purposes of the present study are: a) Pre-viewing text (i.e., engaging learners in planning the cognitive process through generating questions, and monitoring the cognitive process through seeking the answers, thus building metacognitive awareness) (Bouchard, 2005). b) Visualization, which requires readers to use their senses and create an image of what is read so as to make sense of it (Mokhtari & Reichard 2002; Moreillon, 2007). Learners own a great repertoire of visual images from their everyday experience. Using these images to visualize what is happening helps them remember and retrieve information, monitoring and improving comprehension (Harvey & Goudvis, 2014; Sadoski & Paivio, 2001). c) Making connections engages readers in activating background knowledge related to a topic and connecting it to the text to enhance understanding and repair meaning (McNamara, 2007). Good readers make three types of connections: text-to-self (including connections between the reader’s experience and the text), text-to-world (connections between the text and information about the world), and text-to-text (connections between types of texts regarding content, plot, structure and style) (Waller & Barrentine, 2015). d) Summarization, which requires identifying the most important information in a text. Summarizing what one has learnt is an important part of evaluating one’s performance resulting in building metacognition through regulating one’s cognition (Harrison & Vallin, 2018; Mokhtari & Reichard 2002).

According to Veenman, Van Hout-Wouters, and Afflerbach (2006), it is not easy to distinguish between cognitive and metacognitive strategies since metacognition is knowledge about cognition. In effect, metacognition is dependent on cognition and “metacognitive planning cannot be used without cognitive activities referring to the task at hand” (Mehrdad et al., 2012, p. 3758). In a similar vein, Kasimi (2012) points out that metacognitive reading strategies are sequentially applied by learners in order to control the cognitive activities whose ultimate goal is text comprehension. Consequently, both cognitive and metacognitive elements are required to construct understanding.
3. Research methodology

3.1 Research objectives and nature of the study

The aim of the research was to investigate whether the employment of reading strategies improves learners’ reading comprehension and develops their metacognitive awareness in order to enable them to monitor and repair their own comprehension.

The method selected for the purposes of the present study was action research which allows the teacher to gain a deeper understanding of her practice and improve it through using her own instruments to collect, interpret and present data on a cyclical basis which included planning, action, monitoring and reflection (Richards, 2003). At the planning stage, the teacher-researcher decided and planned what needed to be investigated. The action followed was monitored by the teacher-researcher. Finally, the teacher-researcher reflected on the outcome in order to improve her practice.

3.2 Teaching context and participants

The participants of this study were 32 monolingual Greek learners of the waystage/A2 level according to the Common European Framework (Council of Europe, 2001) attending EFL classes in the first grade of Junior High School.

The learners were divided into two homogeneous 16-learner groups: the control and the experimental one. The experimental group participated in the training programme, which lasted a month and was composed of four stages: a) explicit strategy instruction through modelling, b) guided practice, c) collaborative learning, and d) independent practice. The control group engaged only in the first stage of the training programme.

3.3 Data-collection instruments

The research adopted a mixed-methods approach, integrating both qualitative and quantitative data, in order to provide a deeper understanding of the research problem (Teddie & Tashakkori, 2011) as well as “evidence for the validity of research outcomes” (Dörnyei, 2007, p. 45). To that end, both qualitative and quantitative data-collection instruments were designed and employed by the teacher-researcher.

3.3.1 Qualitative research

Pre-, and post-intervention semi-structured interviews were conducted providing the researcher with a comparable set of questions but also allowing participants’ perspectives on the issue to unfold since they could raise issues that had not been pre-planned (Adel et.al., 2015). Since qualitative research interviews are time consuming, a representative sample of learners -8 learners from the control group and 8 learners from the experimental group- was randomly selected to participate. The interviews were conducted in Greek in order to establish a relaxing atmosphere. The pre- intervention interviews primarily investigated the learners’ interest in reading texts in English in order to assist the researcher in selecting the appropriate authentic materials for the tests and the training sessions. Then, they probed into the learners’ attitudes toward reading in English and their learning styles in order to enhance their autonomy and allow the researcher to adjust the training programme to their learning styles. Finally, the learners’ use of and familiarity with cognitive and metacognitive reading strategies were explored.
The post-intervention interviews investigated whether the training programme was effective in assisting the learners in employing the strategies they had been taught.

The second qualitative data-collection instrument employed for the purposes of the present study was observation which helped the teacher-researcher to triangulate the findings derived from the interviews and the tests in order to get insight into what learners really do, instead of relying on what they say they do alone (Dörnyei, 2007). Collecting observational data in a more organized and structured way makes the findings more reliable and comparable (Kawulich, 2012). To that end, observation checklists were designed by the teacher-researcher.

The teacher’s observation checklists, designed in the form of five-scale Likert statements ranging from ‘never’ to ‘always’, were employed in every lesson of the training programme. Each observation checklist focused on the cognitive and metacognitive strategies taught. Concerning prediction, the teacher-researcher had to observe the learners’ ability to use headings, subheadings, pictures and the structure of the text in order to elicit background knowledge and predict content and type of text (Quiroga, 2010). In relation to previewing, the teacher-researcher observed whether the learners could both form questions using the information derived from headings, subheadings, pictures and the first sentence of each paragraph and find the answers to the questions formed (Bouchard, 2005). Regarding skimming, the teacher detected whether the leaners were able to get the general idea of a text and confirm or disconfirm their initial guesses through reading headings, subheadings, the introduction, the first sentence of each paragraph, and the conclusion of a text (Asmawati, 2015). Proceeding to graphic organizers, the learners’ ability to elicit background knowledge and relate it to new knowledge as well as to organize the information in the text through formulating the main idea of the text and showing its relationship to the supporting details was observed (Hall-Kenyon & Black, 2010; Manoli & Papadopoulou, 2012). Concerning visualization, the teacher-researcher observed whether the learners were able to use their senses to create mental images while reading as well as their ability to draw conclusions or recall details using those images (Harvey & Goudvis, 2014). With regard to making connections, the teacher-researcher had to detect the learners’ ability to elicit background knowledge and connect the information in a text to their lives, the real world and other texts (Wahyuni, 2016). Finally, regarding summarization, the teacher had to observe whether the learners were able to distinguish important from less important information and present it in their own words (Harrison & Vallin, 2018).

### 3.3.2 Quantitative research

A pre- and a post-instruction test were constructed in order to check improvement in the learners’ reading comprehension ability and metacognitive awareness (Jang, 2009; Torgesen, 2005). The two tests followed an identical format in order to be comparable. Each test consisted of a pre-, a while-, and a post-reading stage. The pre-reading stage examined the learners’ ability first to predict content and type of text and then to plan the reading process through previewing the text.

At the while-reading stage, the learners were first asked to identify the main idea by skimming through the text and then to find the answers to the questions they had formed at the pre-reading stage. The next task engaged the learners in organizing the information in the text visually through a graphic organizer. The final task checked the learners’ ability to present the mental images they had created while reading.

At the post-reading stage, the learners’ ability to connect the information in the text to their life, the real world as well as other texts was checked. Finally, the learners engaged in summarizing what they had learnt from the text so that they could check their ability to identify the most important information in the text and present it in their own words.
3.4 The training sessions

The training sessions followed the gradual release of responsibility model allowing the teacher to move from teacher knowledge to student understanding (Frey & Fisher, 2007). In the first stage, the strategies were introduced through posters displayed on the interactive board. The learners carried out the tasks themselves and, then, they discussed their answers with the partner sitting next to them. The teacher-researcher acted as a “participant observer” who acted “as a full member of the group” (Dörnyei, 2007, p. 179) and completed the observation checklists in order to identify the learners’ weaknesses in applying the strategies. Explicit strategy instruction through modelling followed, where the teacher employed the think aloud technique. Through the think aloud technique the teacher verbalized her thoughts, explained how the strategies are used and cleared up confusion when it occurred while the learners followed the teacher’s flow of thought and learnt how the strategies are employed, constructing meaning collaboratively (Macaro, 2001; Ys et al., 2018).

The second stage involved guided practice, where the teacher acted as a guide and facilitator gradually releasing task responsibility to the learners, on the one hand, but, on the other hand, providing instructional scaffolds to ensure the students’ success. The learners carried out the tasks themselves and discussed their answers with the partner sitting next to them. The teacher-researcher was a “participant observer” (Dörnyei, 2007, p. 185) and completed the observation checklists.

The third stage engaged the learners in collaborative learning through applying the strategies while collaborating in groups. This stage allowed the learners to consolidate their understanding before they applied it independently in the next stage. The teacher-researcher acted as a “nonparticipant-observer” (Dörnyei, 2007, p. 179) recording the learners’ progress on the observation checklists.

Finally, the fourth stage involved independent practice where the learners applied the strategies on their own, employing the “think aloud” technique, which allowed them to distinguish between reading words and constructing meaning and so developing their metacognitive awareness (Ys et al., 2018). Once again, the teacher-researcher’s role was that of a nonparticipant-observer, recording the learners’ successful application of the strategies on the observation checklists.

Regarding the teaching process, each of the stages of the training programme included two lessons, which followed the pre-, while-, post- reading stage framework in order to activate and build learners’ schemata (Toprak & Almacıoğlu, 2009; Harmer, 2007). Following social-constructivist pedagogy, which posits that knowledge is constructed through interaction with others (McKinley, 2015; Vygotsky, 1978), all the lessons engaged the learners either in pair or in group work.

The texts for all the lessons were authentic materials, that is “ordinary texts not produced specifically for language teaching purposes” (Carter & Nunan, 2001, p.68), selected based on the learners’ interests elicited through the pre-instruction interviews and class discussion. Employing authentic materials can be effective in many ways since they expose learners to real language and to language input used for genuine communication purposes in the target language (Guarente & Morley, 2001; Nadrag & Tihenea, 2017), and allow them to deal with complete messages in a small amount of print (Nunan, 2005). Furthermore, authentic materials help learners contextualize language learning (Nadrag & Tihenea, 2017), and positively affect their motivation and engagement in the reading process (Guo, 2012) while they connect the classroom to the outside world (Nadrag & Tihenea, 2017). The selection of the authentic materials followed Nuttal’s suggested criteria for classroom employment (as cited by Azri & Rashdi, 2014, p. 251): suitability of content (the degree to which a text is interesting to learners and relevant to their needs), exploitability (the way the text is exploited to develop learners’ competence in reading), and readability (a text’s difficulty in terms of
structures, grammar and vocabulary). Regarding readability, Krashen’s Input Hypothesis (1985), which advocates that learners should be exposed to language slightly above their current linguistic level in order for acquisition to take place, was also taken into account. In addition, following Berardo (2006), the learners were exposed to different types of texts to elicit their interest and increase motivation presented in an authentic context, using pictures, tables, photos and so on in order to make them more appealing to the learners. Finally, the texts were retrieved from the Internet, since authentic materials accessed on the Internet are continuously updated, and, visually, more appealing (Berardo, 2006) while they connect learning to learners’ real lives since the Internet forms part of their daily life (Herrington et al., 2014).

4. Research findings

4.1 The qualitative findings

Sixteen learners of the first grade of Junior High School were interviewed before and after the intervention. Based on Dörnyei (2007), the qualitative data was given a textual form through a language-based analysis.

The pre-intervention interviews revealed that most of the learners from both groups did not systematically employ reading strategies. They approached an authentic text by focusing on the recognition and recall of lexical and grammatical forms. In effect, both groups displayed low interest and a negative attitude towards reading in English due to the disappointment, stress and boredom they felt when they could not recognize new words and grammatical structures. Consequently, the majority of learners from both groups were readers of moderate ability since their reading comprehension was impeded by their weakness in vocabulary and grammar.

Concerning learning styles, the learners exhibited diversity in the way they enjoyed reading a text, since the majority preferred reading silently on their own but there were also learners who liked reading in a group, reading aloud and listening to the text.

Finally, both groups read about a variety of topics, such as music, sports, short stories, the environment, the news, and teen columns related to teenagers’ problems.

After the training programme, most of the learners from the control group pointed out that there was no change in the way they approached a text. However, there was a change in the way the experimental group approached a text, since, as the learners pointed out, the programme had allowed them to learn how to apply the strategies taught, which helped them comprehend a text and monitor their comprehension. Finally, a few learners stated that they would like more practice in the previewing and summarization strategy.

Proceeding to the teacher’s observation checklists, the data obtained allowed the triangulation of the findings deriving from the interviews, and the tests regarding strategy use. More specifically, during the model lessons, the majority of learners from both groups exhibited a major weakness in foregrounding background knowledge and applying the strategies taught. However, in the guided and collaborative practice lessons, the experimental group, which engaged in all of the stages of the programme, displayed a
deeper understanding and a significant improvement in their ability to use the strategies. Finally, the data obtained from the independent lessons showed that all the learners in the experimental group exhibited remarkable progress, since they were able to work entirely on their own without any interference from the teacher and apply the cognitive and metacognitive strategies taught effectively.

4.2 The quantitative findings

A summary of the focal points of the data obtained through the pre- and post-intervention tests is presented separately for the control and the experimental group. The paired samples t-test (Dörnyei, 2007) and the independent samples t-test (Dörnyei, 2007) were employed to investigate improvement in the learners’ reading comprehension ability. The Wilcoxon test for dependent samples (Dörnyei, 2007) and the Mann Whitney test for independent samples (Dörnyei, 2007) were employed to investigate the development of the learners’ metacognitive awareness.

With regard to the findings, as Table 1 below reveals, there was no significant statistical difference regarding the reading comprehension scores of the control group [t (15) = 1.546, p = .143] at the pre- and the post-instruction test. Furthermore, there was no significant difference, statistically, between the two groups’ performance at the pre-instruction test [t (30) = .648, p = .522]. However, the experimental group achieved a significantly higher post-instruction test reading comprehension score compared to the pre-instruction test [t (15) = -17.757, p = .000]. Therefore, although the starting point was the same for both groups, there was a statistically significant difference between their reading comprehension scores at the post-instruction test [t (30) = -15.655, p = .000]. What the above analysis suggests is that strategy instruction improves learners’ reading comprehension:

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre - instruction test</td>
<td>16.38</td>
<td>14.53</td>
</tr>
<tr>
<td>Post - instruction test</td>
<td>12.53</td>
<td>66.75</td>
</tr>
</tbody>
</table>

Table 1. Differences between the control and the experimental group’s pre- and post-instruction reading comprehension test scores

With regard to metacognition, our results indicate a clear advantage for the experimental group. Therefore, as displayed in Table 2, there was no significant statistical difference between the pre- and the post-instruction tests scores of the control group (z =-.877, p = .380) concerning the previewing strategy. By contrast, the experimental group exhibited a higher post-instruction test score in the previewing strategy compared to their pre-instruction test score (z =-3.520, p = .000). Furthermore, while there was no significant statistical difference between the two groups at the pre-instruction test (U = 127.500, p = .982), there was a statistically significant difference regarding the employment of the pre-viewing strategy at the post instruction test (U = 2.000, p = .000), as expected, suggesting that metacognitive strategy instruction develops learners’ metacognitive awareness.

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre - instruction test (Previewing)</td>
<td>1.56</td>
<td>1.00</td>
</tr>
</tbody>
</table>
metacognitive awareness can be developed through metacognitive strategy instruction.

Table 2. Differences between the control and the experimental group’s pre- and post- instruction tests scores in relation to the ‘previewing’ strategy

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Experimental</th>
<th>Control</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-instruction test (Previewing)</td>
<td>1.19</td>
<td>2.14</td>
<td>12.31</td>
<td>4.73</td>
</tr>
</tbody>
</table>

Proceeding to the visualization strategy, the data analysis, as presented in table 3, once again shows no significant statistical difference between the pre- and the post- instruction test scores of the control group (z=-.276, p=.783). On the other hand, however, the experimental group achieved a significantly higher post- instruction test score (z=-3.526, p=.000). So while there was no significant statistical difference between the two groups in the pre-instruction stage (U=101.000, p=.323), there was a statistically significant difference between the two groups’ performance in the post instruction test (U=1.000, p=.000). The results confirm the hypothesis that learners’ metacognitive awareness can be developed through metacognitive strategy instruction:

Table 3. Differences between the control and the experimental group’s pre- and post- instruction tests in relation to the ‘visualization’ strategy

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Experimental</th>
<th>Control</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-instruction test (Visualization)</td>
<td>.56</td>
<td>.81</td>
<td>1.00</td>
<td>1.15</td>
</tr>
<tr>
<td>Post-instruction test (Visualization)</td>
<td>.63</td>
<td>.72</td>
<td>7.41</td>
<td>2.15</td>
</tr>
</tbody>
</table>

Table 4 exhibits a great difference in the employment of the ‘making connections’ strategy between the two groups. More specifically, the control group achieved a higher pre-instruction test score in the strategy compared to the post-instruction test score (z=-2.207, p=.027) while the experimental group achieved a substantially higher post- instruction test score than pre-instruction test score (z=-3.519, p=.000). Moreover, while there was no significant statistical difference between the two groups’ performance in the pre-instruction test (U=124.500, p=897), there was a significant statistical difference between them in the post instruction test (U=0.000, p=000), implying, once again, that metacognitive strategy instruction can be effective towards developing learners’ metacognitive awareness.

Table 4. Differences between the control and the experimental group’s pre- and post- instruction tests in relation to the ‘making connections’ strategy

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Experimental</th>
<th>Control</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-instruction test (Making connections)</td>
<td>1.40</td>
<td>1.81</td>
<td>1.25</td>
<td>1.39</td>
</tr>
<tr>
<td>Post-instruction test (Making connections)</td>
<td>.25</td>
<td>.77</td>
<td>7.66</td>
<td>1.76</td>
</tr>
</tbody>
</table>

Finally, as demonstrated in table 5, there was no significant statistical difference between the pre- and the post-instruction tests scores of the control group (z=.000, p=1.000) regarding summarization. However, the experimental group’s post-instruction test scores in the strategy were substantially higher than the pre- instruction test scores (z=-3.500, p=.000). Furthermore, while there was no significant statistical difference between the two groups in the pre-instruction test (U= 120.000, p=.780), a significant statistical difference was observed in the post-instruction test (U= 0.000, p=.000) in terms of the effective employment of the strategy, suggesting, once again, that metacognitive strategy instruction enhances the development of learners’ metacognitive awareness:

Table 5. Differences between the control and the experimental group’s pre- and post- instruction tests in relation to the ‘summarization’ strategy

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Experimental</th>
<th>Control</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-instruction test (Summarization)</td>
<td>.00</td>
<td>.00</td>
<td>.44</td>
<td>1.75</td>
</tr>
<tr>
<td>Post-instruction test (Summarization)</td>
<td>.00</td>
<td>.00</td>
<td>5.87</td>
<td>1.63</td>
</tr>
</tbody>
</table>
Table 5. Differences between the control and the experimental group’s pre- and post- instruction tests in relation to the ‘summarization’ strategy

5. Discussion of the findings

The findings obtained from both the qualitative and quantitative analysis confirm the hypothesis that the employment of reading comprehension strategies improves learners’ reading comprehension and develops their metacognitive awareness.

Regarding the use of strategies, the analysis of both the quantitative and the qualitative data collection instruments employed during the pre-intervention stage determined that the learners attempted to comprehend new reading material relying on lexical, word or grammar knowledge, laying emphasis on specific data from the text, with no interference from their background knowledge which would help them identify how new information fits into their existing schema (Brandao & Oakhill, 2005; Hudson, 2007). Moreover, both the quantitative and the qualitative analysis in the pre-intervention stage and in the beginning of the training programme revealed that the learners did not go through any metalinguistic processes, which would assist them in comprehending a text and monitoring their performance (Chamot, 2005; Saricoban & Behjoo, 2017; Sheorey & Mokhtari, 2001). More specifically, they did not plan the cognitive process using textual cues, visual aids and generating questions in the pre-reading stage. In the while-reading stage, the learners did not monitor comprehension by answering the questions they had formed during the pre-reading stage or creating mental images, which would help them draw conclusions and recall details. Finally, during the post-reading stage, the learners did not draw upon background knowledge to make connections, which would enhance comprehension and repair meaning. Furthermore, they could not evaluate their performance by summarizing what they had learnt from the text. In other words, their metacognitive awareness was not developed.

The data obtained through both the qualitative and the quantitative analysis after the intervention revealed that the training programme was effective in improving the experimental group’s reading comprehension and developing their metacognitive awareness. More specifically, the intervention assisted the learners in developing the ability to make predictions (Acosta & Ferri, 2010), to generate questions about a text and find the answers while reading (Bouchard, 2005; Mokhtari & Reichard, 2002), to obtain an overview of the text through skimming (Ueta, 2005), to represent information in a text through graphic organizers (Grant et al., 2015), to consciously create an image of what is read in order to make sense of it (Woolley, 2010), to make connections and relate the new to the known (Waller & Barrentine, 2015), and to identify the most important information in a text and present it in their own words (Westby et al., 2010). As a result, the learners’ approach to reading a text in English changed dramatically. The interview and the observation checklist analysis confirmed the learners’ ability to interact with the text and derive meaning using strategies and linking new to previously acquired knowledge, which resulted in better comprehension (Ozek & Civelek, 2006; Priebe et al., 2012; Rativa et al., 2012). The quantitative data derived from the test results also verified the learners’ reading comprehension improvement following the effective implementation of the strategies taught. Regarding metacognitive awareness, the findings of the qualitative analysis revealed the learners’ development in planning, monitoring and evaluating the reading process through the effective application of the metacognitive strategies taught (Alsheikh & Mokhtari, 2011; Mokhtari & Reichard, 2004; Zhang & Seepho, 2013). These findings were also verified by the quantitative data analysis. Consequently, the training programme allowed the learners to internalize and employ the cognitive and metacognitive strategies taught effectively, thus improving their reading comprehension and building their metacognition.
6. Conclusion

The study presented in this article focused on action research conducted in a Greek State Junior High School aiming at improving the learners’ reading comprehension and developing their metacognitive awareness through cognitive and metacognitive strategy instruction. If we want our learners to become strategic and autonomous readers, we should allow them to exploit texts in a meaningful and purposeful way by laying emphasis on the reading process. To that end, it seems imperative to provide them with strategy instruction (Akkakoson, 2013), which will assist them in interacting with text and deriving meaning as well as in monitoring their performance through thinking about the learning process, planning for learning, monitoring comprehension or production, and self-evaluating what has been learnt. Expanding the sample of the present research to encompass a larger number of participants in the context of both state and private language schools could shed further light on the issue of strategy instruction. Finally, further research could be conducted in order to investigate the maintenance of strategy instruction gains regarding reading comprehension and metacognition after intervention withdrawal.

Note
Pre- and post-intervention interview transcripts along with the observation checklists and the lesson plans can be found in the author’s Med dissertation hosted in the library of the Hellenic Open University.

References


Appendix I

A sample of the tasks included in the pre-intervention test

Name:                       Date:

Mark:    /100               Time available: 50 minutes

**Pre-reading stage**

**Task:** Look at the heading, the subheadings and the photographs.

a. What do you think the text is about? (   /5)

b. What kind of text do you think it is (letter, poem, article etc.)? (   /5)

**While-reading stage**

**Task:** Read the text and complete the diagram below. (   /30)

![Diagram](image)

**Task:** What images come to your mind while reading? Use your senses and complete the sentences below. (   /10)

i) I see...

ii) I hear...

iii) I taste...

iv) I feel...

v) I smell...

**Post-reading stage**

**Task 1:** What does the text you have read remind you of? (   /10)

**Task 2:** What have you learnt from the text? (   /10)
Appendix II

A sample of the tasks included in the lessons

The modelling lesson

Pre-reading stage

**Predicting**

**Predicting Poster**

![Predicting Poster Image](https://www.google.gr/search?q=posters+for+the+prediction+strategy&tbm=isch&tbo=u&source=univ&sa=X&ved=2ahUKEwi886-7ie3dAhUEiywKHaz8AXwQsAR6BAgGEAE&biw=1536&bih=722#imgrc=ZN6ySm_c80VLDM)

**Task:** Do the following task. Then, listen to me reading the text and modelling the strategy.

**a.** Look at the heading. What do you think the text is about? Discuss your guess with your partner sitting next to you.

I predict that

**b.** Look at the heading and the structure of the text. What kind of text do you think it is? Discuss your guess with your partner sitting next to you.

I think that

Post-reading stage
Task: Do the following task. Then, listen to me reading the text and modelling the strategy.

What does the text you have read remind you of? Make connections to yourself, the world and other texts. Discuss with your partner sitting next to you.

<table>
<thead>
<tr>
<th>Making connections</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Text to self</strong></td>
<td><strong>Text to world</strong></td>
<td><strong>Text to text</strong></td>
</tr>
<tr>
<td>It reminds me of ...</td>
<td>It reminds me of ...</td>
<td>It reminds me of ...</td>
</tr>
</tbody>
</table>

https://www.google.gr/search?q=posters+for+the+skimming+strategy&tbm=isch&tbo=u&source=univ&sa=X&ved=2ahUKEwju3f79me3dAhXiiwKHaNXBUUQsAR6BAgGEAE&biw=1536&bih=722#imgrc=7U0lrjKehIJ__M
The independent practice lesson

Pre-reading stage

Prediction

Task

a. Look at the heading and the photographs. What do you think the text is about? Discuss with the partner sitting next to you.

I predict that

b. Look at the heading, the photographs, and the structure of the text. What kind of text do you think it is? Discuss with the partner sitting next to you.

I think that

Post-reading stage

Making connections

Task

What does the text you have read remind you of? Make connections to yourself, the world and other texts. Discuss with the partner sitting next to you.

<table>
<thead>
<tr>
<th>Making Connections</th>
<th>Text to self</th>
<th>Text to world</th>
<th>Text to text</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Maria Avgeri (avgermaria@sch.gr) holds a B.A. in English Language and Literature from the National and Kapodistrian University of Athens and a M.Ed. in Teaching English to Speakers of Other Languages (TESOL) from the Hellenic Open University. She works as a permanent teacher of English in secondary education in the Greek public sector. Developing learners’ skills and autonomy constitutes the core of her teaching practice.
Improving lexical access and acquisition through reading the news: case studies of senior high school students in China

Yuandi ZHANG & James MILTON

This paper reports an empirical study investigating the effects of extensive reading on vocabulary uptake, and speed of word access, among of Chinese high school students using English as a second language. It builds on a study by Masrai & Milton (2018) which reported that an informal program of watching the news on TV and reading subtitles, carried out outside class, could result in significant and measurable improvements in both orthographic vocabulary size and speed of lexical access. This study investigates whether these findings can be repeated with learners from a different L1 background, and reading process background, and where the vocabulary noticing focus of the intervention is removed. The participants were given tests of written vocabulary size, aural vocabulary size, and reading speed. These tests were conducted before and after a six-week intervention which involved reading current news materials. The results showed an increase in reading speed which is assumed to indicate an improvement in the speed of lexical access, but no obvious growth in vocabulary size. It is suggested that this is a learning method which can help facilitate Chinese students' ability to acquire their second language ability in English.

Key words: ELT, Second language acquisition, Chinese education, vocabulary uptake, senior high school, lexical access, reading speed

1. Introduction

Informal learning is a term suggested by Milton (2008) to characterise learning activities conducted outside the classroom and for fun. It is distinguished from incidental learning. Informal learning has a clear learning goal such as learning new words in a foreign language, whereas incidental learning can be inside or outside class, is not necessarily done for recreation, where the activity is meaning focused, and where the acquisition which occurs, something like L2 vocabulary knowledge, is not the goal of the activity. Prior research (Chen 2018, Milton 2012, 2018, Nation 2007, Pigada & Schmitt 2006) has confirmed the positive effect of informal learning in a number of different aspects of language. For example, informal learning can help boost participants' interest and motivation; extensive reading is helpful for different ages of people, including both children and adults (Liu & Zhang 2018). There is a connection between informal learning and vocabulary size gains (Milton 2018). Informal learning is a
broad topic, therefore, that can encompass a variety of approaches and activities and which is thought to produce a similar variety of positive effects. Both individual studies such as watching DVD films, singing songs (Milton 2008) and reading a comic-book (Horst & Meara 1999), and meta-analyses, such as Liu & Zhang (2018), confirm that informal learning can result in vocabulary learning.

While many of these studies investigate vocabulary size gains, the Masrai & Milton (2018) study extends this focus to the fluency dimension in Daller et al’s (2007) three-dimensional model of the mental lexicon, and to the dual-processing model of the lexicon (Coltheart 1978) which considers the lexicon in terms of orthographic and phonological halves. Masrai & Milton’s study, therefore, examined the effects of an intervention involving daily watching of news with English subtitles. It investigated whether this intervention would impact differently in gains in the orthographic and phonological side of the lexicon; confirming the Coltheart and Rastle view of two separable halves of the lexicon. They also investigated whether this intervention could result in changes in the speed of processing of the written word; Daller et al’s fluency dimension. To achieve this, the study made use of some recently developed software to measure language improvement, such as the XK-Lex test (Masrai & Milton 2012) to calculate participants’ orthographic vocabulary size, and A-Lex (Milton & Hopkins 2005a), a test designed to measure phonological size. It also used ReadLex (Milton & Hopkins 2005b), a test designed to measure reading speed in English. The results suggested the reading intervention resulted in a measurable improvement in orthographic vocabulary size, but not phonological vocabulary size. It also resulted in a measurable improvement in reading speed. However, the case study format (one L1 Arabic speaker), requires repetition to confirm whether these results can be generalized.

The study reported in this paper is intended to extend the Masrai & Milton (2018) study. It investigates vocabulary learning which occurs among senior high school students in Mainland China from a program of news reading outside class. Five participants were given a six-week news reading activity, including 36 pieces of news, with 200 to 400 words each, and were given pre-test and post-tests to help the research investigate their language ability change. The investigative methods are taken from the original paper and involve the use of the XK-Lex, A-Lex, and ReadLex vocabulary tests to measure orthographic vocabulary size, phonological size and reading speed. It was intended to investigate whether extensive reading, even without the vocabulary noticing element of reading subtitles which was used in the original study, can be beneficial in promoting vocabulary knowledge and use.

The study is essential for the following three reasons. Firstly, it is widely known that what is taught is not what is learned (Milton 2009). Students only obtain some of the knowledge that teachers provide. The uptake from formal classroom teaching is usually not enough to achieve fluency. Masrai & Milton (2018) report that the uptake of orthographic vocabulary is around 2.5 words per contact hour. This is smaller, typically about half the volume of the words made available for learning. Informal learning, which can allow students to enjoy English after school and in addition to classroom time, may allow learners to acquire words more efficiently. Because learners can choose the topic and content of materials which they like, students are likely to have higher motivation for learning English, they may improve their language ability more quickly, and benefit from a wider variety of learning contexts than is possible in the classroom. Children are likely to be more willing to read where the materials are something new and are fun, than from the unchangeable and limited classroom English textbooks which are provided for non-native speakers. The study can tell us whether an approach of this kind is beneficial to learners.

Secondly, there is little research examining the effect of news reading as an activity in language learning, and little on the effect of such an activity on lexical processing. In the past, researchers have tested movie-watching (Milton 2008) in some situations, informal online learning through mobile phones (Jurkovič 2019), and even extensive reading with graded readers (Carney 2016). As technology
improves, people are more willing to choose more up-to-date and exciting methods of learning compared with traditional news reading, however, the availability of these methods in the Chinese classroom seems limited. Such methods may be time-consuming and are not accessible for Chinese senior high school students where, typically, learners are not allowed to use mobile phones in school. A brief, leisure-time activity, such as extensive news reading which can take only 15 minutes every day using technology which the students already possess, seems a suitable approach for Chinese students where learners already have the motivation to improve their English and perform better in exams. This investigation can add to research and tell us whether such activities are attractive to and practical for learners.

Thirdly, this study can help expand the methodological horizon for English teaching in China. Most formal teaching can still be characterized in terms of traditional English teaching or learning approaches. Most Chinese students, too, still utilize the traditional formal teaching-learning activities they are familiar with. Pressure to succeed may make both teacher and learners conservative in their choice of learning activities and may favour these formal activities, and hard work, making them fearful of trying anything new. However, these approaches, although well-used, may not make not be optimal. Chen (2018) argues that the informal teaching approaches might boost students' learning motivation in English due to the variety of topics for reading learners can choose. Renandya and Jacobs (2016) take this a step further and suggest informal learning seems to improve the traditional teaching methods to some extent, and enlarge students' English vocabularies more efficiently. This study may help extend the range of methodology used in EFL teaching in China and make it more effective.

This paper will firstly focus on a previous relevant literature review describing the importance of learning vocabulary, especially on learners' language abilities like reading speed and reading comprehension. This is followed by an empirical study with a thorough discussion of the results.

2. Literature Review

2.1 Vocabulary uptake

Vocabulary is usually considered to be fundamental in developing language proficiency in L2 speakers (Milton 2009). Research has established that there is a connection between vocabulary and language ability where, generally, the more vocabulary a learner knows, the better they can perform communicatively. Van Zeeland & Schmitt (2013) stated that 2000-3000 word families are needed to reach a basic understanding when facing listening tasks. Coxhead & Boutorwick (2018) stressed the importance of high-frequency vocabulary in dealing with complicated text, especially in maths and science. In mainland China, English is regarded as indispensable for academic and career advancement (Lawson & Hogben 1996). Most Chinese teachers hold the view that vocabulary is a crucial factor in learning L2 English for native Chinese students (Chen 2009). More research is stressing the importance of English vocabulary learning in China.

For example, Wei (2007) did an examination of vocabulary learning in Chinese College-level Learners to figure out the most efficient vocabulary learning strategy. This study used a learning strategy questionnaire to collect data from mainland Chinese college participants. The results showed that contextualized activation and management strategies were seldom used. This is a pity because such activities are considered to aid long-term retention and use of vocabulary. Wei also found that female students in both English majors and non-English majors were more aware of the importance of management in vocabulary learning, which means they use vocabulary learning strategies more frequently than male students.
This is not just a matter of purely research interest. Central government in China also stresses that English is vital in dealing with the growing demand for global communication. China uses an English Language Ability (CSE for short) standard system as a reference tool for teaching, learning, and assessment and this standard is similar to the function of The Common European Framework of Reference (CEFR) and has been linked with the European system. CSE is designed for Chinese students to be classified at three broad language levels: basic, intermediate, and advanced. It gives descriptions of different abilities, including overall ability, reading, writing, speaking, and listening for all students’ grades. This policy stresses that future English teaching in China should focus on language ability and other vital competencies like logic and calculating to cultivate more English-proficient speakers (Chen 2018). Cai & Yun (2020) recently provided some tentative teaching methods within the CSE’s framework to improve students’ intercultural communicative competence. This research illustrates that more language input is necessary for Chinese speakers learning English to reach a high language ability.

China has linked the CSE with the widely-used IELTS test framework, to connect Chinese tests and British tests. The link is presented below (see Table 1). Shangchao (2019) focused on the consistency and feedback of the IELTS reading test results and China’s Standards of English Language Ability (CSE). The results were consistent in most of the levels except CSE 8. This consistency could help Chinese students know their English levels in British tests. For example, the recommended level for senior high Chinese students is level 3, according to CSE. Students are likely to acquire 3500 words in English to enter university. Thus, Chinese students who hold 3500 vocabulary size are more likely to perform less than level 4 in CSE or less than 4.5 in IELTS overall grading.

<table>
<thead>
<tr>
<th>IELTS Score</th>
<th>CSE Level 4</th>
<th>CSE Level 5</th>
<th>CSE Level 6</th>
<th>CSE Level 7</th>
<th>CSE Level 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Score</td>
<td>4.5</td>
<td>5.5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Listening</td>
<td>5</td>
<td>6</td>
<td>6.5</td>
<td>7.5</td>
<td>8.5</td>
</tr>
<tr>
<td>Reading</td>
<td>4.5</td>
<td>5.5</td>
<td>6</td>
<td>7</td>
<td>7.5</td>
</tr>
<tr>
<td>Speaking</td>
<td>5</td>
<td>5.5</td>
<td>6</td>
<td>6.5</td>
<td>7</td>
</tr>
<tr>
<td>Writing</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>7.5</td>
</tr>
</tbody>
</table>

Table 1. Results of Linking IELTS to China’s Standards of English Language Ability

2.2 Vocabulary uptake from formal and informal instruction

Vocabulary is increasingly understood as complex and multidimensional and because of this, it requires equally multidimensional methods to teach it (Kim et al 2018, Pellicer-Sánchez 2016). Formal classroom learning is one method that can enhance knowledge uptake. For example, Yang (2009) illustrated that explicit correction by the teacher and prompting in the classroom leads to a higher uptake rate than recasts. However, formal teaching has some limitations. Cobb (2009) has stated that the vocabulary teaching that textbooks provide is limited because textbooks recycle only a small number of vocabulary items, and did little to expand their vocabulary. Milton (2009) makes the same point. The number of words needed for fluency is so large, and word knowledge so complex, that there is probably insufficient time in the classroom, and insufficient space in the textbook, to present, explain and recycle every vocabulary item.

Attention in vocabulary has moved away from the classroom, therefore, to consider the possibilities that informal learning can offer outside the classroom (for example, Masrai & Milton, 2018; Milton, 2012; Nation, 2007). To the potential benefits in terms of offering wide contexts for learning, and form enhanced motivation, can be added the extra time made available for vocabulary learning. It is sometimes asserted that vocabulary learning in these contexts is incidental and requires little effort or attention on the part of the learner (for example, Harris and Snow 2004). However, research
suggests that while informal vocabulary learning such as reading activities, which has a clear vocabulary learning focus, can be highly successful in enhancing vocabulary uptake (for example, Milton 2008, Chen 2018). Purely incidental vocabulary learning from extensive reading, without this clear learning focus, can be much less successful (Horst et al 1998, Carney 2016) at least in terms of growing a large vocabulary for communication.

The relationship between formal and informal learning, even in an area where this is researched, is still not well understood. Both seem to be needed if something as large and as complex as a foreign language lexicon is to be fully mastered and it can be argued that the two teaching methods are of equal importance (Renandya & Jacobs 2016). There remains an idea, however, even if this is not supported by research, that the lexicon is acquired mostly incidentally and by informal learning outside the scope of the formal classroom curriculum, as in Harris & Snow (2004).

2.3 Informal vocabulary learning and reading speed

The research which is the focus of this paper, and where we attempt to test its conclusions, is Masrai and Milton (2018). This paper conducted a study of informal vocabulary learning in an L1 Arabic speakers learning English who watched movies and the news on TV in Arabic but with subtitles in English. The subject watched 38.87 hours of movies and 8.6 hours of news broadcasts. His language ability was tested through in three different ways: vocabulary size measured by recognition of written words (orthographic vocabulary size), vocabulary size measured by listening to words (phonological vocabulary size), and reading speed. Comparing the data from pre-test and post-tests, their results showed a substantial, 900 word gain in orthographic vocabulary size, a slight improvement in orthographic vocabulary size, and an increase of over 30% in reading speed. It appeared that an informal activity focussing on processing the orthographic form of words, and linking them to meaning, produced benefits in terms of knowledge and access in the orthographic half of the lexicon. The phonological side of the lexicon was, as they expected, relatively unchanged by the activity since there was no obvious L2 listening focus in the activity. Masrai and Milton were particularly interested in this result because of the way they understand native Arabic speakers transfer their L1 Arabic reading skills to reading in L2 English. This kind of intervention, they hoped, would provide useful guidance for such L2 learners of English in overcoming difficulties in reading quickly and fluently in their second language.

Another similar study was conducted by Masrai (2020), which investigated the relationship between phonological vocabulary knowledge and extensive listening. The subjects in the study listened to English L2 news and movies and read the translation in L1 Arabic in the subtitles. They took the same pre- and post-tests as in the original Masrai & Milton (2018) study and the expectation was that phonological vocabulary size would be enhanced with this treatment but the orthographic vocabulary size would not. The results suggested phonological vocabulary was enhanced by the intervention. The rate of learning was higher than that observed in classroom learning, although slower than for the uptake of words in written form in Masrai & Milton’s (2018) study and other studies.

3. Research Questions

The review of the academic literature in the area has suggested the potential benefits of extensive reading in L2 English learning are worth further exploration. For many L2 learners in China, the learning of vocabulary still relies heavily on memorizing words and on translation. These are mechanical and tedious processes. If extensive reading can be shown to be an effective method in L2 vocabulary learning then more people can learn better and in a more relaxed way.
The research in this paper is intended to add to the literature by investigating vocabulary learning from extensive reading, but in a population which does not have L1 Arabic, with its attendant reading processes, by way of background. It can be argued that learners from a Chinese background, where the L1 is logographic, may produce rather different results. The research is intended, too, to investigate whether incidental learning from a reading activity can include the gains in lexical access that Masrai & Milton observe in their study of informal learning through reading. In this research, therefore, we seek to answer the following questions:

**Before the reading intervention**

1. Do L1 Chinese learners of English display the unusual phonologically and orthographically balanced lexicon noted in Milton & Hopkins (2006) and Milton (2018)?
2. How quickly do the participants read successive 10-word lines in a passage of prose?

**After the reading intervention**

3. Does the participant grow the lexicon's orthographic side at the rates noted in other studies?
4. Does the participant grow the phonological vocabulary after extensive reading?
5. Does the participant read the same passage faster?

**4. Material and Methods**

**4.1 Participants**

In this research, the participants were five female volunteers from grade one senior high school, aged about 16. Three of them began studying English in junior high school, which means they had experienced three years of formal English learning. They were believed to have gained a 1,500 word basic vocabulary from junior high school. The other two students started learning English in elementary school, and had more than six years of English study. Their vocabulary was expected to be higher than the others, therefore, and they aspired to attain a level of 3,500 words to attend university.

**4.2 Methodology and tests**

To carry out the programme of research, three vocabulary tests are used. Each of the tests focuses on one specific aspect of vocabulary ability:
- XK-Lex for reading vocabulary size,
- A-Lex for listening language size, and
- ReadLex for learners' reading speed.

XK-Lex (Masrai & Milton 2012) is a Yes/No test which makes an estimate of a learners' knowledge of the most frequent 10,000 lemmatised words in English. It takes a principled sample of 100 words across the frequency range to form an initial estimate of size. It also includes 20 highly infrequent English words, which are never usually known even by native speakers, to form a judgement of over-estimation, and this is used to produce an adjusted score. The version used in this study was a pencil and paper test rather than a computer test, which means testees could go back to revise their answers during the testing.
A-Lex (Milton & Hopkins 2005a) is a computerised Yes/No test which measures phonological knowledge of the most 5000 frequent lemmatized words of English. The test comprises a principled sample of 100 real English words, drawn from across the frequency range, and 20 non-words that sound similar to the real items. Like the function of non-words in XK-Lex, the inclusion of the pseudowords is used to prevent guesswork from participants.

The final test, ReadLex (Milton & Hopkins 2005b) is a computer delivered test of reading speed. The test comprised 40 lines of an English text with 10 words per line, presented one line at a time. The participants were asked to read each line in turn and click a button to load the next line. The programme measures the time taken to load each successive line. The scores are thought to be a measure of lexical access; the automaticity with which words in the lexicon can be recalled for use when needed.

4.3 Material and data collection

The participants were asked to undertake a news reading task six days a week for six weeks. The news presented 200 - 400 words and was taken from a current news online source, assessed at A2 and B1 level, and suitable for senior high school students. The reading task took the participants 15-30 minutes and varied according to their reading ability. Thus, the learner would read about 9,000 words in 36 days and could use a dictionary to look up any unknown words. There was no further intervention. The vocabulary tests were administered both before and after this six week intervention.

5. Results

This study investigates the effect of a reading intervention on three aspects of vocabulary knowledge and ability. Owing to the small sample sizes in this study statistical analysis has not been carried out. Caution must be taken, therefore, in drawing conclusions from the results. The results of pre- and post-tests are shown and discussed below.

Table 2 gives the scores from XK-Lex, the orthographic test of vocabulary size, which has a ceiling of 10,000 words.

<table>
<thead>
<tr>
<th>XK-Lex</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>participant 1</td>
<td>1500</td>
<td>1800</td>
</tr>
<tr>
<td>participant 2</td>
<td>2800</td>
<td>3500</td>
</tr>
<tr>
<td>participant 3</td>
<td>3900</td>
<td>3100</td>
</tr>
<tr>
<td>participant 4</td>
<td>3900</td>
<td>3500</td>
</tr>
<tr>
<td>participant 5</td>
<td>3700</td>
<td>3900</td>
</tr>
<tr>
<td>mean</td>
<td>3160</td>
<td>3160</td>
</tr>
<tr>
<td>sd</td>
<td>1033</td>
<td>811</td>
</tr>
</tbody>
</table>

Table 2. Vocabulary size from XK-Lex

Individual scores on the orthographic vocabulary size range from 1500 to 3900 before the intervention, and from 1800 to 3900 after the intervention. This range reflects the variation in time spent learning prior to testing, and overall language level, as explained above. The participants, based on their vocabulary size appear to be, as expected, at about CEFR A2 level for the lowest scoring learner and about CEFR B2 level for the higher scorers. The higher scoring participants' vocabulary size is roughly that of the requirement of Chinese university entrance of 3500 according to the demands of the Ministry of Education. This number also means that they could have a basic understanding of
comprehension and communication (Milton 2008). The mean score both before and after the intervention is 3160.

Table 3 gives the scores from A-Lex, the phonological test of vocabulary size, which has a ceiling of 5000 words.

<table>
<thead>
<tr>
<th>A-Lex</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>participant 1</td>
<td>650</td>
<td>900</td>
</tr>
<tr>
<td>participant 2</td>
<td>1250</td>
<td>850</td>
</tr>
<tr>
<td>participant 3</td>
<td>650</td>
<td>1300</td>
</tr>
<tr>
<td>participant 4</td>
<td>1450</td>
<td>2100</td>
</tr>
<tr>
<td>participant 5</td>
<td>2350</td>
<td>2750</td>
</tr>
<tr>
<td>mean</td>
<td>1270</td>
<td>1580</td>
</tr>
<tr>
<td>sd</td>
<td>701</td>
<td>824</td>
</tr>
</tbody>
</table>

**Table 3. Vocabulary size scores from A-Lex**

Individual scores on the phonological test vary from 650 to 2350 on the pre-test, and from 850 to 2750 on the post-test. Again, this variation is thought to reflect the range of learning time and level in the learners. The mean score increases from 1270 before the intervention, to 1580 after the intervention and this reflects the fact that four of the five participants increased their individual scores in this area of their knowledge.

Table 4 gives scores from ReadLex, the test of reading speed or, as Masrai & Milton (2018) describe is, speed of lexical access.

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>participant 1</td>
<td>5.8</td>
<td>3.7</td>
</tr>
<tr>
<td>participant 2</td>
<td>6.8</td>
<td>4.3</td>
</tr>
<tr>
<td>participant 3</td>
<td>6.1</td>
<td>7.1</td>
</tr>
<tr>
<td>participant 4</td>
<td>6.0</td>
<td>4.5</td>
</tr>
<tr>
<td>participant 5</td>
<td>6.4</td>
<td>3.1</td>
</tr>
<tr>
<td>mean</td>
<td>6.2</td>
<td>4.5</td>
</tr>
<tr>
<td>sd</td>
<td>0.39</td>
<td>1.53</td>
</tr>
</tbody>
</table>

**Table 4. Reading speed scores from ReadLex in minutes**

Individual scores on the pre-test are quite consistent and vary from 5.8 to 6.8 minutes. There is rather more variation in the results of the post-test where individual scores vary from 3.1 to 7.1 minutes. Four of the five participants improve their reading speed substantially taking, sometimes, only half the time to read the test passage after the intervention. The mean score diminishes substantially too, therefore, and is reduced from 6.2 to 4.5 minutes.

For ease of comparison, the information in these tables is summarised in Figures 5 and 6 allowing pre- and post-test results to be compared.
6. Discussion

There are five specific research questions asked in this paper and the answers provided by the results are dealt with in turn in the sections below.

6.1 Does the learner display a balanced lexicon?

The results suggest that participants have a much larger orthographic vocabulary size than phonological vocabulary size. This conclusion can be reached when pre-intervention and post-intervention scores are compared. Table 7 illustrates this.

<table>
<thead>
<tr>
<th>Participant</th>
<th>A-Lex pre-test</th>
<th>XK-Lex pre-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>participant 1</td>
<td>650</td>
<td>1500</td>
</tr>
<tr>
<td>participant 2</td>
<td>1250</td>
<td>2800</td>
</tr>
<tr>
<td>participant 3</td>
<td>650</td>
<td>3900</td>
</tr>
<tr>
<td>participant 4</td>
<td>1450</td>
<td>3900</td>
</tr>
<tr>
<td>participant 5</td>
<td>2750</td>
<td>3700</td>
</tr>
<tr>
<td>mean</td>
<td>1270</td>
<td>3160</td>
</tr>
<tr>
<td>sd</td>
<td>701</td>
<td>1033</td>
</tr>
</tbody>
</table>

Table 7. Comparison of pre-test XK-Lex and A-Lex scores

The two scores are not strictly comparable, of course, because the ceiling in the XK-Lex test is double that of the A-Lex test. However, these are the tests used in the Masrai and Milton (2018) study which provides the basis for this research. Nonetheless, it is possible to draw persuasive conclusions here. In the original 2018 study the subject’s A-Lex scores were about 90% of the XK-Lex scores which allowed Masrai and Milton to conclude that the subject possessed a balanced lexicon. This was expected given the subject’s L1 background where this type of lexicon is commonly observed (for
example, Milton et al 2010, Milton & Riordan 2006). In this study the subjects’ A-Lex scores are of the order of 40% of the XK-Lex scores which is very similar to the results for non-L1 Arabic learners investigated in the Milton et al 2010 study, and characterized as an unbalanced lexicon. This means the learners know a lot more words in written form than they do by sound.

6.2 How quickly do the participants read successive 10-word lines in a passage of prose?

The reading speed test is thought by its creators (Milton and Hopkins 2005b) and by Masrai and Milton (2018) to give an indication of the speed of lexical access. A feature of knowing a word is the ability to recall it for use automatically when it is needed. This is fluency in the Daller et al (2007) three dimensional model of word knowledge. Segalowitz & Hulstijn (2005) call this automaticity. Alhazmi & Milton (2020) record that educated native speakers on this test take, on average, about 2 – 3 seconds per line and about 2 minutes for the whole test which is roughly in line with other measures of reading speed in English native speakers (Harrison 2001). The subjects in this study are roughly intermediate level learners with comparatively small lexicons and are not expected to match native-like performance. Reading speed will, of course, be affected both by vocabulary size, whether the words a text contains are known at all, and by the learners’ fluency, whether the words that are known can be called to mind for use automatically.

The learners in this study take, on average and before the intervention, over 6 minutes to read the passage. This is slightly longer that the subject in the Masrai & Milton study who was at a comparable level of language ability and with a comparably sized English orthographic L2 lexicon. This outcome likely reflects both the level of vocabulary knowledge these learners have and their comparative lack of fluency in word use. The literature suggests that individuals at the beginning of L2 learning, with small vocabularies, focus their attention in reading on their lower-level processes, which can slow down the reading speed (Segalowitz & Hulstijn, 2005). The literature also suggests that learners have to take the time to develop their reading skills and fluency in the new language (Favreau & Segalowitz, 1983). Another possible explanation comes from Fraser (2007), where his study found a strong association between L1 and L2 speed across different types of reading tasks. If the students read less in their first language, and are slow readers generally therefore, they are more likely to perform comparatively slowly when reading in the L2.

6.3 Does the participant grow the orthographic side of their lexicon at the rates noted in other studies?

Extensive reading is generally argued to aid the growth of a learner’s lexicon (for example, Krashen 1989). To this can be added a caveat, on the basis of research into vocabulary acquisition. Research appears to show that where the focus of a reading activity is on noticing new vocabulary, an informal learning task therefore, then substantial vocabulary gains can be observed, as in the Masrai & Milton (2018) study which informs this study. Where there is no such focus, the activity is for meaning only and any vocabulary learning will be purely incidental, then vocabulary gains may be negligible (Horst et al 1998). The caveat, then, is that extensive reading can be effective in growing a lexicon provided the focus of reading includes noticing new vocabulary.

In this study there is no requirement for learners to notice new vocabulary, they are reading the news for information. It is an incidental learning activity, therefore, at least from the point of view of increasing vocabulary breadth. This may explain the outcome where, on average, there is no observable gain in vocabulary size following the intervention in this study. Three subjects in this study improve their XK-Lex scores but these gains are always smaller than in the Masrai & Milton study, while two learners exhibit similarly marginal decreases in score.
6.4 Does the participant grow the phonological vocabulary after extensive reading?

The Masrai & Milton (2018) and Masrai (2020) studies suggest that if systematic gains in the phonological side of the lexicon are needed then a listening activity that includes noticing the sound form of new words is probably needed. In this study, the activity is an orthographic one for information, with no prompting to notice the phonological form of words. Nonetheless, the results do suggest a gain in the phonological size of the learners’ lexicons. On average, the A-Lex scores increased by about 20% after the intervention. This is about double the size of the gains in phonological size noted in the learner in the Masrai & Milton study. It appears that the intervention may have had the unexpected effect of increasing phonological vocabulary size. It is not clear why this should have occurred. Possibly, the learners in this study sub-vocalised while reading, thus enhancing their knowledge of the phonological form of words in their lexicon known predominantly in orthographic form. Possibly, and this is discussed below, the additional activities that the intervention provoked, including dictionary look-up for pronunciation, may have contributed to this unexpected outcome. The small number of subjects, as with all the conclusions in this study, means that any interpretation can be only tentative.

6.5 Does the participant read the same passage faster?

Table 4 indicates that, on average, the intervention resulted in a decline in the reading speed of the test passage. The mean reading speed fell from 6.2 to 4.5 minutes. After the intervention the learners in this study read the passage, on average about 30% faster, and this is very similar to the learning in the original, Masrai & Milton study. There appears to be no obvious growth in vocabulary size which might help explain such a decline. Our tentative interpretation of this is that, following the reading intervention, the learners in this study, as in the Masrai & Milton study, improved their speed of lexical access.

While there has been no obvious vocabulary noticing or learning focus in the intervention, and the learners are not required to notice new words, they are required to read regularly. The activity may not be focused on expanding vocabulary breadth therefore, but might be argued to focus on the skills required to enhance the vocabulary fluency dimension. This may explain the development of this aspect of the learners’ ability.

6.6 General comments

In studies that address the learning of a foreign language lexicon, there is a considerable focus on how best and most effectively to grow the breadth of the lexicon. This is understandable given the importance of knowing enough words, and knowing the right words, to manage communication. This is, understandably too, a principal target of language learning if it is to be effective. There is an emphasis, additionally, on measuring and finding ways to improve learners’ depth of vocabulary knowledge. This is important because, for fluency, the words a learner knows must be used in the right way and in appropriate structures. The learning of word knowledge, word breadth, in particular appears to be associated with very deliberate processes of word noticing and then word retrieval and use. There is less emphasis on the fluency or automaticity dimension of word knowledge where, perhaps, learning is less deliberately focussed but relies more on the processes of repetition and practice which turns declarative knowledge of and about words, into the procedural knowledge where these words can be used.

Our tentative interpretation of the results of this study gives credence to the potential usefulness of the extensive reading which many teachers and academic believe is so important for learning a foreign language to fluency. The process of regularly reading original and authentic material in a foreign


language, even without a deliberate vocabulary learning focus, can add to the overall goal of vocabulary learning. It appears that learners may not learn new words by the process of extensive reading, but the words they do know are accessed more readily and more quickly, adding to the fluency dimension. The pauses that mar fluent communication, where a word has to be retrieved with time and effort, are diminished. It seems to us that the intervention activity, where the learning of new words would be incidental, can be viewed as an informal learning activity form the point of view of developing the skill of reading. This conclusion is, probably, in line with the observations by Milton et al (2012) and Milton (2012), where game playing in a virtual world foreign language environment, resulted in little obvious vocabulary growth, but measurable increases in the amount of language that could be produced in any given time frame.

Part of the rationale for conducting this study with L1 Chinese participants was to revisit the Masrai & Milton (2018) study with learners from a different L1 background. The motivation for their study was to address perceived issues with the reading process which their L1 Arabic learners, almost uniquely, possessed. Out interpretation of the data in this study is that their conclusions are generalizable to learners of other languages. An extensive reading intervention in this study, as in their study, has resulted in measurable improvements in reading speed which can be interpreted as improved lexical access.

We have described our interpretation of this data as tentative and this is because, as with the Masrai & Milton (2018) study where the authors are equally tentative, the numbers involved in this study are very small. The methodology used to investigate this kind of learning is derived from the case study approach used in Horst & Meara (1999) and Milton (2008) where the lexical micro-environment of the learner can be controlled and understood, to better explain the language learning processes that are occurring. It has proved to be a most insightful methodology but does require repetition if conclusions drawn from the approach are to be confirmed. To this might be added that the intervention in this study was quite short, 6 weeks compared to the 20 week intervention and the absence of measurable vocabulary size gains may, conceivably, be a product of this very short learning time frame.

Finally, to us, the intervention appears to increase is potential for self-learning motivation. The participants have seldom experienced extensive English reading after school. The subjects in this study generally report they would finish the English homework set by their teachers and then would not use English until their next class. This kind of learning method is new for them, and appears to trigger high curiosity and motivation. These news reading activities were welcomed by the students guiding them to the news material, in this case Western news media, which they had not previously accessed. They did not report the material boring or uninteresting, which was feared at the outset of the study. Rather, it prompted them to further, useful, learning activity. Some reported searching the dictionary to confirm the pronunciation of words and to check meaning, to understand the passage entirely. This dictionary search activity may have contributed to the phonological vocabulary improvement observed after the intervention. Some of them even searched the equivalent Chinese news online actively to see possible differences. But, more than this, the activities proved sufficiently motivating in themselves for the participants to see the intervention through to the end with, we think, the lexical access improvement that resulted.

7. Conclusion

This research suggests that extensive reading may be beneficial to the process of vocabulary learning. It appears that it is an activity that could, very usefully, be added to the rather traditional approaches to language learning which are generally to be found in Chinese schools. Even among learners of intermediate level, and whose use of English language is not fluent, the reading of news media proved
to be accessible, interesting and motivating. Previous studies have suggested that a word noticing focus can make this type of activity an effective vehicle for the development of vocabulary breadth. This study adds to this by suggesting that, even without such a focus, the process of practicing the skill of reading in the foreign language improved the speed of accessing words and the fluency of the learners in the use of their L2 lexicons. The intervention also prompted a range of additional, language related activities, such as dictionary look-up and further reading, which can be considered generally beneficial to the learning process.

References


Masrai, A. (2020). Can L2 phonological vocabulary knowledge and listening comprehension be developed


**Yuandi Zhang** was an Ma student in Teaching English as a Second Language at Swansea University between 2019 and 2020 where his dissertation investigated vocabulary learning from extensive reading. He now teaches, and continues to research, in China.

**James Milton** ([j.j.milton@swan.ac.uk](mailto:j.j.milton@swan.ac.uk)) is Emeritus Professor of Applied Linguistics at Swansea University. He has written extensively on vocabulary learning in a second language. His books include: *Modelling and Assessing vocabulary Knowledge* (CUP), *Measuring Second Language Vocabulary Acquisition* (Multilingual Matters), *Dimensions of Vocabulary Knowledge* (MacMillan Palgrave), and *Vocabulary in the Foreign Language Curriculum* (MacMillan Palgrave).
Investigating the Parameters Accounting for Language Proficiency: the case of an English language test used for Academic purposes

Dina TSAGARI & Theodosia DEMETRIOU

In language testing it is essential to understand the validity of test scores (Kane 2013 & 2016). However, the focus on whether test items/tasks lead to the target language use expected (Bachman and Palmer, 1996) has been somewhat limited especially in validation studies undertaken by examination boards. The present study explores the (content) validity, e.g. how different linguistic parameters account for language proficiency in high-stakes international English examination for Academic Purposes (the Pearson Test of English Academic - PTE Academic). Various measures of proficiency were taken into account for Writing and Speaking. Results showed that vocabulary mainly accounts for language proficiency and can be used as a predictor variable for the Writing and the Overall Scores in the test. Fluency could also predict some of the variability in the Speaking Scores. The paper contributes to ongoing research on how various language measures can discriminate between levels of proficiency and proposes a statistical model (regression analysis) that can predict speaking and writing scores. Our research intends to provide feedback to test developers or other stakeholders regarding the PTE Academic and offers research and methodological recommendations for the study of content validity of high-stakes tests in other contexts.

Key words: language proficiency; writing; speaking; vocabulary; fluency; regression analysis

1. Introduction

In the field of language testing, it is good practice for examination boards to conduct validation studies when making claims about the validity or reliability of their language tests (Weir, 2005). Kane (2013, 2016) suggests that any claims by exam organisations should be supported by evidence to the validity and reliability exam scores while Weir (2005, p. 16) proposes that:

“Test validation is the process of generating evidence to support the well-foundedness of inferences concerning trait from test scores, i.e., essentially, testing should be concerned with evidence-based validity.”
Validation research in the field of language testing has tried to identify variables that affect language proficiency (e.g. Elder & O’Loughlin, 2003; Green, 2009; Hughes Wilhelm, 1997). Motivated by work on the features of spoken and written language (Demetriou, 2016; Banerjee et al., 2007; Barkaoui, 2013; Bosker, 2014; Liontou and Tsagari, 2016; Mayor et al., 2007; Read & Nation 2002; Riazi & Knox, 2013; Seedhouse et al., 2014), the current research study set out to investigate and identify specific speaking and writing features that account for language proficiency in the Global Scale of English (henceforth GSE) used by Pearson Test of English Academic (PTE Academic). Our research intends to use the validation framework proposed by Kane (2015) on validity as a property of score interpretation and provide evidence for writing and speaking test score interpretation in the PTE-Academic.

The study aims at contributing to previous work on how measures of lexical diversity can discriminate between levels (e.g. CEFR levels, see Treffers-Daller, Parslow & Williams, 2016) and offers a statistical model (e.g. regression analysis, using the most important variables/features) that can predict PTE Academic Speaking and Writing Scores. The present study is not an investigation of the relationship between the automated scoring systems employed by Pearson and human raters, but rather an exploration of what is actually captured by the scores allocated. In this study, we are trying to explain which of the variables that are used for the automated scoring are the most important in terms of scoring (in other words, what accounts more for getting higher scores).

The study builds on two main themes: i) the importance of vocabulary in distinguishing between proficiency levels, an area highlighted by various researchers (Crossley, Salsbury, McNamara and Jarvis, 2011a; 2011b; Iwashita et al. 2008; Milton, 2010), and ii) the investigation of other variables accounting for proficiency (Barkaoui, 2013; Griffiths, 1992; Pimsleur et al., 1977; Rimmer, 2006; Tauroza & Allison, 1990; Vanderplank, 1993). These will be presented in detail in the following section.

2. Importance of vocabulary in distinguishing between language proficiency levels

The importance of vocabulary in distinguishing between proficiency levels (Crossley, Salsbury, McNamara & Jarvis, 2011a; 2011b; Iwashita et al., 2008; Masrai & Milton, 2018; Milton, 2013) and the significant relationship between vocabulary richness and ratings has been highlighted in the literature (Engber, 1995; Lee et al., 2009; Malvern & Richards, 2002; Milton, 2009; Morris & Cobb, 2004; Yu, 2009). Daller and Phelan (2007) also suggest that the use of infrequent words could be an indicator of language proficiency. In addition, the literature also showed that the more words (tokens) produced by a learner, the higher the level they achieved (Morris & Cobb, 2004). Iwashita et al. (2008) found out that the features of vocabulary and fluency as individual detailed features of spoken language produced by test takers have the strongest correlation with levels of speaking performance (also in Crossley et al., 2011a; McNamara, Crossley, & McCarthy, 2010). Adam’s (1980) study also showed that vocabulary and grammar were the main components that identified different levels of proficiency.

Furthermore, Hawkey and Barker (2004) concluded that in higher IELTS proficiency levels, essays were longer and employed broader vocabulary. O’Loughlin (2013) analysed PTE Academic writing tasks and found a strong correspondence between the holistic scores of essay responses and academic vocabulary use in terms of tokens and types. Their study showed that frequency
and breadth of academic vocabulary (Academic Word list-AWL tokens and types) were important markers of quality in the essay responses.

In the following studies, almost half of the variation is explained by vocabulary knowledge. For example, research undertaken by Demetriou (2016) on the relationship between vocabulary measures and IELTS Writing Task 2 ratings produced a model that explains nearly 50% of the variance of IELTS Writing Scores and confirms that vocabulary is indeed one of the most important factors with a strong relationship with all other language skills. Schmitt (2010) also reports that findings from previous studies showed that vocabulary accounts for 37-62% of the variance in proficiency scores. Similarly, Crossley et al.’s (2011b) findings revealed that lexical diversity could explain over 45% of the variation in human ratings in general and TOEFL scores in particular.

Treffers-Daller et al. (2016) using PTE Academic essays written by 179 English learners showed that lexical diversity measures can discriminate between CEFR levels. Analysing also learners’ scores for each essay, the study showed that the best predictor of CEFR levels that explained 22% of the variance in the scores was the count of words. In addition, basic measures such as the number of different words, Type-Token Ratio (TTR) and Guiraud proved to be better measures than D (Malvern et al. 2004), the Hypergeometric Distribution (HDD - McCarthy & Jarvis, 2007) and Measure of Textual Diversity (MTLD - McCarthy, 2005). Huhta (2014) also found that one of the best predictors for performance on L2 writing tasks were tests of other English skills such as vocabulary.

It is evident from the above empirical research that vocabulary measures are important parameters to consider when investigating the linguistic parameters that account for language proficiency because they can explain a very large percentage of the variance in proficiency ratings/scores.

3. Investigation of other constructs accounting for language proficiency

There is also a large body of literature that investigates other aspects or constructs that account for language proficiency such as fluency, grammar (or grammatical accuracy) and cohesion and coherence. For example, regardless of its problematic definition and its impact on research results (Chambers, 1997; Foster & Skehan, 1996), it is generally accepted that fluency is one of the descriptors of oral proficiency. Chambers (1997) explains the importance of defining fluency in speech as temporal variables (pauses or lengths of runs between pauses) because it “provides a useful anchorage for a concept which is prone to vagueness and multiple interpretations” (1997, p. 538). As temporal variables are “empirically identifiable and quantifiable”, therefore fluency in the current research project will be defined as a temporal variable.

Speech rates or the speed of speech is an additional factor in speaking and understanding English (Griffiths, 1992; Pimsleur et al., 1977; Tauroza & Allison, 1990; Vanderplank, 1993). There are various definitions of speech rate (Chambers, 1997; Levelt, 1989; Riggenbach 1991;). Towell et al. (1996) stress that speech rate alone does not contribute to fluency but other aspects such as frequency of pauses and the length of run (the number of syllables between pauses) are significant factors that need to be taken into account. In addition, Raupach (1987), Towell (1987) and Towell et al. (1996) identified the mean length of runs as the main factor contributing to improvement in fluency.
The importance of fluency in communication has also been highlighted by various other researchers (Foster & Skehan, 1996; Schmitt 1990; Skehan, 1992 & 1996). Most studies on fluency (Ejzenberg, 2000; Freed 1995, 2000; Lennon, 1990; Riggenbach, 1991; Towell et al., 1996) agree that the best predictors of fluency are speech rate (number of syllables articulated per minute) and the mean length of runs (average number of syllables produced in utterances between pauses of 0.25 seconds and above) (Kormos & Denes, 2004, p. 148). In addition, Vanderplank (1993) indicates pacing (the number of stressed words per minute) as another good predictor. Kormos and Denes (2004) also argue that the temporal variable of pace (number of stressed words uttered per minute) is an important predictor of fluency. A more recent study by Révész, Ekiert and Nessa Torgersen (2014) also identified fluency (and lexical diversity) as significant predictors of adequate oral production. The current study also focuses on fluency and investigates whether and how fluency relates to the PTE Academic Scores.

Furthermore, according to Rimmer (2006), grammar also plays an important role in the interpretation of scores. Rimmer’s research showed that “grammatical ability correlates highly with overall proficiency and distinguishes between different levels of test-taker performance” (2006, p. 497). Bygate (1999) also stresses that both grammatical accuracy and fluency are important for language proficiency. According to Foster and Skehan (1996), accuracy is defined differently by different researchers (Crookes, 1989; Ellis, 1987; Robinson, Ting, & Urwin, 1995) The current research project study followed the suggestions by Foster and Skehan (1996:304) who recommended that the calculation of error-free clauses has merit as a measure of accuracy. Also the definition of grammatical accuracy used in this study is the proportion of error-free clauses relative to the total number of clauses (Bygate, 1999; Foster & Skehan, 1996; Skehan & Foster, 1997) where an error-free clause is defined as: “a clause in which there is no error in syntax, morphology, or word order” (Foster & Skehan, 1996, p. 310).

Finally, research has highlighted the importance of text features of cohesion and coherence as aspects of proficiency. For example, Banerjee et al. (2007, p. 12) working on cohesive ties, counted all instances of demonstratives (this, that, these, those). Their results showed that “the use of demonstratives seems to tail off at higher levels of language proficiency, suggesting that other cohesive ties come into use” (ibid., p. 61). In the current study, we check if the same applies to PTE Academic Scores. Barkaoui (2013) also found that coherence and cohesion are some of the features that increase in proportion with scores. The researcher operationalised cohesion and coherence using three measures: ‘Connectives density’, (provides an incidence score for all connectives i.e. causal, additive, temporal and clarification connectives) ‘Coreference cohesion’ (refers to the phenomenon of when a noun, pronoun or noun phrase refers to another constituent in the text, also in Crossley et al., 2011) and ‘Conceptual cohesion’ (refers to how semantically or conceptually similar the content of sentences or paragraphs is). These were all calculated using the Coh-Metrix (a software also used in other studies) to provide measurements of various linguistic indices (e.g. Crossley et al., 2011; McNamara et al., 2010; Riazi & Knox, 2013). In the current study, we also use Coh-Matrix and similar operationalisations of cohesion and coherence.

4. Research Questions and Hypotheses

The present study was motivated and informed by the literature reviewed in determining which variables account for language proficiency in the PTE Academic. The study also aims at creating a predictive model consisting of variables that account the most for gaining high PTE Academic Scores. To accomplish its aims, the current research set out to answer the following questions:
1. *Which variables correlate the highest with Scores obtained on PTE Academic?*

We assumed that all variables would correlate with the PTE Academic Scores but since the literature highlights the importance of vocabulary in proficiency ratings, the researchers expected that vocabulary would have more predictive validity than other measures. Therefore, it was expected that vocabulary values would increase in proportion with scores.

2. *Can a statistical model using the variables under investigation explain the variance in the PTE Academic Scores and to what extent?*

Grammar, oral fluency, and written discourse (Coherence and Cohesion) were expected to be features that increase in proportion with scores. A model consisting of variables deriving from all these constructs should be able to predict (to a large extent) the PTE Academic Writing and Speaking Enabling Scores and Overall Scores.

5. **Methodology**

5.1 **Participants and Materials**

The participants in the study were 100 learners of English from 11 different countries (See Appendix 1 for countries of origin). 72 of the participants were male and 28 were female. All students took the Pearson Test of English Academic (PTE Academic) and were allocated a particular score for writing and speaking skills. Pearson provided the researchers with Writing essay scripts from PTE Academic with their accompanying scores (Scores for Enabling Skills and Overall Score)\(^1\) and Speaking test recordings with their Scores (scores for Enabling Skills\(^2\) and Overall Score) from students\(^3\).

Overall, the data consisted of 100 essay scripts based on different topics and 200 spoken responses from the same students on two different tasks entitled: ‘Describe image’\(^4\). In such tasks candidates usually describe in detail an image (e.g. chart, graph, picture, table or map) related to an academic theme drawn from the fields of humanities, natural sciences or social sciences.

5.2 **Measures**

The measures chosen for the operationalisation of the constructs of the grading criteria were decided following suggestions and findings from the studies reviewed. These are presented in Table 1.

---

1. These were the test takers’ Overall Scores as well as their scores on a number of variables and their Writing and Speaking Score (see PTE Academic Score Guide https://pearsonpte.com/wp-content/uploads/2017/08/Score-Guide.pdf).

2. Enabling Scores were provided for: Fluency, Grammar, Pronunciation, Spelling, Textual skills and Vocabulary.

3. These Speaking and Written Scores are based on all items in PTE Academic which contribute to speaking and written respectively, not only on the essay. Similarly, the Overall Score is based on all items of PTE Academic.

4. For each exam there are 6 or 7 questions of this type - only two of them were used in this study.
<table>
<thead>
<tr>
<th>Variables</th>
<th>Type of Analysis</th>
<th>Method</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written Data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral Data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>Rate of Speech</td>
<td>Total number of syllables /total time expressed in seconds multiplied by 60 to give a figure/number expressed in syllables per minute</td>
<td>Kormos and Dénes (2004) Riggenbach (1991)</td>
</tr>
<tr>
<td>Mean Length of Runs</td>
<td>Total number of syllables produced in utterances between pauses of 0.25 seconds and above</td>
<td>Towell et al. (1996)</td>
<td></td>
</tr>
<tr>
<td>Pace</td>
<td>Total number of stressed words per minute</td>
<td>Vanderplank (1993)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Overview of measures

5.3 Data Treatment and Analysis

The 100 PTE Academic Writing essays and the 200 test recordings (100 audio files) from the Speaking Part of the exam were extracted from the Pearson database. They were all transcribed and formatted into the CHAT (Codes for Human Analysis of Transcripts, MacWhinney, 2000)
transcription format. The latter is one of the most widely used methods of transcribing oral and written data (MacWhinney, 2000). The data was later analysed using CLAN tools. CLAN (Child Language Analysis) is a program that was designed for the creation and analysis of transcripts in the CHILDES (Child Language Exchange System) database (http://childes.talkbank.org/). It comprises various commands for analysing language including \textit{vocd} (McCarthy and Jarvis, 2007) used also in the present study.

The PTE Academic written essays were subjected to analysis of quantitative measurements for the constructs of Vocabulary, Grammar and Written Discourse. The speaking transcripts were subjected to quantitative analysis of Oral Fluency and Vocabulary (see Table 1 for an overview of the measures and methods used).

Vocabulary, Grammar, Written Discourse, and Oral Fluency are 4 of the 6 enabling skills (scoring criteria) used for scoring the PTE Academic. The overall score is based on performance on all test items (tasks in the test consisting of instructions, questions or prompts, answer opportunities and scoring rules). Each test taker does between 70 and 91 items on any given test and there are 20 different item types. For each item, the score given contributes to the overall score. The score range is 10–90 points\footnote{http://pearsonpte.com/wp-content/uploads/2014/07/PTEA_Score_Guide.pdf} (See Appendix 2 for the PTE Academic test format and an extract from the PTE Academic Score Guide explaining how the overall score is calculated\footnote{The Pearson system uses scoring engines such as the Knowledge analysis Technologies (KAT), Intelligent Essay Assessor (IEA), Reading Maturity Metric (RMM) and Versant Technology.}). Our aim was to check for correlations between the measurements and the Scores (Enabling skills Scores and Overall Score).

The calculations of the measurements for each construct (based on the transcriptions of the 100 written essays, recordings, and calculations of the various measures) were analysed in SPSS along with the scores received for each essay (data extracted from the Pearson database). Descriptive statistics tested the correlations between the different measures and the test scores. Correlations between measures and scores were calculated and multiple regression analyses produced various statistical models of predictive validity.

6. Results

6.1 Descriptive Statistics

After all the exclusions, 97 Writing, Speaking and Overall Scores were used in the study (out of the original 100). Participants 10, 27 and 95 had to be excluded from the analysis due to insufficient data. The mean score for Writing, Speaking and the Overall Scores are presented in Table 2 which shows the total number of essays and recordings analysed. The mean for the Overall Score was found to be lower compared to Writing and Speaking Scores due to the fact that the calculations of the Overall Score were based on all item scores obtained by every candidate. However, only 2 score items were used in our study, e.g. essay for Writing and description of image task for Speaking.

\begin{table}[h]
\centering
\begin{tabular}{lcccc}
\hline
& N & Minimum & Maximum & Mean & Std. Deviation \\
\hline
Descriptive Statistics & & & & & \\
\hline
\end{tabular}
\end{table}


6 The Pearson system uses scoring engines such as the Knowledge analysis Technologies (KAT), Intelligent Essay Assessor (IEA), Reading Maturity Metric (RMM) and Versant Technology.
### Writing Score

| Writing Score | 97 | 39 | 88 | 60.10 | 12.856 |
| Speaking Score | 97 | 23 | 90 | 58.02 | 16.827 |
| Overall Score | 97 | 33 | 90 | 57.64 | 12.562 |

**Table 2. Descriptive Statistics**

**Correlations**

The pairwise correlations for the Writing Scores and the writing variables can be seen in Table 3.

<table>
<thead>
<tr>
<th>Writing Score</th>
<th>Pearson Correlation</th>
<th>Sig. (2-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing Score</td>
<td>1</td>
<td></td>
<td>97</td>
</tr>
<tr>
<td>Grammatical Accuracy</td>
<td>.552**</td>
<td>.000</td>
<td>97</td>
</tr>
<tr>
<td>Cohesion – Demonstratives</td>
<td>.122</td>
<td>.235</td>
<td>97</td>
</tr>
<tr>
<td>Cohesion – Connectives</td>
<td>-.243*</td>
<td>.016</td>
<td>97</td>
</tr>
<tr>
<td>Cohesion – Coreference</td>
<td>-.47</td>
<td>.647</td>
<td>97</td>
</tr>
<tr>
<td>Cohesion - Conceptual Sentences</td>
<td>.153</td>
<td>.135</td>
<td>97</td>
</tr>
<tr>
<td>Cohesion - Conceptual Paragraphs</td>
<td>.145</td>
<td>.156</td>
<td>97</td>
</tr>
<tr>
<td>Vocabulary - W - Tokens</td>
<td>.337**</td>
<td>.001</td>
<td>97</td>
</tr>
<tr>
<td>Vocabulary - W - Types</td>
<td>.502**</td>
<td>.000</td>
<td>97</td>
</tr>
<tr>
<td>Vocabulary - W - TTR</td>
<td>.119</td>
<td>.245</td>
<td>97</td>
</tr>
<tr>
<td>Vocabulary - W - Guiraud</td>
<td>.485**</td>
<td>.000</td>
<td>97</td>
</tr>
<tr>
<td>Vocabulary - W - Guiraud Advanced</td>
<td>.426**</td>
<td>.000</td>
<td>97</td>
</tr>
</tbody>
</table>
According to Table 3, the measures with the highest correlations are *Grammatical Accuracy* (0.55), *Types* (0.50), *Guiraud* (0.48) and *Guiraud Advanced* (0.42) with statistically significant results and a negative correlation with the *Cohesion-Connectives* (-0.24). This result was quite unexpected and will be revisited in the discussion section.

### 6.2 Correlations between Speaking Scores and Speaking variables

The pairwise correlations for the Speaking Scores and the speaking variables can be seen in Table 4.

**Table 3. Correlations between the Writing Score and writing variables.**

**Table 4. Correlations between the Speaking Score and speaking variables.**

Table 4 shows statistically significant results of positive correlations between the Speaking Scores and *Rate of Speech* (0.56), *Tokens* (0.50), *Types* (0.49), *Pace* (0.44) and *Guiraud* (0.28) which means that these measures increase when Speaking Scores increase.
To answer our first research question (which variables correlate the highest with scores obtained on PTE Academic) we also checked the average lexical scores obtained by all candidates and compared writing with speaking. Table 5 presents the descriptive analysis regarding the average lexical scores.

As can be seen from Table 5 the mean and standard deviation are higher for the Writing lexical scores than the Speaking Scores. In order to check whether this difference was statistically significant, a paired samples t-test was conducted (Table 6).

Table 6 shows that the difference between the scores of lexical measures for Writing and Speaking are statistically significant.
7. Regression Analyses and Inference

7.1 Predictive Model for Speaking Scores

After all the correlations were checked, data from the whole population was analysed using multiple regression employing all the previous variables (all fluency and all vocabulary and speaking measures) as predictor variables for the PTE Academic Overall Scores. Since all the measures of Vocabulary and Fluency were used as predictors in the regression model for the Speaking Scores, variance inflation factors were calculated to check the presence of multicollinearity (the phenomenon where two or more predictor variables in regression analysis are highly correlated which means that one can be predicted by the other).

After the first regression, there seemed to be a problem with the variable Types. Its Variance Inflation Factor (VIF) was too high (VIF: 420.479). Therefore, another regression was carried out which excluded the variable Types. In the second regression, the TTR VIF was above 10. To prevent collinearity issues, the value needed to be under 10 (Myers, 1990 in Field 2005). Therefore, the variable TTR was excluded (As a rule of thumb, variables higher than 10 were excluded from the model).

A third regression followed without the inclusion of the variable TTR making sure that all VIF values were lower than 10. However, another problem arose. The variable Tokens had to be excluded from the model because it was not statistically significant (p-value was high, e.g. p=.982). Thus, another regression was performed with the exclusion of Tokens. The Mean Length of Runs variable was not statistically significant (p=.907) and needed to be excluded. The results of the next regression analysis showed that Guiraud Advanced had a high p-value (p=.59). Consequently, it was excluded from the model to improve its validity. The results after the exclusion of Guiraud Advanced revealed that Guiraud had a high p-value (p=.217) and had to be excluded as well. After these eliminations, only two variables remained in the analysis: Rate of Speech and Pace. We removed Pace because it did not have a significant value either (p=.170). The results of the final regression model (see Table 7) showed that Rate of Speech was the variable that could explain 31.9% of the variability in the Speaking Scores (R Square=.319).

<table>
<thead>
<tr>
<th>Model Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

* Predictors: (Constant), Fluency – Rate of Speech

Table 7. Final Regression Model (Speaking Score) summary.

Therefore, the fitted regression model for the Speaking Score is as follows:

**PTE Academic Speaking Score: 8.161+ 0.27*Rate of Speech**

7.2 Predictive Model For Writing Scores

The same steps were followed for the creation of the model for predicting the Writing Scores. A regression analysis using backward elimination was carried out using all the writing variables as predictors of Writing Scores. For the same reasons explained in the previous section (high VIF values), the variables Types, Cohesion-Conceptual Sentences and Guiraud were removed to improve the model. After the last regression, all VIF values were under 10 but some other variables had to excluded because they...
were not statistically significant (high p-value), in the following elimination order: Guiraud Advanced (p=.095), Cohesion-Demonstratives (p=.790), Cohesion-Coreference (p=.640) and Cohesion-Connectives (p=.118). All the remaining variables were highly significant in predicting the Writing Scores. Grammatical Accuracy, Cohesion-Conceptual Paragraphs, Vocabulary Tokens, and Vocabulary TTR were the measures that could explain 50.6% of the variability in the Writing Scores (see Table 8, R Square=.506).

<table>
<thead>
<tr>
<th>Model Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

* Predictors: (Constant), Vocabulary – W – TTR, Grammatical Accuracy, Cohesion – Conceptual Paragraphs, Vocabulary – W – Tokens

**Table 8. Final Regression Model (Writing Score) summary**

Therefore, the fitted regression model for the Writing Score is as follows:

\[
PTE \text{ Academic Writing Score} = -43.801 + 24.493 \times \text{Grammatical Accuracy} + 36.986 \times \text{Cohesion-Conceptual Paragraphs} + 0.146 \times \text{Vocabulary Tokens} + 90.171 \times \text{Vocabulary TTR}
\]

**7.3 Predictive Model For Overall Scores**

Finally, the same procedure was repeated for the creation of the final model for predicting the Overall Scores.

After the last regression analysis, all the VIF values for the remaining variables were lower than 10. However, we excluded variables that were not statistically significant starting with the exclusion of Cohesion-Demonstratives (high p-value, e.g. p=.832). After the final regression analysis, all remaining variables were also highly significant in predicting the Overall Scores. Grammatical Accuracy, Vocabulary – W – Tokens, Vocabulary – W – TTR, and Fluency-Rate of Speech were the variables that could explain 54.5% of the variability in the Overall Scores (see Table 9, R Square=.545).

<table>
<thead>
<tr>
<th>Model Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

* Predictors: (Constant), Fluency – Rate of Speech, Vocabulary – W – TTR, Grammatical Accuracy, Vocabulary – W – Tokens

**Table 9. Final Regression Model (Overall Score) summary**

Therefore, the fitted regression model for the Overall Score was as follows:

\[
PTE \text{ Academic Overall Score} = -25.467 + 24.026 \times \text{Grammatical Accuracy} + 0.088 \times \text{Vocabulary-W-Tokens} + 53.725 \times \text{Vocabulary-W-TTR} + 0.110 \times \text{Fluency Rate of Speech}
\]

**8. Summary of results and discussion**

In this section each research question will be addressed separately, presenting the results and discussing any implications and relations with the literature reviewed.

The initial assumption with regard to the first research question (which variables correlate the highest with scores obtained on PTE Academic) was that all variables would correlate with the
PTE Academic scores. However, since the review of the literature highlighted the importance of vocabulary in proficiency ratings, it was expected that vocabulary would have more predictive validity compared to other measures and, therefore, vocabulary values would increase in proportion with scores.

After the analysis of the Writing Score, the results revealed that the measures with the highest correlations were Grammatical Accuracy (0.55), Types (0.50), Guiraud (0.48) and Guiraud Advanced (0.42) with statistically significant results and a negative correlation with the Cohesion-Connectives (-0.24). There were three vocabulary variables (Types, Guiraud and Guiraud Advanced) that correlated with the Writing Scores. This result was not surprising since the expectation was that mostly vocabulary measures would correlate with the Writing Scores. A study by Adam (1980) also showed that grammar and vocabulary were among the main components that distinguished between levels of proficiency. Grammatical accuracy seemed to correlate with Scores as well (also in Bygate 1999 and Rimmer 2006) and one of the Cohesion measures, in particular, Connectives. This result is also in line with findings from Barkaoui (2013) who also found that measures of Cohesion would correlate with the IELTS Writing Scores.

There are various types of connectives (as explained under Cohesion and Coherence) that are included in the count provided by Coh-Metrix under this index. This includes causal (as, for, because, etc.), additive (and, or so, also, furthermore, etc.), temporal (first, finally, meanwhile, etc.) and clarification (therefore, as a result, etc.) connectives. As the literature suggests, the use of these connectives is what makes a text coherent. Therefore, the more of these connectives found in a text, the higher the proficiency level a candidate should be placed at. However, our analysis showed a negative correlation regarding the last variable (Cohesion-Connectives) which was quite surprising and unexpected since the negative correlation with the Writing Score means that the more connectives found in an essay the lower the score that essay is assigned to. It could be the case that if there is incorrect use or maybe overuse of these connectives, the essay scores are negatively affected. In other words, the mere existence of these connectives is not enough to account for the score, but their correct use is. Nevertheless, this finding needs further investigation (e.g. the type of connectives) to understand what might have led to this result.

Regarding the Speaking Score, the results revealed positive correlations with the variables Rate of Speech (0.56), Tokens (0.50), Types (0.49), Pace (0.44) and Guiraud (0.28). Again, as expected here, there are three vocabulary variables (Tokens, Types, and Guiraud) that correlate with the Scores and two fluency variables. This result once again highlights the importance of vocabulary in proficiency ratings and confirms the hypothesis that vocabulary measures increase in proportion with scores.

The results are also supported by previous studies on the relationship between various variables and proficiency scores. Tokens and Guiraud were expected to have a correlation with the Overall Scores (Trefers-Daller, 2016; Demetriou, 2016). Tokens was also found to be a significant factor that correlated with scores in the studies by Hawkey & Barker (2004), O’Loughlin (2013), and Morris & Cobb, (2014). O’Loughlin (2013), in particular, found a correlation between the variable Types and Overall Scores. In addition, it is not surprising that two measures of fluency correlate with the Speaking Scores since the relationship between Fluency and Speaking Scores was also previously highlighted by Iwashita et al. (2008). In particular, it was expected that Speech Rate would be one of the fluency measures that would correlate with the Overall Scores since previous studies (Allison, 1990; Pimsleur et al., 1977; Tauroza & Griffiths, 1992; Vanderplank, 1993) revealed its importance as one of the main factors in speaking English. This was also confirmed in the current study.
As for the second research question (whether there can be a statistical model that can explain the variance in the PTE Academic Scores and to what extent), three predictive models (for Speaking, Writing and Overall Scores respectively) were created using inferential statistics (via SPSS). The three predictive models generated after the analyses conducted are the following:

**PTE Academic Speaking Score** = 8.161 + 0.27 * Rate of Speech
*Rate of Speech* is the variable that can explain 31.9% of the variability in the Speaking Scores.

**PTE Academic Writing Score** = -43.801 + 24.493 * Grammatical Accuracy + 36.986 * Cohesion-Conceptual Paragraphs + 0.146 * Vocabulary Tokens + 90.171 * Vocabulary TTR
Grammatical Accuracy, Cohesion-Conceptual paragraphs, Vocabulary Tokens, and Vocabulary TTR are measures that can explain 50.6% of the variability in the Writing Scores.

**PTE Academic Overall Score** = -25.467 + 24.026 * Grammatical Accuracy + 0.088 * Vocabulary-W-Tokens + 53.725 * Vocabulary-W-TTR + 0.110 * Fluency Rate of Speech
Grammatical Accuracy, Vocabulary –W-Tokens, Vocabulary-W-TTR, and Fluency-Rate of Speech are variables that can explain 54.5% of the variability in the Overall Scores.

The results for the Writing and the Overall Scores were the expected ones. The Writing Score model consists of at least one variable of each of the three constructs under investigation (*Grammatical Accuracy, Vocabulary and Cohesion*) and comprised more vocabulary variables (*Tokens* and *TTR*) than the other constructs. The fact that vocabulary seems to account more for the ratings is also supported by previous studies (Treffers-Daller, Parslow and Williams, 2016; Demetriou, 2016) which also showed the significance of these vocabulary measures and their importance for proficiency scores. Furthermore, according to the literature, *TTR* may be affected by text length. However, we did not analyse samples of equal length but opted for the whole essay for each candidate. This was so because previous research by Demetriou (2016) showed that its text dependence flaws make it a good predictor because the better texts are usually longer.

Nevertheless, what was very surprising was the result of the Speaking Scores. There was only one variable in the predictive model for Speaking Scores that accounted for the variability in the score. This was a measure of Fluency (*Rate of speech*). Even though this finding was in line with previous findings from a study by Révész, Ekiert and Nessa Torgersen (2014) who investigated linguistic features for adequate oral production and identified fluency as one of the significant predictors, it was quite surprising that this was the only predictive variable in the present study. One possible explanation for this could be the nature of the task. The nature of language produced in written speech is different from oral speech (Bygate, 2009; 2010; Luoma, 2004;). This could have affected our measurements and results. Furthermore, the ‘Describe Image’ task is scored on content, fluency, and pronunciation. If the content poorly matches the image, then students will get a low score. Pearson’s automated speech recognition system is quite sophisticated and looks also at prosody but this was not investigated in this project.

In addition, time is another important factor. In other words, how quickly one speaks (or how much more one speaks) during a speaking task could be more important than other constructs. Especially on the PTE Academic task used in this study (e.g. ‘Describe Image’) where the candidates have limited amount of time to speak (25 seconds), it may seem more important to

---

produce as much as one can and as fast as they can. Therefore, the Rate of speech could be one of the most significant factors accounting for the Speaking Score.

Another possible explanation could be that for the Speaking measures most calculations were done manually, therefore subjectivity could be an issue here. Most of the fluency measurements for the oral data were extremely hard to transcribe because of the strong accents of most of the participants which made it hard for the researchers to distinguish, particularly the count of stressed syllables. Also, the difficulty that was encountered for calculating Pace, for example, could have introduced a measurement error (e.g. uncertainty regarding the relationship between the dependent and independent variables) that might have affected the results. There is also a possibility that the way the transcriptions were prepared (by the researchers rather than the examination board) might have had an impact on the output of the analysis.

Lastly, another possible explanation could be that even though all these measures could work on their own, they cannot be used in combination for the creation of the predictive model. To check this, a new regression analysis was run between the Speaking Score and Pace. Pace was the last variable that had to be excluded from the model because of its high p-value. Therefore, we used it on its own to check if it would yield different results. The results and model produced can be seen in Tables 10 and 11.

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
<th>Collinearity Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
<td>Tolerance</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>22.109</td>
<td>7.162</td>
<td>3.087</td>
<td>.003</td>
</tr>
<tr>
<td></td>
<td>Fluency - Pace</td>
<td>.751</td>
<td>.146</td>
<td>.460</td>
<td>5.132</td>
</tr>
</tbody>
</table>

* Dependent Variable: Speaking Score

Table 10. Regression Coefficients with variable ‘Pace’

<table>
<thead>
<tr>
<th>Model Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

* Predictors: (Constant), Fluency – Pace

Table 11. Final Regression Model (Speaking Score) summary

As suspected, even though the p-value is a bit lower than the predictive model of the first Speaking Score, the variable Pace has now a significant value and can explain 21% of the variability of the Speaking Score. This means that the faster the candidates speak, the more proficient they are judged. This confirms the hypothesis that even though all these measures are good predictor variables (for the Speaking Scores) when used on their own, they do not work when combined. Therefore, building a statistical model to predict the Speaking Scores may not be as straightforward as once thought. In order to build a statistical model more predictor variables may need to be taken into account and perhaps a more complicated statistical analysis (e.g. checking the residual plots to adjust the model) should be conducted to achieve this.
9. Conclusion

This study focused on the investigation of different features of written and oral speech and their relationship with PTE Academic Scores. Vocabulary seems to be one of the most important variables which accounts for language proficiency and can be used as a predictor variable for the Writing and the Overall Scores. The fact that Fluency was the only score that predicted some of the variability of the Speaking Score raised some interesting questions that we addressed but this issue needs to be investigated further. For example, different operationalisations of the fluency construct could be explored, for example adding other aspects of fluency such as repair fluency (Skehan 2003) which was also used in recent studies (Révész, Ekiert, and Nessa Torgersen, 2014). In addition, a larger sample (data based on responses from each candidate on all 6-7 speaking tasks) could be used in the analysis to investigate if this could influence the results. One recommendation for further research would be to obtain and analyse the transcriptions for all the speaking and writing tasks from this dataset (100 candidates) and then compare them with transcriptions and results from the examination board. It may also be interesting to look at participants’ L1 and how this might affect the results since the participants’ L1 may intervene in how their English as an L2 develops (Murakami, 2016). Finally, it would also be useful to analyse a balanced sample of test-takers responses across different GSE ranges.

In addition, given that the vocabulary lists used in this study for the calculations of Guiraud Advanced are not updated, we would like to suggest that future research uses more recent vocabulary lists such as the GSE Vocabulary List (Benigno & DeJong, 2017). The GSE List was created by exploring L1 corpora of spoken and written English and aligning vocabulary to the CEFR and the GSE based on combined criteria of frequency and usefulness. One other venue of research would be the investigation of the use of different words from the List based on candidate responses from different proficiency levels or different L1s. It would also be very interesting to look at the average lexical and other scores obtained by each proficiency group and compare speaking with writing. Lexical diversity scores, for example, are expected to be higher for writing than for speaking. Furthermore, since TTR is known to be affected by text length, we would recommend further analysis of samples of equal length to check if the model would remain the same. One last suggestion could be to work on the middle parts of the essays to make sure all essays have an identical number of words (Treffers- Daller et al., 2016).

To conclude, we would like to note that the researchers are aware of the limitations of the study which is not experimental but a construct validity study. Therefore the results of the current study cannot be overgeneralised as the models created here were based on a particular dataset with particular tasks. However, based on cross-sectional data, the study has shown the most significant explanatory composites of L2 speaking and writing skills. Nevertheless, as Ortega and Iberri-Shea (2005) argue, cross-sectional studies cannot be replaced by longitudinal studies for capturing the complex and dynamic processes of L2 development, therefore a longitudinal design can be helpful in understanding such processes (Schoonen et al., 2011). The issues and limitations of this study need to be addressed in further research where different constructs could be added (e.g. spelling or pronunciation) in order to improve the model.

However, the study contributes in identifying the linguistic parameters that account for language proficiency. The findings of the study explain what is actually captured by the allocated scores in the PTE academic and which of the variables that are used for the automated scoring are the most important in terms of scoring, in other words, what accounts more for getting higher scores. According to Pearson, “several proprietary, patented technologies are used to
automatically score test takers’ performance on PTE Academic® and it is particularly important for test developers and other stakeholders to be provided with validity evidence for such widely used tests. The intention of our research was to provide feedback to the test developers whether the algorithms used are reasonable, or whether they need to be revisited and if needed amended. Our findings provide important insights into the construct validity of the PTE Academic examination and, in particular, to the construct of writing and speaking. Finally the paper hopes to have provided a model of validation for anyone studying specific correlates of specific skills in high-stakes exam validation.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Pearson Education Ltd, Research Calls 2016.

References


Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *1*, 2, 57-78.


Appendix 1

Countries of origin and number of participants

<table>
<thead>
<tr>
<th>Countries</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>India</td>
<td>44</td>
</tr>
<tr>
<td>Australia</td>
<td>41</td>
</tr>
<tr>
<td>United States</td>
<td>4</td>
</tr>
<tr>
<td>United Arab Emirates</td>
<td>4</td>
</tr>
<tr>
<td>Singapore</td>
<td>2</td>
</tr>
<tr>
<td>Nigeria</td>
<td>1</td>
</tr>
<tr>
<td>Indonesia</td>
<td>1</td>
</tr>
<tr>
<td>Turkey</td>
<td>1</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>1</td>
</tr>
<tr>
<td>Jordan</td>
<td>1</td>
</tr>
</tbody>
</table>

Appendix 2

PTE Academic content and scoring system - Test Format

Part 1: Speaking & Writing (77 – 93 minutes)
- Personal introduction
- Read aloud
- Repeat sentence
- Describe image
- Re-tell lecture
- Answer short question
- Summarize written text
- Essay (20 mins)

Part 2: Reading (32 – 41 minutes)
- Multiple choice, choose single answer
- Multiple choice, choose multiple answers
- Re-order paragraphs
- Reading: Fill in the blanks
- Reading & writing: Fill in the blanks

Part 3: Listening (45 – 57 minutes)
- Summarize spoken text
- Multiple choice, choose multiple answer
- Fill in the blanks
- Highlight correct summary
- Multiple choice, choose single answer
- Select missing word
Highlight incorrect words
Write from dictation

Scoring system

The overall score is based on performance on all test items (tasks in the test consisting of instructions, questions or prompts, answer opportunities and scoring rules). Each test taker does between 70 and 91 items in any given test and there are 20 different item types. For each item, the score given contributes to the overall score. The score range is 10–90 points.

Overall Score: 56

The Overall Score for the PTE Academic is based on the test taker’s performance on all items in the test. The scores for Communicative Skills and Enabling Skills are based on the test taker’s performance on only those items that pertain to these skills specifically. As many items contribute to more than one Communicative or Enabling Skill, the Overall Score cannot be computed directly from the Communicative Skills scores or from the Enabling Skills scores. The graph below indicates this test taker’s Communicative Skills and Enabling Skills relative to his or her Overall Score.

When comparing the Overall Score and the scores for Communicative Skills and Enabling Skills, please be aware that there is some imprecision in all measurements, depending on a variety of factors. For more information on interpreting PTE Academic scores, please refer to Interpreting the PTE Academic Score Report which is available at www.pearsonpte.com/pteacademic/scores.

Skills Profile

```
<table>
<thead>
<tr>
<th>Skills</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communicative Skills</td>
<td></td>
</tr>
<tr>
<td>Listening</td>
<td>51</td>
</tr>
<tr>
<td>Reading</td>
<td>69</td>
</tr>
<tr>
<td>Speaking</td>
<td>47</td>
</tr>
<tr>
<td>Writing</td>
<td>74</td>
</tr>
<tr>
<td>Enabling Skills</td>
<td></td>
</tr>
<tr>
<td>Grammar</td>
<td>70</td>
</tr>
<tr>
<td>Oral Fluency</td>
<td>37</td>
</tr>
<tr>
<td>Pronunciation</td>
<td>47</td>
</tr>
<tr>
<td>Spelling</td>
<td>61</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>74</td>
</tr>
<tr>
<td>Written Discourse</td>
<td>61</td>
</tr>
</tbody>
</table>
```

Dina Tsagari (dina.tsagari@oslomet.no) is a Professor at the Department of Primary and Secondary Teacher Education, Oslo Metropolitan University, Norway. Her research interests include language testing and assessment, materials design and evaluation, differentiated instruction, multilingualism, distance education and learning difficulties. She is the editor and author of numerous books, journal papers, book chapters, project reports etc. She coordinates research groups, e.g. CBLA SIG – EALTA, EnA OsloMet and is involved in EU-funded and other research projects (e.g. NOHED, KriT, DINGLE, ENRICH, TALE, DysTEFL, PALM, etc)

Theodosia Demetriou (demetriou.th@unic.ac.cy) is a Lecturer at the Department of Education at the University of Nicosia, Cyprus. Her area of expertise is Language Testing and Assessment and her research interests include vocabulary measurement, dimensions of vocabulary knowledge, language testing and assessment and predictive models for IELTS scores. She worked as a Lecturer of English and Linguistics at the University of the West of England, Bristol and as a postdoctoral researcher at the University of Cyprus, (2017). She is currently participating in various research projects in teacher’s assessment literacy in Greece and Cyprus, perceptions of senior academics towards research evaluation and multilingualism and identity styles in Cyprus.
Exploring teachers’ stances regarding the development of digital literacy in the 6th grade textbook of the Greek Primary school

Stergios ROUMELIOTIS & Alexandra ANASTASIADOU

The present paper investigates the promotion of digital literacy in the textbook “English 6th Grade” which is used in primary schools in Greece and explores whether this book should be revised to a certain extent. Eventually, suggestions are provided, based on the accrued data, on possible changes in the specific book. In particular, this study employed a mixed-method approach, as 89 teachers who work in the 6th grade of Greek state schools completed an online questionnaire and, later, 10 of them were interviewed. The participants’ responses yielded that the textbook lacks key features that could contribute to the development of students’ digital literacy skills and the teachers are not offered appropriate guidance on the practical integration of digital literacy in their teaching practice. Moreover, the digital platforms which have been developed to supplement the textbook are not widely used by teachers and no reference to them is made in the Teacher’s book. The findings indicate that the present textbook, as it is, fails to adequately promote digital literacy; thus, specific supplementation needs to be made.

Key words: digital literacy, young learners, textbook, ICT, teacher’s views

1. Introduction

It could be said that there is no perfect textbook that addresses all students’ needs and interests, so teachers tend to design extra material to supplement any inadequacies. Nevertheless, textbooks offer both teachers and learners numerous benefits; having, thus, a central role in education. On the other hand, computers and the Internet are becoming more and more accessible by people of all ages, transforming modern lifestyle which significantly affects educational agendas as they have to integrate the emergent innovations in curricula. Hence, textbook writers incorporate such features in new teaching materials.

In this line, the present paper, which is part of a larger study which was carried out in early 2019, will investigate the teachers’ views about the textbook “English 6th Grade” (Efraimidou, Frouzaki & Zoi-Reppa, 2010) as the particular textbook has been used since 2010 for the teaching of English in state Greek primary schools and no updated versions have been published which incorporate features that
are related to digital literacy. The authors will provide the theoretical background of textbooks and digital literacy and a description of the integration of digital literacy in Greek primary schools, which includes some updates concerning the practices which were employed due to the covid-19 pandemic which spread all over Greece in March 2020. The description of the research methodology is elaborated on, followed by the presentation and discussion of the major findings of the study. Finally, pedagogical implications are offered and possible alterations in the content of the particular textbook are suggested.

2. Theoretical and pedagogical background of incorporating digital literacy in textbooks

Textbooks constitute an essential component of the teaching process, as they provide a framework of what has to be taught during a specific academic year (Richards, 2014), while the accompanying teacher guides advise teachers how to present certain learning objects and familiarize them with new teaching paradigms (Ansary & Babaii, 2002). Textbooks also benefit learners as they give information on the course structure as well as the progress they have made so far (El-Dakhs, 2011), attempting to include at the same time meaningful activities that facilitate the achievement of the goals of the relevant curriculum (Richards, 2001).

Nevertheless, teachers may feel overdependent on the textbooks and not adjust their teaching to their learners’ needs (Porfyri, 2019), restricting this way their own self-development (Richards, 2001). Moreover, textbook content can be outdated or introduce controversial or stereotypical references (Kodriyah, Dayu & Hakim, 2018). For instance, Teliousi, Zafiri and Pliogou (2020) argue that the texts and illustrations in the 4th, 5th and 6th-Grade EFL textbooks of Greek state schools tend to associate males with professions that require manual or mental skills, while women are mostly associated with occupations that concern care and arts.

Meanwhile, developments in the Web have significantly enriched FLT, as teachers can have free access to various information and multimedia, enabling them to supplement textbooks with material that is appropriate to their learners’ needs and wants. Furthermore, students are getting more and more exposed to digital information and multimodal forms of communication, so they need to master certain skills to become competent digital citizens (Liton, 2015). As a result, the focus of modern curricula is no longer restricted to the mere ability to master grammar and the four skills, but they promote multiliteracies, too. One of them is digital literacy which implies the ability to “find, understand, evaluate, create, and communicate digital information” (American Library Association; 2013, p. 1). In other words, individuals need to be active processors of the vast amount of information they have access to and be able to judge its validity so as to link it to their knowledge of the world. Not all digital materials promote digital literacy, though, so teachers need to be familiar with this particular concept to properly incorporate it in their lessons. As some teachers may be inexperienced or lack sufficient training, textbooks can play a significant role in teaching by offering pre-tested activities as well as a rationale on why and how digital literacy should be incorporated in the teaching practice (Hismanoglu, 2011). The demand for digital literacy has greatly increased with the emergence of the new pandemic since 2020 which necessitated engagement in distant learning for both teachers and students worldwide. This required from teachers to revise their teaching practices and search for digital tools and material that could fit in this new reality. Since the covid pandemic induced, social distance educators and students had to become familiar with the Internet in order to participate in online communities and manage the vast amount of information it encompasses.
3. Description of the research context

Having referred to the importance of integrating digital literacy in the mainstream textbooks, the focus will now shift to the Greek primary education.

3.1 ELT curricula in Greek primary schools

In the Greek context, the Ministry of Education is responsible for the design of curricula for state primary schools nationwide in an effort to provide equal opportunities to learning for all students in the country (Anastasiadou, 2015). The present paper focuses on the last two curricula for the teaching of English in primary education. The first one is the Cross-Thematic Curriculum Framework (DEPPS henceforth) (Pedagogical Institute, 2003) which determined the design of the textbook “English 6th Grade”. The second one is the “Integrated Foreign Languages Curriculum” (IFLC henceforth) (Government Gazette of the Greek Republic, 2016) which has been in effect since 2016. The DEPPS (2003) curriculum attempted to introduce ICT as a component of the teaching practice, acknowledging the necessity for students to critically process digital information and get familiarized with the new information society. Nevertheless, these elements were confined to a passive processing of digital texts.

Later on, the IFLC (2016) which was put into action integrated the main features of the preceding curriculum adding a number of innovations, though. According to Anastasiadou (2015), it makes a clearer reference to multimodal information and multiliteracies (especially digital literacy). Furthermore, the IFLC stresses the importance for teachers to incorporate ICTs into their teaching practice using any possible means. This can be achieved through access to the suggested digital platforms (i.e. Photodentro, Aisopos, the British Council) as well as the utilization of digital material that a teacher can select based on their students’ needs.

3.2 The textbook

The textbook under investigation is “English 6th Grade” (Efraimidou, Zoi-Reppa, and Frouzaki, 2010) which adopted the main principles of the DEPPS (2003) curriculum. Although the curriculum for the teaching of foreign languages has changed since 2016 (see section 3.1), the particular textbook is still employed. The content of this coursebook has been designed based on the can-do statements of the CEFR (Council of Europe, 2018) for A2-level learners.

The textbook adopts a cross-curricular approach and consists of ten units whose topics are related to the assumed students’ interests. Furthermore, there are activities such as projects, revision exercises, and self-assessment activities that attempt to promote discovery learning and cognitive and metacognitive awareness. Each unit integrates all four language skills, while grammar and vocabulary also play a major role in the development of structures and carrying out certain functions that are related to the can-do statements of DEPPS (2003). It is important to mention, though, that little reference to ICTs is made throughout the textbook, and no activities are included that require the use of certain skills and strategies for the process of digital information. The current curriculum, i.e., the IFLC (2016), recommends that teachers exploit the platforms that are suggested by the Ministry of Education and attempt to incorporate ICTs in their teaching practice so as to supplement the textbook.

Regarding the textbook “English 6th Grade”, there has not been, to the best of the researchers’ knowledge, any extensive research that investigates its effectiveness, as far as digital tools are concerned, even though there are some studies which examine some aspects of the textbook and
the researchers deem that these findings can accumulatively provide an overall estimation of its potential or lack thereof.

Koutsoupia (2016) investigated teachers’ views on the effectiveness of the reading activities of the particular textbook, a skill that is directly linked to digital literacy. The results pinpointed that its texts contain very demanding vocabulary in relation to the target students’ proficiency level, learners are not asked to be actively involved in information processing, students are not motivated to participate in these activities, the texts do not address different learning styles, and only one-third of the teachers think that the topics appeal to the students’ interests. Another study conducted by Schortsinaniti (2016) concerning the listening aspect of the textbook, reveals that the textbook requires extended supplementation, or a new textbook should be designed in order to include more interesting authentic-in-nature content. Finally, Chatzi (2018) researched the impact of games in the textbook and according to her findings, these games do not seem to appeal to their learners’ preferences, do not appeal to different learning styles, and do not effectively promote the development of language skills.

The aforementioned findings put forward the necessity to investigate the existing literature regarding the teachers’ familiarization with digital literacy and their attitudes towards it. Additionally, reference should be made to the effectiveness of relevant digital platforms to upgrade the present textbook and promote digital literacy. Finally, the already conducted research concerning the quality of infrastructure of Greek state schools will be put under the lens.

3.3 The educational digital platforms

In 2010, the Ministry of Education introduced the program “Digital School Platform, Interactive Books, and Learning Object Repository” (Digital School, 2010) in an attempt to further integrate ICT in state education. One of its aims was the development of different educational digital platforms; one of these is the portal Photodentro (2014) which is a repository hosting Open Educational Resources (Megalou et al., 2016).

In Tsavalia (2017), EFL teachers acknowledged the quality items included in the Photodentro and their effectiveness on the students’ progress. Nevertheless, most of the participants acknowledged that they had little or no knowledge of the Photodentro before their participation in the study and argued that lack of time, proper school infrastructure, or relevant teacher education prevents their use. Similarly, in Batsios (2014) and Kostaki and Kalogiannakis (2020), teachers of Physics in Greek schools stated that they were not aware of the existence of the Photodentro, while they pointed out that the particular material was not well represented making it unattractive for both students and teachers. Moreover, it is mentioned that the Learning Objects of the Photodentro need to be evaluated and classified and that the platform mainly appeals to teachers who have good knowledge of ICT and are willing to devote their free time to integrate it into their teaching practice.

Despite the little popularity of the particular digital platforms in the past, the emergence of the pandemic in 2020 completely changed the situation, since teachers were forced to seek additional online teaching material to the one they used in their classes so as to exploit it in distant teaching. There is no adequate data concerning the popularity of the Photodentro during the pandemic period, though Klouvatos (2021) mentions a huge increase in popularity for the digital platform e-me, so it could be deduced that there was a similar tendency in other platforms as well.

Since only few of these studies concern the subject of English, this study investigated the extent to which the Photodentro effectively supplements the textbook “English 6th Grade” concerning any additional promotion of digital literacy, in an attempt to fill the gap in the relevant literature so far.
3.4 Implementation of ICTs in Greek state schools

In the existing literature, there is a common consensus that digital literacy can advance the learning process due to the possibilities it provides teachers with. In Karkoulia (2016), Tsourapa (2018), and Tzotzou (2018), teachers seem confident about their knowledge of ICTs and most of them use it in their lessons 1-2 times per week. Moreover, the most popular digital tools are videos (especially on Youtube), word processors, and emails. They also argue that ICTs can promote critical thinking and intercultural awareness, aid students with learning disabilities, attract interest, facilitate collaboration, and enable meaningful discussions.

Nonetheless, the implementation of ICTs is challenging in the Greek context due to various factors. Studies conducted by Kokkinaki (2010), Karkoulia (2016), Tsourapa (2018), and Tzotzou (2018) show that the major obstacles teachers face are the poor technological infrastructure of their schools, time pressure, inadequacy of ICT training, and lack of technical support.

Throughout the years, the Ministry of Education has organized a number of ICT training programs in an effort to familiarize teachers with technological tools. One of them was the pilot program “Odyssey (1996-2001) which was heavily criticized as the teachers disapproved of its design and efficiency and questioned its capacity to support them in the use of technology in the teaching practice (Giavrimis, Papastamatis, Valkanos & Anastasiou (2011). Later on, various other programs were carried out, aiming at the development of advanced ICT skills. However, studies conducted by Gousiou (2016), Patseadou (2016), and Tzotzou (2018) demonstrate that access to such programs is very limited for teachers of English as participation is determined through a random selection process, meaning that a large amount of the teachers is excluded. Furthermore, the participants mentioned that their training was partially an outcome of school advisors’ initiatives, which in many instances did not seem to boost their confidence to use ICTs in a proper way, while these programs did not provide them with a sufficient theoretical background to integrate ICT tools in their teaching practice. Addionally, even though, distance learning was put in effect in March 2020, only in April 2021 did the Ministry of Education launch specialized training “Intensive training of teachers’ in distance learning teaching” (https://t4e.sch.gr)1. In this sense, this study makes an effort to fill this gap.

Teachers’ attitudes towards ICTs, the existing challenges in their implementation, and the issue of ICT training will also be investigated in the present study as they are directly related to the promotion of digital literacy in the teaching practice.

4. Research Methodology

Having presented the literature of digital literacy and its application in the 6th grade of the Greek primary school, this part will elaborate on the research methodology.

4.1 Aims of the study and research questions

This study mainly aims to investigate the extent to which teachers of English consider the promotion of digital literacy in the textbook “English 6th Grade” appropriate and adequate. To this end, the following research questions will be addressed:

---

1 Εγκύκλιος ΥΠΑΙΘ 6731/Ε3/20-01-2021. Αθήνα: Εθνικό Τυπογραφείο
1) To what extent are EFL teachers aware of the applications of digital literacy in education?
2) How does this group of teachers supplement the sixth grade EFL textbook to empower their students to develop digital literacy?
3) Which supplementary material could enrich the textbook under investigation in order to boost digital literacy?

As mentioned in the previous sections (3.2, 3.4), there is some research concerning the present textbook and although it does not investigate the aspect of ICT integration or digital literacy, some of its components have been negatively evaluated by teachers. Taking into account that the textbook was released in 2010 and the fact that it was written based on the DEPPS (2003), the researchers’ expectation for this study was that the teachers would be dissatisfied with the promotion of digital literacy in “English 6th Grade”. Furthermore, the studies mentioned so far highlighted that not all the teachers were very familiar with Photodentro before the pandemic. These results would mean that the textbook would require small or extended changes in its content in order to adequately foster digital literacy.

4.2 The participants’ profile

The questionnaire sample consists of 89 teachers of English, 85 women and 4 men, the majority of whom (78.7%) have taught in the 6th grade for at least 6 years.

Almost 60% of the participants feel they are familiar with technology and about 54% acknowledge familiarity with the notion of digital literacy.

![Figure 1. Familiarization with digital literacy](image)

Half of them argue that they use ICT in their teaching practice very often, even though about one-fourth of the participants (25.8%) mention that the technological equipment in their classes is not in good condition.

Almost half of the teachers (56.2%) answered that they have only participated in 1 or 2 ICT-related seminars.
From the above sample, 10 teachers accepted to take part in the interviews. The sampling for both procedures was random with no geographical, educational, experience or age constraints.

4.2 Research instruments

The present study employed a “mixed-method approach”, combining both quantitative and qualitative data in order to enhance the validity and reliability of the accrued findings and provide greater insight into the present situation (Dörnyei, 2007).

For the collection of quantitative data, the authors designed an online questionnaire on Google Forms, as it enabled them to easily share it through email and social media (Cohen, Manion & Morrison, 2017) as well as it facilitated data processing. Moreover, in order to enable participation for people who do not access social media, emails were sent by the researchers to hundreds of schools around Greece, so participation in it was random and voluntary. Then, the qualitative data were garnered through online interviews which were carried out via Skype in order to record them with the participants’ consent. Both instruments had been initially piloted with three teachers, and the feedback was exploited in the correction of items to improve their clarity, validity and reliability.

4.3.1 The questionnaire

The questionnaire was selected as questionnaires are a suitable and extensively utilised quantitative methodological instrument, owning to its versatility and effectiveness to yield quantifiable data in a short period of time (Cresswell, 2011).

The questionnaire was developed after a careful study of the textbook, the IFLC, the digital platforms, and the existing literature. The items were developed so that they could respond to the research questions, without violating the questionnaire principles as mentioned in Cohen et al. (2017).

The questionnaire included 40 compulsory closed-ended questions in total so that it would take little time to answer. The questionnaire consists of 4 parts; Part A concerns demographic information of the participants, Part B includes questions about the respondents’ practices as far as ICT and digital literacy are concerned, Part C is about the teachers’ attitudes towards digital literacy and new technologies, and Part D examines their views on the textbook “English 6th grade”. In parts C and D (Appendix I), the implementation of Likert scales allows the participants to express varying opinions on the included statements flexibly and straightforwardly.
Additionally, three statistical tests were performed to determine the validity and reliability of the findings. Cronbach’s Alpha coefficient test was employed for the analysis of data to measure the internal reliability of certain groups of items in the questionnaire. One-sample Wilcoxon signed-rank test compared the hypothetical and the observed median of specific items to analyze their significance. Finally, Goodman-Kruskal Gamma sought possible correlations between certain items and some demographic features.

4.3.2 The interviews

Since reliance on one methodological tool might limit the findings of a research (Cohen et al., 2017), a mixed method approach can guarantee more reliable data through triangulation (Altrichter et al., 2008). In this line, an interview was employed as the qualitative tool.

To this end, the authors invited the questionnaire respondents, 10 of whom accepted the invitation, through Facebook, Yahoo Groups and emails to participate in the interviews. These were in semi-structured form as there were 10 main open-ended questions for all participants, while some extra ones were included based on the responses that were provided to ensure the flexibility of the procedure and fully investigate the areas the interviewees focused on (Kajornboon, 2005). These questions concerned the participants’ teaching situation, their views on ICT, and the textbook as well as their suggestions for further incorporation of ICT into the teaching materials. To ensure the confidentiality of the procedure each interview was assigned a single number.

Through the data that were gathered from both the questionnaire and the interviews, the researchers could propose practical ways the current textbook could be adapted so that it would incorporate the respondents’ suggestions.

5. Results and discussion

The combination of the quantitative and the qualitative data provided the researchers with invaluable insights concerning the existing situation as far as the promotion of digital literacy is concerned. Based on the findings of this analysis the researchers will attempt to answer the research questions mentioned in the previous section.

5.1 Integration of data from the questionnaire and interviews

1) To what extent are EFL teachers aware of the applications of digital literacy in education?

The findings suggest that despite the little ICT training most of the participants had received, they felt confident to use technological tools in their teaching practice (47.19% Agree, 31.46% Strongly Agree), while they appreciated the overall beneficial effect of digital literacy on their learners (35.96% Agree, 60.67% Strongly Agree), its significance in the school curriculum (37.08% Agree, 50.18 Strongly Agree), and its contribution to the development of language skills (42.70% Agree, 55.06% Strongly Agree).

These findings confirm the ones mentioned in Tsourapa (2018) and Tzotzou (2018) and can be explained by the increased role of computers and the Internet in people’s everyday life which has helped teachers familiarize themselves with various digital tools. Nevertheless, it is important to mention that some of the participants in the interviews did not seem able to clearly distinguish digital literacy from the mere use of technology, which suggests that knowledge of technology is not
enough in order for teachers to effectively integrate it in the teaching practice as they need to be aware of the key principles of digital literacy and be sufficiently trained for its proper implementation. Another finding is that most of the teachers do not feel that the syllabus imposes significant time constraints that prevent them from using ICT, which contradicts the findings in Kokkinaki (2010), Karkoulia (2016), Tsourapa (2018), and Tzotzou (2018). This could be attributed to the fact that, in recent years, teachers do not use the present textbook as the primary material and depend on supplementary material.

2) How does this group of teachers supplement the sixth grade EFL textbook to empower their students to develop digital literacy?

In the questionnaire, the majority of the participants acknowledged the usefulness of the digital platforms suggested in the curriculum (51.69% Agree, 11.24% Strongly Agree), something that can be associated with the sense of safety that ready-made materials create (37.08% Agree, 14.61% Strongly Agree) since it is hosted on an authorized digital platform.

![Figure 3. Usefulness of the suggested digital platforms](image1)

However, more than half of the participants in the interviews pointed out that they were not familiar with the Photodentro and they preferred to use their own material, which confirms the findings in Tsavalia (2017), Batsios (2014), and Kostaki and Kalogiannakis (2020). Some of them explained that using their own material is a quicker and more reliable option, since even the content in the Photodentro required some adaptation in order to use it. The large majority of the respondents confirmed that tendency to develop their own material considering their learners' needs (37.08% Agree, 14.61% Strongly Agree).

![Figure 4. Teacher-made digital materials](image2)
These findings could suggest that teachers should be further familiarized with the aforementioned digital platforms through ICT training and be encouraged to develop material to be hosted on them. Even though these platforms became more popular during the pandemic, many teachers depended on their own material or ready-made material posted on online teacher communities.

3) Which supplementary material could enrich the textbook under investigation in order to boost digital literacy?

Half of the respondents reported that the utilization of digital tools is not fostered through the textbook (38.20% Agree, 11.24% Strongly Agree), while the same number does not believe that the achievement of the aims and objectives of the IFLC in relation to digital literacy is facilitated by the textbook (11.24% Strongly Disagree, 40.45% Disagree) or that digital literacy is sufficiently integrated into it (7.87% Strongly Disagree, 43.82% Disagree).

What is more, the majority (64.02%) of the teachers argued that the textbook lacks key features that concern digital literacy such as multimodal information to process digital texts.

These responses confirm the insufficient promotion of digital literacy in the textbook and this is why the IFLC (2016) suggests the use of supplementary digital material. Besides, most of the interviewees also admitted that digital literacy is not of high priority for the particular textbook. These demonstrate that the textbook does not aim at enhancing the students’ computer skills, which are a major component of the current curriculum, and confirm the participants’ belief that the textbook needs to be revised. It is also important to mention that a large number of the respondents mentioned that the textbook writers do not provide them with adequate tools to integrate digital literacy in the teaching practice (8.99% Strongly Disagree, 30.34% Disagree).
This is also confirmed in the interviews where the respondents argued that the textbook makes no reference to the suggested digital platforms and does not advise them on how to integrate digital tools in their lessons. This can be confusing for both teachers and learners as browsing through the Photodentro or any other platform can be time-consuming if there are no direct prompts to certain learning objects.

In the interviews, several suggestions were made concerning possible changes in the textbook. For instance, it was suggested that the textbook could become more appealing and motivating for learners by including updated material and web links. Since information easily becomes outdated, there could be some fixed web links whose content could be regularly updated. Another point they made is the inclusion of direct references to the suggested digital platforms. This would help teachers easily and quickly browse through various learning objects and choose the ones they believe will meet their learners’ needs and interests. Moreover, several teachers agreed that the project tasks in the textbook should encourage the use of modern digital tools (such as Google Docs instead of the traditional pen-and-paper format) in order to enhance the students’ computer skills. Finally, strategies need to be included to guide learners towards the processing of digital information. These strategies could concern the use of the highlighter, the way they can use hyperlinks, or how to evaluate the validity of digital information.

5.2 The statistical tests

In Cronbach’s Alpha, the Alpha coefficient for each group of items was 0.821, 0.773, and 0.735 respectively, demonstrating a satisfactory coefficient of internal consistency for all three groups. In One-sample Wilcoxon Signed Rank, all key items deviate from the hypothesized median value, considering that their p-value is below .05, which demonstrates that the responses are of high statistical significance. Finally, the Goodman-Kruskal Gamma shows that none of the items was influenced by certain factors such as teaching experience or attending ICT training programs, which means that the findings of the present study concern all the participants (Appendix I).

5.3 Recommendations for the digital supplementation of the book

In Section 2, it was discussed that digital literacy concerns the identification, evaluation, creation and communication of digital information, and only few of these features can be found throughout the
textbook, while learners are not guided in the development of such skills. In order to address this issue, the textbook needs to incorporate reading activities along with specific strategies - such as highlighting, exploitation of hypertexts, and evaluation of digital information. These reading activities could be uploaded on the Photodentro or another digital platform and be renewed on a regular basis, since digital information can become outdated easily nowadays. Teachers could play an important role by sharing such activities with other teachers so that different interests and needs are covered. The Covid-19 experience showed that maybe the Photodentro and the Aesop platforms could be transformed into digital communities in which teachers will interact, share, and review material, since such communities solved several problems teachers faced in the initial phase of distant teaching.

Another suggestion that can be made based on the participants’ responses is for the textbook to include prompts to the material that is hosted on the aforementioned digital platforms. Even though the employment of distant teaching has increased the popularity of these platforms, browsing through them to find the appropriate learning objects can still be quite time-consuming and difficult for both teachers and learners. These prompts will offer guidance to the material and will stress its significance for the achievement of the aims and objectives of the curriculum regarding digital literacy.

Moreover, the Teachers’ Book should assist teachers in exploiting ICT, as access to ICT training is very limited as mentioned in previous sections. There could be some definitions for teachers to understand the nature of digital literacy, advice on the strategies they should focus on and inclusion of possible activities that promote the development of computer skills.

Finally, the textbook should encourage the exploitation of various digital tools such as word processors, blogs, and wikis as they supply numerous benefits in relation to the traditional pen-and-paper writing. It is important to mention that distant teaching gave prominence to even more digital tools such as Wordwall, Wakelet, Jamboard, and Microsoft Sway which can make lessons more interactive and integrate multimodal features to address different learner needs. These tools could be combined with the Photodentro, the e-me, and the Aesop platform in order to maximize the benefits for both teachers and students.

In Appendix II, there is an example of such changes on page 22 of the “English 6th Grade” Student’s Book so that learners interact with the digital information, evaluate it, and capitalize on a real and more meaningful digital environment.

Even though this research attempted to investigate the issue of the promotion of digital literacy in the textbook “English 6th Grade”, further research would provide a better insight into this issue. Furthermore, the findings of this research can potentially be exploited along with the findings of other studies regarding the textbook “English 6th Grade” in order to develop proposals for a revised textbook.

6. Conclusion

The findings of the present study suggest that the textbook “English 6th Grade” fails to promote digital literacy to a satisfying degree with respect to the aims and objectives of the IFLC. Additionally, the existing digital platforms partially address this issue due to the fact that no clear guidance is provided to either teachers or students on how to exploit them in order to develop the required skills. Therefore, the extensive revision of the present textbook or even writing a new one should be
considered by the educational authorities. Hence, apart from the aforementioned recommendation, the following points should also be considered.

In order to address the limited access to ICT training programs in a cost-effective way, the Teacher’s Book needs to provide teachers with guidance concerning the integration of digital literacy in their teaching practice. Furthermore, teachers and students should be familiarized with the Photodentro and other suggested digital platforms, by being frequently prompted to its digital material through the textbook and be supported on how to exploit it. Finally, teachers should be trained to efficiently use various digital tools (Wordwall, Padlet to mention but a few) in order to create their own material that will in turn enable them to cater to their students’ needs and preferences.

References


Kokkinaki, A. (2010). Obstacles to the integration of ICTs in Primary Education in Greece. *Interactive Conference on Computer Aided Learning*.


### Appendix I: Questionnaire Sections C & D

<table>
<thead>
<tr>
<th>Items</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Wilcoxon</th>
</tr>
</thead>
<tbody>
<tr>
<td>I believe that digital literacy is necessary for students.</td>
<td>0</td>
<td>0</td>
<td>3.37</td>
<td>35.96</td>
<td>60.67</td>
<td>.000</td>
</tr>
<tr>
<td>I consider digital literacy an important part of the school curriculum.</td>
<td>0</td>
<td>1.12</td>
<td>5.62</td>
<td>37.08</td>
<td>50.18</td>
<td>.000</td>
</tr>
<tr>
<td>My students are motivated when the lesson includes the use of technology.</td>
<td>0</td>
<td>0</td>
<td>5.62</td>
<td>32.58</td>
<td>61.80</td>
<td>.000</td>
</tr>
<tr>
<td>I feel that the textbook publishers provide me with adequate tools to integrate digital literacy into my lessons.</td>
<td>8.99</td>
<td>30.34</td>
<td>40.45</td>
<td>19.10</td>
<td>1.12</td>
<td>.007</td>
</tr>
<tr>
<td>Time pressure doesn’t allow me to prepare lessons which include applications that promote digital literacy.</td>
<td>4.49</td>
<td>17.98</td>
<td>33.71</td>
<td>31.46</td>
<td>12.36</td>
<td>.011</td>
</tr>
<tr>
<td>I find it difficult to integrate digital literacy into my teaching practice.</td>
<td>15.73</td>
<td>37.08</td>
<td>26.97</td>
<td>13.48</td>
<td>6.74</td>
<td>.002</td>
</tr>
<tr>
<td>I struggle to use technology because I have little time to cover the given syllabus.</td>
<td>14.61</td>
<td>32.58</td>
<td>24.72</td>
<td>20.22</td>
<td>7.87</td>
<td>.046</td>
</tr>
<tr>
<td>I feel confident to use technological tools in my teaching practice.</td>
<td>1.12</td>
<td>5.62</td>
<td>14.61</td>
<td>47.19</td>
<td>31.46</td>
<td></td>
</tr>
<tr>
<td>I believe that digital literacy can contribute to the development of language skills.</td>
<td>0</td>
<td>0</td>
<td>2.25</td>
<td>42.70</td>
<td>55.06</td>
<td>.000</td>
</tr>
<tr>
<td>I mainly use technology when I want my students to relax.</td>
<td>13.48</td>
<td>37.08</td>
<td>37.08</td>
<td>8.99</td>
<td>3.37</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table 1. Percentages of items from Section C and analysis of their significance with One-sample Wilcoxon test
Table 2. Percentages of items from Section D and analysis of their significance with One-sample Wilcoxon test

<table>
<thead>
<tr>
<th>Items</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Wilcoxon</th>
</tr>
</thead>
<tbody>
<tr>
<td>My teaching mainly follows the structure of the textbook.</td>
<td></td>
<td>4.40</td>
<td>17.98</td>
<td>30.34</td>
<td>42.70</td>
<td>4.49</td>
</tr>
<tr>
<td>The current textbook is appropriate for the needs of the learners of the 6th grade.</td>
<td>16.85</td>
<td>47.19</td>
<td>26.97</td>
<td>8.09</td>
<td>0</td>
<td>.000</td>
</tr>
<tr>
<td>The textbook doesn’t promote the use of technological tools.</td>
<td></td>
<td>2.25</td>
<td>16.85</td>
<td>31.46</td>
<td>38.20</td>
<td>11.24</td>
</tr>
<tr>
<td>The textbook helps me achieve the aims and objectives of the current curriculum concerning digital literacy.</td>
<td></td>
<td>11.24</td>
<td>40.45</td>
<td>40.45</td>
<td>7.87</td>
<td>0</td>
</tr>
<tr>
<td>I find the digital platforms provided on the websites suggested in the curriculum (e.g. Aesop, Photocentro) useful in my teaching practice.</td>
<td></td>
<td>3.37</td>
<td>6.74</td>
<td>26.97</td>
<td>51.69</td>
<td>11.24</td>
</tr>
<tr>
<td>I usually develop my own materials, which promote digital literacy, by taking into account my learners’ needs.</td>
<td></td>
<td>1.12</td>
<td>10.11</td>
<td>13.48</td>
<td>40.44</td>
<td>25.84</td>
</tr>
<tr>
<td>I feel that the textbook adequately incorporates digital literacy.</td>
<td></td>
<td>7.87</td>
<td>43.82</td>
<td>37.08</td>
<td>10.11</td>
<td>1.12</td>
</tr>
<tr>
<td>I feel safer when I am provided with proper ready-made digital materials.</td>
<td></td>
<td>4.49</td>
<td>11.24</td>
<td>32.58</td>
<td>37.08</td>
<td>14.61</td>
</tr>
<tr>
<td>The textbook incorporates activities in which the students seek information online.</td>
<td></td>
<td>3.37</td>
<td>24.72</td>
<td>42.70</td>
<td>26.97</td>
<td>2.25</td>
</tr>
<tr>
<td>The textbook doesn’t need to be revised as long as I am able to supplement it accordingly.</td>
<td></td>
<td>29.21</td>
<td>41.57</td>
<td>19.10</td>
<td>8.99</td>
<td>1.12</td>
</tr>
<tr>
<td>The textbook exposes the students to uses of digital tools that are relevant to the needs of students in Greece.</td>
<td></td>
<td>11.24</td>
<td>41.57</td>
<td>40.45</td>
<td>4.49</td>
<td>2.25</td>
</tr>
<tr>
<td>The textbook takes into account various sources of information from the Internet.</td>
<td></td>
<td>5.62</td>
<td>30.34</td>
<td>41.57</td>
<td>19.10</td>
<td>3.37</td>
</tr>
<tr>
<td>The content of my teaching practice follows the content of the textbook.</td>
<td></td>
<td>5.62</td>
<td>6.74</td>
<td>37.08</td>
<td>47.19</td>
<td>3.37</td>
</tr>
<tr>
<td>The Teacher’s Book provides me with guidance on how to exploit digital tools.</td>
<td></td>
<td>25.84</td>
<td>39.33</td>
<td>28.09</td>
<td>6.74</td>
<td>0</td>
</tr>
<tr>
<td>The textbook integrates real instances of use of digital platforms.</td>
<td></td>
<td>13.48</td>
<td>40.45</td>
<td>33.71</td>
<td>11.24</td>
<td>1.12</td>
</tr>
<tr>
<td>The textbook doesn’t suggest strategies to process digital texts.</td>
<td></td>
<td>2.25</td>
<td>11.24</td>
<td>21.35</td>
<td>50.56</td>
<td>14.61</td>
</tr>
<tr>
<td>The textbook exposes students to hypertexts (i.e. links between different digital texts).</td>
<td></td>
<td>7.87</td>
<td>34.83</td>
<td>33.71</td>
<td>22.47</td>
<td>1.12</td>
</tr>
<tr>
<td>The textbook exposes students to multimodal information (i.e. language input that consists of combinations of sound, text, video, graphics, etc.)</td>
<td></td>
<td>12.36</td>
<td>39.33</td>
<td>24.72</td>
<td>23.60</td>
<td>0</td>
</tr>
</tbody>
</table>
Appendix II: Original version

UNIT 2

Lesson 3  E-shopping

1. The Internet site
   A. Have you ever used the Internet to do your shopping online? Why?

   B. John and Mary are looking for some toys on the Internet. Here is a site where they can look at, order and buy toys online at:
   www.countdowncreations.com/toys.html
   Work in pairs and say what kind of toys they are looking for.

   C. If you click on the picture you can find more information about the toy you are interested in. Visit the webpage to find the following information:
   ⇒ How much is the astronaut?
   ⇒ How much is the space shuttle?
   ⇒ Which of the two toys comes in pieces you put together?
   ⇒ How many pieces is it?

   D. PROJECT
   You want to buy presents for the pupils of a primary school in Cyprus. Work in pairs and:
   ⇒ Visit the webpage to buy your presents online
   ⇒ Choose the toys you like
   ⇒ Fill in the order form on the right but don’t order!
   (if you don’t have Internet access, turn to pp. 156-157)
UNIT 2

Lesson 3  E-shopping

1. The Internet site

A. Have you ever used the Internet to do your shopping online? Why?

B. John and Mary are looking for presents on the Internet. Here is a site where they can look at, order and buy electronics online at:
http://example.to/3Wc4ARn
Work in groups, browse the website and discuss what kind of electronics you can buy in this link.

C. If you click on a product, you can find more information about it. What features do you look at to judge the product? Share your ideas with your group and then with the entire class.

D. Visit the link again. On the left side there are some options. How helpful are they? Why?

E. PROJECT
Take a look at the different products and choose one that you like. Take a screenshot, post it on the e-me wall of your class and write a brief post justifying your choice. Then, your class can create a poll on e-me with all the products to vote the best one. Finally, work together, write a review on the final choice and post it on the e-me blog of your class.

F. EXTRA
Visit the e-me blogs of other schools and write comments on the products they have selected. Write why you like or don’t like them.
Stergios Roumeliotis (stergios-r@hotmail.com) is a graduate of the Faculty of English Language and Literature of the National and Kapodistrian University of Athens and holds a M.Ed. in TESOL from the Hellenic Open University of Patras. He works as a teacher of English in primary education. His research interests concern educational technology and teaching material development.

Alexandra Anastasiadou (alexanastasiadou@yahoo.gr) holds a B.A. in English Language and Literature and a B.A. in the Pedagogical Department for Primary Education from Aristotle University, an M.Ed. in TESOL from the Hellenic Open University of Patras, an M.A in Cognitive Development from the Pedagogical Department of Western Macedonia and a Ph.D. in Teaching Writing from Aristotle University. She works as a state school advisor-teacher trainer in the Regional Directorate of Education for Central Macedonia and as a dissertation supervisor in the postgraduate programme of the Hellenic Open University in Patras. Her research interests include Teaching writing, Teaching young learners, Curriculum design, Teacher Training, Critical literacy, CLIL.
VR and Social Identity
The utilisation of Virtual Reality Technology for the study of Social Identity in children with a migrant background situated in Greece

Kyriaki THOIDOU & Vasilios ZORBAS

This case study examines the use of virtual reality for comprehension of children’s social identity. More specifically, it investigates whether virtual reality can be assistive in social identity discovery. In this context, virtual reality refers to the utilization of 360° panoramic photography as well as commonly used (smartphones) and inexpensive devices (VR Headset for mobile phones). Three different qualitative tools were employed: semi-structured interviews, social location maps and identity texts. Data were collected from the two siblings with a migrant background as well as their immigrant parents. The findings showed that VR was assistive in the vivid, in-depth detection of social identity. Moreover, their creations highlighted the progression of the participants' social identity through time and space. Their social identity changed just by virtually visiting different places. Symbolic representations regarding essential aspects of social identity as well as the need for inclusion emerged from this research. The children demonstrated their ideal inclusive Greek environments, which were different from their reality. Furthermore, theoretical reservations against the use of VR were also considered.

Key words: Social Identity, Virtual Reality, 360° panoramic photography, migrant background, social location maps, identity texts

1. Introduction

“People often say that this or that person has not yet found himself. But the self is not something one finds; it is something one creates.”
(Thomas Szasz, 1973, pp. 49, in The Second Sin)

Every individual possesses their own sense of self. Humans have spent centuries developing theology, philosophy, psychology, science, policy and art in order to understand their identity. Exploring one’s identity in all of its forms can provide valuable and surprising knowledge contributing towards deeper human comprehension. The study of migrants' identity can offer valuable insights into their mindset, background and ideology. The deeper divide between people with migrant and non-migrant
backgrounds derives from social discrimination. In the literature, the study of discrimination is closely linked to social identity (Hoff & Pandey, 2005; Mummendey & Schreiber, 1983; Oakes & Turner, 1980). Social identity is the type of identity that is developed through social interaction. It involves the groups to which an individual belongs.

So far, to the best of my knowledge, the detection of social identity has never been linked to virtual reality. Although identity and virtual reality have been used in research to elicit empathy for migrants, there is a significant difference between the two concepts. The current case study attempts to provide a small innovative spin to the research of social identity. The present modification aims to build upon previous research (Papalexatou & Zorbas, 2020; Zorbas & Karras, 2020) and detect the social identity of the participants in a different way. Three different qualitative tools are combined with a VR activity in order to probe into the social identity of two siblings of an immigrant family. The family is of Romanian origin and currently resides in Greece.

2. Theoretical Landscape

2.1. Virtual Reality in Research

A more profound comprehension of this research can be achieved by examining virtual reality and its applications in the literature. According to the Cipresso et al. (2018) the concept of virtual reality was initially utilized in the late 80s, although the idea of virtual reality was formed during the 60s. According to Lanier (1992), virtual reality is defined as the gear, clothing supplies, or machinery necessary to support the belief of living or existing in a cyber environment. In recent years, virtual reality has progressed to involve higher levels of immersion with the invention of the VR Headset. Moreover, Zhang et al. (2018) wrote that VR enables users to fully immerse themselves in the virtual environment without using plugin peripheral input devices (e.g., mouse, keyboard). The virtual environment can be viewed instinctively, just by shifting one’s head.

Virtual reality has also been used in research. Spiridon and Sandu (2016) utilized it in order to engage museum visitors in cultural communities. Kors et al. (2016) had students play a VR game visualizing the experiences of a refugee. Hence, VR technology has been used in research in order to raise cultural sensitivity and spatial immersion. This study will go a step further and try to learn about students’ culturally sensitive issues rather than teaching participants about cultural sensitivity. Kors’ et al. (2016) study can be an inspiring example that demonstrates how visual immersion can trigger reflection and engagement. There are also similar studies educating participants about migrant experiences (such as Grover & Pennay, 2020; Camilleri et al, 2016).

As in most new technological advancements, VR technology has received some criticism. Chawla et al. (2013) mention that some individuals are anxious that VR will reduce human interaction. However, in this study, the opposite effect is expected. Kenwright (2018) has also explored some ethical challenges regarding the use of VR. He stresses the need to further investigate the dangers of desensitization, post-traumatic stress and children’s recognition of virtual characters as real. Although these concerns constitute valid grounds for investigation, they seem to refer to prolonged exposure and virtual reality games. Virtual reality games have a plot where users interact and change the virtual environment (e.g. Jang & Park, 2019; Skarlatos et al., 2016). However, in this study, VR is used for a short period of time and is only interactive in viewing the 360° photography of real-life photos. It does not provide an augmented reality. Hence, the issues raised by Kenwright (2018) do not seem applicable in this situation.

In her article, Rosalia (2019) used VR documentary stories of migration in the classroom. VR documentaries seemed to enhance the interaction with the outside world rather than minimize it.
This idea was also supported by Aman and Shiratuddin (2020). Voutira and Dona (2007) recognized that the use of virtual reality in research would include a pragmatical alteration of the methodological process. There are also plenty of researchers who utilized virtual reality to teach a target language (Cowans, 2018; Lin & Lan, 2015; Nersesian et al., 2018) or induce empathy for migrants (Camilleri et al., 2016; Grover & Pennay, 2020; Tikhonova et al., 2020).

2.2. Research on Migrants’ Social Identity

Social identity refers to the shared identifiers or groups brought on by society and internalized by individuals (Islam, 2014; Tajfel, 1978; Tajfel & Turner, 1979). Jenkins (2014) implies that these identifiers can refer to stereotypical categories that can influence behavior. In the literature these identifiers are seen as dynamic (Daynes, 2007; Haslam et al., 1992; Hogg et al., 1995; Jenkins 2014; Laitin & Watkins, 1998; Mueller et al., 2019; Norton, 1997; Verkuyten, 2016). The social identity of migrants is shaken by opposing forces of their social environment. Ochs’ (1993) term “social personae” (p. 288) referred to community prestige, positions, connections, roles, as well as corporate identifiers, which an individual selects or receives through social interaction. Social categorization can be different in the residential country than in their country of heritage, as such migrants construct their social identities differently from the natives.

Why is it important to study the social identity of children with a migrant background? Vietze et al. (2021) stated that being dismissive of the differences between children with a migrant background and native students can be catastrophic. She highlighted that expectations of uniformity between social groups could hinder inclusion. Different social categories include ethnicity, socio-economic status, geographic location, language and communication skills. In the vast diversity of the literature, researchers have focused on different aspects of social categories. On the one hand, ethnicity is argued to be one of the essential social categories employed in everyday life (Worrell, 2015) and for the construction of the self (Deaux, 2001). On the other hand, Gumperz (1982) proposed that social identity is predominantly founded and preserved through language. Furthermore, Miller (2000) implied that connections between the target language and social identity of migrants are associated with empowerment and enhancement of self-realization.

Praharso et al. (2017) indicated that loss of social identity could have a negative impact on their well-being. Social factors can influence migrants’ mental and physical health. Jetten et al. (2017) demonstrated that positive in-group associations could provide individuals with purpose, approval and power. Hence, positive in-group associations can help individuals that have been through difficult situations. Jetten et al. (2017) refer to this concept as a social cure, which significantly affects health. However, the writers state that when in-group categorizations are stigmatized, they can actually endanger or damage the individuals’ welfare. As a consequence, it is not only essential to perceive and understand the social identity of others, but also to recognize that some social categorizations might provide additional struggle. Finally, Alfadhli and Drury (2018) found that in many studies exchanging concerns, anxiety, or even anguish among people who migrated to another country, could lead to anticipation of assistance, which might transform the inflicted group into instigators of change and aid.

Moreover, the impact of social identity on migrants has been studied from different perspectives. For instance, Akerlof and Kranton (2000) discuss the correlation among social identity, social exclusion and poverty. The socio-economic factor of social identity in migrants is often an area of concern in the literature (e.g., Wang & Ni, 2005) as it represents power dynamics. Afridi et al. (2015) uses social identity to unravel inequalities that have a deep impact through China’s hukou system. In other cases, social identity is used to comprehend stereotypes about migrants and find alternative perspectives and strategies to tear down these stereotypes (e.g., McNamara, 1987; Stoyanov, 2018). Additionally,
Miller (1999) highlighted the idea that teachers’ understanding of social identity could lead to more effective second-language teaching for migrants. This is also demonstrated in Wang’s et al. (2019) work on academic performance among migrant children in Beijing.

Conclusively, social identity seems to affect a variety of significant aspects of a migrant’s life. These include socio-economic status, language acquisition, academic performance, health, levels of empowerment and self-realization. Thus, negative social identity classification should viewed as signs of trouble. Ignoring these indicators encourages further desensitization of the issues at hand and makes us accomplices to the creation of a cruel reality.

2.3. The assistive use of place identity elements

So far, virtual reality has not been used in research to detect the social identity of children with migrant backgrounds. One academic concern is the possibility of achieving the connection of places in 360° photography with the social experiences and expectations of children with migrant backgrounds. One way to achieve this is by examining elements of place identity. Proshansky et al. (1983) defined the term place identity as an assortment of memories, ideas, perceptions, understandings and emotions connected to a place. Since then, place identity has been extensively explored in the literature (e.g. Csikszentmihalyi & Halton 1981; Cuba and Hummon, 1993; Dixon & Durrheim, 2000; Johnson & Bibbo, 2014; Molony, 2010; Rosbrook & Schweitzer, 2010). In this study, Taylor’s (2009) and Arvanitis and Yelland’s (2019) dimensions of the relational, spatial, material, and temporal home are employed.

3. Research Methodology

The scope of the study is to construct a new and different way to investigate the social identity of migrant students by using VR technology. More specifically, it investigates the following:

- To what extent can the use of virtual reality and 360° photography assist students in revealing their identity?
- What are the parameters that should be considered?

The main research questions are investigated through a constructivist outlook. Constructivism is based on the premise that there is plurality in the definition of truth. Truth is created through social interaction. The most common way to exercise constructivism is through qualitative research. According to Labaree (2009), qualitative research focuses on the methods, clues and characteristics of an examined phenomenon. This study centers on two siblings of a migrant family and constitutes a case study.

3.1. Participants

This case study is intended to use a new combination of established methods to study the social identity of two male siblings with a migrant background. The brothers (N and F) are 11 and 15 years old. Their parents have emigrated from Romania to Greece. However, they live in a complex linguistic environment with different languages spoken by the parents. Thus, in order to have a holistic view of the social identity of their children, it is also important to engage with the parents. This family was chosen because it constitutes an interesting example. The father is a migrant Minister. Both parents are college-educated and, at the same time, very close to their religion. They are migrants, but their father has a leading role in their community as a Minister. This initial information is enough to spark an interest in these possible contradicting social identity aspects and their effect on the social identity of their children.
3.2. Research Tools

The data collection is conducted through three different research tools: semi-structured interviews, social location maps and identity texts. A semi-structured interview contains some flexible themes or questions prepared (known as an interview guide). Interviews with the parents and the children are conducted based on predesigned different interview guides. Firstly, interviews with the children were carried out in order to collect some introductory information about them. At the same time, the main questions of the interview guide were used to form the VR experience. Secondly, interviews were carried out with each of the parents individually. To achieve adequate detection, personal identity as well as social identity wheels, which were created from the LSA Inclusive Teaching Initiative of the University of Michigan (2017), were used as interview guides. Additional connective questions between the parents and the children interview were added.

Following the interviews, the VR activity was carried out with the children. The activity was conducted on a different day than the interviews in order to ensure that the students were well rested. Each child worked individually with the VR. They created social location maps and identity texts of their choosing. In the end, the students came together for a discussion and a comparative reflection. The term social location map has been adopted from Zorbas’ and Karras’ (2019) work. A social location map (SLM) consists of a net illustration with diverse destinations corresponding to aspects such as class, gender, social status, ethnicity. Identity texts are defined by Cummins et al. (2011) as a creative, original means for student participants to showcase their identity. Identity texts, not unlike SLMs, are flexible research tools. They can take on any form the participants choose.

3.3. Data Analysis Validity in Research Practice

Creswell (2014) underscores the importance of validity. Validity refers to the dependability of the data and their examination. In the present study, thematic analysis is employed in order to organize the information derived from the semi-structured interviews. SLMs and IDTEs are analyzed more closely using themes as well as keywords. The coding of the data is manual as well as inductive. The collection of themes is interpreted through Creswell’s (2014) established approaches that secure genuineness. These approaches include detailed and dense descriptions, classifications as well as triangulation. Triangulation is exceptionally significant and can be achieved because of the multiplicity of the methodological tools. Rich descriptions are also employed to secure deep analysis and close examination of the data. Other strategies proposed by Creswell (2014) are prolonged engagements, clarification of bias and peer debriefing.

4. Results and Discussion

4.1. Some initial data from the interviews

In all the interviews, it seemed that there were different levels of masking difficult situations or emotions. It was apparent that in many instances, the participants tried to avoid answering questions associated with cultural origin. Instead, they deflected the discussion to more personally comfortable areas of social identity such as religion. The semi-structured interviews provided a variety of information about the participants. The most significant issues associated with the research was the detection of introvert characteristics as well as the importance of religion and language.

4.1.1 Detection of Introvert Characteristics
The children’s statements revealed that they faced cultural struggles, but they may not realize or desire to share where this difference in treatment derives from. This could also be connected to the lack of discussion of such issues with their families. According to Worrell (2015) and Deaux (2001), ethnicity is of monumental importance for constructing the self. Moreover, overlooking or concealing aspects of social identity, as seen here, was described as potentially harmful.

Another explanation for their reluctance was extracted from M(Mom)’s response. The parents did not see their kids as outsiders of their community. M characterized herself as Romanian in her identity wheel, but when talking about her children, she said that they were half and half. It seems that she sees her children gaining more of an insider status compared to her. Hence, it might be that the parents did not expect that the children would face similar extended difficulties as they did. This idea was also supported by the father (P). He painted a picture of "things" now being better than before regarding discrimination in Greece. Praharso et al. (2017) and Vietze et al. (2021) have previously highlighted the negative effect that the dismissal of one’s social identity can have on one’s well-being.

4.1.2 The Parent’s Personal and Social Identity Wheels

During their interviews the parents verbally completed personal and social identity wheels. As previously mentioned the wheels derived from the University of Michigan (2017) and were translated appropriately to fit the linguistic repertoire of the participants. The wheels assisted in indicating their personal and social identities as well as function as conversation starters.

Image 1. M’s Personal Identity Wheel
Image 2. M’s Social Identity Wheel

Image 3. P’s Personal Identity Wheel
4.1.3 The Importance of Religion

Religion is one of the most prominent identities showcased by the parents. Both parents have a bachelor’s degree in Theology. M mentioned that she enjoys reading religious books while her husband enjoys listening to religious music. Additionally, M chose religion as the most important identity in her social identity wheel. However, there are significant parts, which separate P and M from the stereotypical profile of a priest and the wife of a priest because they are also migrants. It seems that F balances both identities by combining and expanding fundamental values that engulf them. He notably said:

“Look, I try the same. To teach about love, so that the other person understands it…. We have an obligation to make humans understand what love is. And I always say in my sermons that Christ is love. Not only of Religion, but it really is. If you love the other, you will achieve everything. If you do not love, you will not achieve anything in this life”

M also mentioned that her personal motto is that people should not judge each other. She seemed to receive also significant societal expectations as the wife of a priest. Hence, religion is very significant for the parents, but it is also being filtered through the lens of migrant life. The importance of religion is also reflected by in the children’s choices in their SLMs and IDTE’s. Moreover, F and N indicated in their interviews stressed their involvement in religious practices during their interviews. For instance:

“I: ... What did you like most about the monastery?
F: The lifestyle that we had there. We would help with the cows and :::: the hay… eh: and the sermons, we conducted them later around 8:00. Thus, we woke up later. Better.
I: That’s nice. When you are here, what time do you wake up?
F: Here I wake up at 6. We start at 7.
I: Do you conduct the sermons with your father? Do you help him?
F: Yes"

4.1.4 Importance of Language

Gumperz (1982) posits that social identity is predominantly founded and preserved through language. On their previously presented social identity wheels both parents list Romanian as their first language. However, P also listed Moldovan as his first language.

«Romanian because Moldova is the same. It is not that Eh:: has another language. Romanian itself. We write the old Cyrillic (script). Now Romanian. (Moldova) It was an old Romanian state and then they separated and it is an independent state now.»

Through additional probing, it was revealed that P originated from Moldova but is a Romanian citizen. Additionally, P noted that he considers Greek, Russian, Italian and Portuguese as second languages. M also referred to Greek as a second language along with English and French. She can speak, read and write all of these languages. Consequently, the linguistic environment that the children grow up in is extremely rich.

4.2. Virtual Reality Activity

Interviews were carried out in order to extract some initial information from the children. This information was used to shape and personalize the VR experience. During the interview, the children were asked to choose three places to visit virtually. They could choose one place from their country of heritage (Romania), one from the country they live in (Greece) and one from any part of the world (preferably a place they long to visit or stay in the future). In brief, N chose a Monastery at Curta de Arges (Romania), a monastery on Mount Athos (Greece) and a hotel or view of Tokyo (Japan). On the other hand, F decided on Bran Castle (i.e., Castle of Dracula) (Romania), the Mycenaean Ancient Ruins (Greece) and the Camp Nou stadium of Barcelona (Spain). The VR activity was conducted on a different day in order to ensure that the students are well-rested. Each child worked individually with the VR. They created social location maps and identity texts of their choosing.

More specifically, after virtually visiting each place, the brothers created an SLM depicting their social identity in each place. During each VR visit, they were asked to notice their feelings, memories, languages they would use in the place, social status and other social identity aspects. These themes are all valuable for the detection and realization of social identity. It is important to note that the participants were not obligated to respond. They could respond, or quietly contemplate these aspects. The creation of each SLM was carried out immediately after each virtual visit. This design detail is significant. The children should have the opportunity to immediately note the aspects of their social identity. If this step is not adequately followed, memories, experiences, and ideas can be intermixed between places, or other aspects might be forgotten or overshadowed.

After completing all three social location maps on the same piece of paper, the children were given time to reflect on their work. They had the opportunity to view their three social location maps as a whole, while noting the connections and contradictions happening between the different places.

The next step for each child was to create their own identity text. As defined by Cummins et al. (2011), these are creative, original means for student participants to showcase their identity. In the end, the brothers came together for a discussion and a comparative reflection. The structure of the study was designed to have self-reflection as well as mutual reflection.
Before the activity, the participants were asked about their perceived significance of the VR. N said that using the VR would be different from just thinking about places. In N’s own words:

“Eh:: I will see the place in reality and it will be like I am there... I will feel much better because when I am not here::: and I do not feel very well (whispering the last phrase)”

When F was asked if his feelings would change after using the VR glasses, he said:

“eh maybe a little yes... I would want to visit that place more”

4.3 Social Location Maps (SLMs)

4.3.1 N’s Social Location Maps

The first image demonstrates N’s creation. The red arrow shows the order in which N saw the places. They correspond to the places he chose in Romania, Greece and Japan accordingly. Each bubble demonstrates social identity and place identity aspects connected to the location he visited virtually.

While he was creating each part, he mostly made connections between the aspects that corresponded to the same place. When he was finished, he was given time to think if there were any cross-location connections. These connections across aspects of different themes signify either commonality or change. N made six cross-location connections. Five between the Monastery in Romania and the Monastery in Athos and only one between all three of them.

To untangle N’s complicated SLMs, a more organized form was created. All aspects were translated into English. Inter-location connections were marked with the same color as their location category, while cross-location connections were marked with green in order to stand out.
Place identity values we’re used to create a more comfortable environment and immerse the child into the VR experience, making their memories more vivid. Dixon and Durrheim (2000) characterized place identity as undeniably social. More vivid experiences assisted in the gradual progress from easy to sensitive social identity aspects.

4.3.2 F’s Social Location Map

F’s SLMs were structured similarly to N’s. There was an arrow delineating the order in which the places were seen. The first place was in Romania, the second was in Greece and the third was in Spain. F also placed each aspect (bubble) roughly underneath the title of the location it belonged to. Contrary to N’s SLMs, he did not use connectors between all the bubbles and the titles. F used only connectors to demonstrate a relationship between the aspects, such as commonality or change.
A translation was made in order to make F’s SLMs more comprehensible (image 5). Similarly, different colors were used to indicate where each aspect belonged. Overall, he made 14 individual connections. These connections can be grouped into five categories. F found more cross-location connections than his brother. However, both brothers have overlooked some of the connections they could have made. That can be thought of as a valuable hint about the connections they recognized and those they did not.

Similarly, to his brother, F recognized more connections between the first two places than connections involving the third place. For both participants, the first two places represented the present and the past. On the other hand, the third place represented the future. Hence, it is natural to find more connections between places that were in their life at the moment than places they aspired to visit.

4.4 Discussion of SLMs

Multiple researchers have characterized social identity as dynamic (Daynes, 2007; Haslam et al., 1992; Hogg et al., 1995; Jenkins 2014; Laitin & Watkins, 1998; Mueller et al., 2019; Norton, 1997; Verkuyten, 2016). In theory, one is given the impression that social identity can change gradually through time or due to an unexpected event. However, as demonstrated in the findings, social identity can change just by visualizing oneself living in a different place. It was extraordinarily remarkable to see participants being immersed in their virtual environment and realizing that some of their identities would shift depending on the place.

Ochs’ (1993) term social personae (p. 288) subsumes identifiers created through social interaction. However, in this case, there was no real-life social interaction. Just by virtually visiting different places (even places the children had never visited before), the brothers noted changes in socio-economic levels, languages, feelings and freedoms. There were also aspects, which the brothers expected to be consistent at least between two places. These referred to community prestige and connections associated with respect, hospitality, religion or their positive expectations regarding their future life. Their SLM’s provided valuable information on the children’s viewpoints and understandings of migration. The brothers traced many fundamental shifts that migration entailed. This could be the result of their experiences filtered through the lens of their
migrant background. In future research, it would be interesting to compare how natives would respond to a similar activity. Would they be able to detect as many changes in social identity as children with a migrant background? This idea could provide information on how children with a migrant background view their social world.

The brothers were more engaged when given time to work on their social identity through the VR activity combined with their SLMs. It became apparent that their social identity was not flat but flexible. Their flexibility was showcased through their depictions of social identity, which developed differently in various virtual geographical locations. Moreover, their SLMs’ flexibility was also evident through their ability to include a variety of elements. These elements include aspects such as smells and views, which are closely linked to place identity.

Their preferred Greek environment was also illustrated through their SLMs. By vicariously traveling to their favorite Greek places, the participants showcased representations connected to their ideal Greek reality (see table 1).

<table>
<thead>
<tr>
<th></th>
<th>Words used in each desired Greek place</th>
<th>Perceived ideal Greek place</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>• respect</td>
<td>An outlook of relaxation with family and friends in a secluded place away from the city</td>
</tr>
<tr>
<td></td>
<td>• the smell of incense</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• greenery</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• friendly</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• resting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• high</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• empty</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• family &amp; friends</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>• freedom</td>
<td>A safe, simple place in nature surrounded by freedom and different cultures.</td>
</tr>
<tr>
<td></td>
<td>• friendly</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• without dangers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• simple foods</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• the smell of flowers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• History</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• multiple languages</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Information about the Greek Virtual Places retrieved from the SLMs

F’s ideal Greek reality seemed to signify change from their current environment. His ideal place portrays a safe, multicultural reality, where he can be free—an environment, which is desired by every child, is culturally and linguistically diverse and safe. On the other hand, N’s ideal place in Greece was a more secluded, relaxed space that seems to represent detachment from his Greek community. N’s ideal place symbolized the enjoyment of family and friends but away from the rest of the world. Thus, in his place, safety and happiness were succeeded through isolation from the Greek reality. Comparing the two perspectives, it seems that F’s “reality” was more hopeful, while N’s was more drastic and protective.

4.5 Identity Text Representations

Prior to creating their own identity texts (IDTE), students were given different examples of identity texts of other students in order to understand what an identity text was. Identity texts were presented in different forms. They both individually decided to use a piece of paper and a black crayon. In
comparison, however, their SLM were a bit more colorful. Nevertheless, their IDTEs' simplistic style did not detract from the significance of the results.

4.5.1 N’s Identity Text

In both participants IDTEs, some elements reaffirmed some associations made in their SLMs. However, at the same time, additional themes emerged. N’s IDTE (image 6) included some of the social identity aspects of his SLMs. It is important to highlight that N decided to represent himself with the Japanese sword, the katana. This is also an inspiration deriving from his favorite anime show. However, it can also be associated with power. His logic behind his IDTE seemed to be that he placed words, which were most important to his identity, closer to the sword, and words that were less important but still part of his identity, away from the sword. Hence, the most essential identities for N were described by the words: boy, anime, Romanian, English, Greek, Japanese, katana, PC and his brother’s name, which was redacted. It should be highlighted that this is the second time his brother’s name appeared in a significant representation.

Apart from his gender and computer, all the other aspects important for his identity can also be found in his SLMs. This reaffirms the importance of these elements. Other elements such as Christianity, Greece, Romania, and Japan were placed in a medium distance from the sword. However, new elements began to appear as well, like his favorite food, his love for music and smartphones as well as the adjective smart. The word “smart” is written in bigger letters indicating confidence about his mental abilities. Further away from his representation are few words such as respect, friendly, tablet and little money. The words little money are also included in his SLM where he wrote that we would make less money if he lived in Romania. It is no coincidence that his choice of little money is closer to the word Romania. This seems to represent that N expresses support towards the heritage part of his identity, even if it will bring a lower economic status. Thus, for him heritage is a more significant identifier than his socio-economic status.
4.5.2 F’s Identity Text

Regarding F’s identity text, there are some similarities but also some interesting information about him. F drew a picture of himself in the middle of the paper. The artwork is of no importance, but the facial expression is quite surprising. As mentioned before, F seemed to be an introvert and did not express any similar emotions during the other tasks. Moreover, his quiet personality on the surface seems to conceal emotions associated with quick-temperament, which refer to the words used on the top of his IDTE.

Furthermore, it is essential to examine the identities that seem more valuable to F. Most of the words have close proximity to his representation. Hence proximity cannot be taken as the only deciphering factor. A significant difference is that the initial letter of some words was capitalized. By comparing these words, it seems his choice is not associated with grammar. The words with the initial letter capitalized are boy, Romania, orthodoxy, English, trips, movies, Netflix, respect, to make a lot of money, family, friends, clarinet and Greece (which was written twice). By comparing these words with
N’s most important identities, a pattern seems to emerge. Both F and N seem to include gender, languages, religion, family and entertainment (e.g., pc, movies) as the core part of their identity. A significant difference is that F seems to give more importance to the value of money than his brother. This is expected as F is a teenager and has a better grasp of economic issues. This was also evident in his SLM, where he noted that there were not many jobs in Greece. Thus, his future wages would be lower.

4.7 The Process of Reflection

During the final reflection, F and N were a bit playful. They laughed and teased each other. They exchanged some meaningful information, but some of their guarded characteristics returned. F hypothesized that his brother chose the two different monasteries in Greece and Romania because they conjured up pleasant memories. N also tried to find social location destinations with which he could tease his brother, such as a "moldy smell". However, his playful demeanor unintentionally gave away more data than he had intended. Following that, they managed to recognize common identities such as the Orthodox religion, their sex (boy), music, their languages (English, Romanian & Greek), as well as objects such as the computer or smartphone. The identities deriving from objects seemed to be really strong and more comfortable to talk about. Overall, they recognized 11 common identities. This was a substantial amount. However, it seemed more important for them to maintain their uniqueness than finding a common ground.

It seemed that this kind of open and vulnerable engagement was something quite unfamiliar to them. The children were very reserved at times, not making eye contact with each other and trying to avoid sensitive subjects. Instead of becoming instigators, as Alfadhli and Drury (2018) suggest, they released feelings of awkwardness. In future practice, the structure of the reflection should be altered or organized differently in order to assist introverted children. New methods can be employed for children who are reluctant to share their emotions or engage in the reflection more deeply.

5. Final Triangulations and Data Discussion

The children were significantly more reluctant to talk about Romania during the interviews than when creating their SLMs and IDTEs. After completing their SLMs and
IDTEs, the children were more relaxed and open. They revealed more personal elements about themselves. In short, the most significant social identities of the children included languages, religion, ethnicity, geographic-location, socio-economic status and in some cases technology-media consumption.

IDTEs revealed additional parts of the participants’ identity. IDTEs showcased the aspects of their social identity that have meaning for them personally. Some of the identity constituents were identical in both their SLMs and IDTEs. This reaffirmed their significance of their social identity. Some features revealed in the IDTEs seem to be connected more closely to their personal identity than their social identity. Jenkins (2014) stated clearly that social identity differs from personal identity. Personal identity is ultimately centered on their unique identifiers (Goffman, 1968; Jenkins, 2014). On the other hand, social identity refers to the identities shared by other people or groups created through societal interaction (Islam, 2014; Tajfel, 1978; Tajfel & Turner, 1979). Hence, some of the additional aspects revealed in the IDTEs could be under the umbrella of personal identity. The supplementary data illustrated in the IDTEs can be considered as an additional benefit of the study.

Conclusively, semi-structured interviews and identity texts were flexible (Cummins et al., 2011) but only revealed some aspects of one’s identity. This approach was more static. One has to keep in mind that the elements of one’s identity can change over time. However, SLMs showcased the constant change of social identity in practice. All directions have their benefits. IDTEs included more broad ideas of one’s identity, while SLMs demonstrated ways one’s identity can change. Nevertheless, all tools were valuable for the triangulation of the results.

Regarding the first research questions, virtual reality and 360° photography were indeed assistive in revealing and demonstrating the participants’ identity. This is illustrated by the abundance of information and elements produced by the children through their SLMs and IDTEs. Even information from the semi-structured interviews pointed towards the fact that the addition of VR can assist in the recollection of the participants’ memories. In some areas, the findings were reaffirmed by the literature. For instance, language and ethnicity have been the most critical factors of social identity. It was also shown that the geographical location, which is a part of one’s social identity, can influence other social identity aspects significantly. Moreover, place and social identity worked very well together. Thus, the incorporation of VR could be one viable choice in the detection of one’s social identity.

Moreover, regarding the second question, the intended slight innovative spin was more significant than expected. The combination of place identity, SLMs and the VR activity brought forth richer and more profound results than was expected. Another unanticipated feature referred to the demonstration of social identity fluidity in the SLMs. The minute-by-minute changes of social identity were incredibly apparent. The children with migrant backgrounds who participated in the study seemed to understand that some of the aspects of their social identity could change depending on the location or time. Finally, the participants’ ideal Greek places proved very valuable in understanding their view on the ideal host environment.

6. Conclusions & Limitations

As in every study, it is apparent that there are limitations. In this study there is danger of epistemological relativism. Prolonged engagement is needed to be able to generalize the benefits to a wider population. However, this study can be a promising start. Additionally, the structure of the study makes it easily replicable and approachable. The low cost of the material needed (hardware and software), combined with the use of everyday devices (e.g. smartphone), contributes to the
accessibility of the study which can be easily replicated in schools by teachers or other researchers. Although more expensive equipment may have made the experience even more appealing, it would only be used by the few who can acquire it. Thus, services or local schools would not pragmatically be able to benefit from this experience. Education and understanding of minorities should be attainable for everyone. This research showcases that remarkable results can be deduced without having to spend a fortune. Based on that, the application of VR and social identity can be implemented in refugee camps. VR can assist refugees in virtually visiting places, which are inaccessible to them. Thus, their social identity can be more vividly portrayed while providing a unique experience.

References


Kyriaki Thoidou (kyriaki_thoidou@hotmail.com): She holds two MAs, one in Inclusive Education & Technology from King’s College London and a second in Language Education for Refugees and Migrants from the Hellenic Open University. She has worked for the last seven years in special and general education and volunteered for the education of marginalized groups. She also won the King’s Experience Global Award.

Vasilios Zorbas (vzorbas@enl.uoa.gr): He obtained a B.A and M.A in English Literature from New York University. He then furthered his graduate studies at the English department of Columbia University (Teachers College) where he pursued an Ed.M in TESOL and completed his graduate work with an Ed.D in English Education specializing in Educational Drama. As an NYC permanent licensed English teacher (with tenure), he taught English language, literature and drama courses. He also taught English Education courses at Hunter College and was a research associate at the Research Centre for English Language Teaching, Testing and Assessment of the University of Athens. Currently, he is a tutor at the Hellenic Open University. His main research interests are: Drama in Education, Alternative Forms of Assessment, Curriculum Design and Teacher Education, Qualitative Research. He has publications in these areas in local and international journals.
The battle of language learning apps: a cross-platform overview

Athanasios KARASIMOS

Learning foreign languages has become a very appealing and important ability in the contemporary world. Technology, nowadays, permeates all aspects of people’s lives, and it has consequently also reached the domain of education. Oftentimes, learning a foreign language involves using language learning platforms (LLPs). There is an abundance of language learning software tools available, but we can only find a small amount of research specifically dealing with such tools and investigating their effectiveness in language learning. This paper narrows down the focus on five language learning platforms. It is a comprehensive evaluation of the popular and rapidly growing online language learning platforms, such as Duolingo, Rosetta Stone, Memrise, LingQ and Busuu. In the process of this evaluation, the participants will present features of the five platforms and underline their strengths and weak points, while addressing certain limitations and providing suggestions for further improvement of the learning platform.

Key words: language learning platforms, asynchronous education, distance learning, Duolingo, Memrise, Rosetta stone

1. Introduction to machine learning dominance

Mobile-assisted language learning (MALL) is one of the most important recent developments in the field of computer-assisted language learning. In MALL literature there is an increasing number of studies that examine various mobile devices in formal and informal learning environments. Along with the advancement of mobile devices such as smartphones and tablets, a large number of educational apps have been developed and are widely available in several popular platforms and operating systems (Ho 2018; Sánchez et al. 2019). The number of language learning apps is also rapidly growing, and it is a dominant trend for promoting autonomous and self-teaching learning (Rodgers & Weatherby, 2021; Tao et al. 2020). This chapter explores mobile apps for language learning and centres on evaluating these apps. It introduces a language learning app review form, which guides language teachers to critically evaluate the pedagogical and technical aspects of language learning apps. The review form contains fifteen evaluation criteria: purpose, accuracy, usefulness, flexibility, authenticity, engagement, feedback, integration, support, price, reliability, presentation, organization, navigation and multimedia. The chapter also presents a list of example
apps selected for reading, writing, listening, speaking, vocabulary, grammar, pronunciation and culture learning activities with a particular focus on learning English as a second/foreign language.

In the decade or so since smartphones were introduced, these and subsequent devices such as tablets have become quite popular and abundant: they are now used by large proportions of the population in most developed countries. Mobile applications (commonly referred to as apps) can be installed on these devices, and offer an enormous market of educational resources, including language learning. Early claims about the potential of apps for language learning included the possibility to learn, practise, and enhance both language skills and cultural knowledge (Rosell-Aguilar, 2009; 2018).

2. The theoretical background: re-defining distance learning

2.1. Towards a theoretical foundation

Learning in the 21st century has developed beyond physical and traditional borders. Learners have access to knowledge since technology and Web 3.0 (nowadays we are in the dawn of Web 4.0) is present and dominant in the learning procedure. The Gen-Z (and only) learners urged by the pandemic and some additional reasons choose distance, autonomous and asynchronous learning rather than face-to-face teaching (Garcia & Guzmán, 2021; Lei & Liu, 2020, Yin & Chik, 2021). King et al. (2001) provided separate definitions for the terms distance learning and education since they defined “distance learning” as the activity of learning at a distance and not an alternative version of “education”. Additionally, place, medium and time are the major limitations which are associated with the term distance (Drutsko, 2020; Dubakov & Kuzmenkina, 2021; Guilar & Loring, 2008). All these dominant features of distance learning are usually present in any language learning platform (LLP) application.

On the other hand, in asynchronous education, the learners may attend courses, participate, finish their tasks and activities at their own pace and according to their schedule. On the contrary, synchronous education is immediate, unpersonalized, real-time and easier for the learners to interact with their instructors or co-learners. Asynchronous education is based on three significant characteristics, such as flexibility, availability and autonomy, which are implemented in LLP tools. It is a balanced combination of three dominant and well-known theories: (a.) theory of independence and autonomy (Wedemayer, 1981), (b.) theory of industrialization (Peters, 1988) and (c.) theory of interaction and communication (Holmberg, 1986). Despite the enormous tech evolution of distance education platforms, apps and services, the theory behind-the-scenes of distance learning remains chaotic, uncharted and confused. As Moore (1997) opposes, there is no national policy nor anything approaching a consensus among educators of the distance learning’s value, a commonly accepted definition, a concrete or broad methodology or even the concept of distance education. Since Shale’s (1990) last call for theoreticians and practitioners to stop emphasizing points of difference between distance and traditional education, there was not much effort to identify common educational problems and to build a theoretical framework to support all these apps, tools, services and features. Nevertheless, during the last five years several research (Gonzalez-Vera 2016; Hohol, 2020; Koriahina 2021; Sam, Li & Chatwin, 2016; Worden 2021) have provided a common ground of characteristics and features that are shared in LLPs.

2.2. Identified essential elements of autonomous (distance) learning and language learning apps

The core notion behind autonomous learning has been structured with all the necessary and
essential elements to create and support the independent learners. Furthermore, distance learning promotes a greater student responsibility and a widely available instruction to cover each user’s needs, preferences and different learning patterns and behaviors. For instance, the framework of RASE (Resources, Activity, Support, and Evaluation) learning design (Churchill, King, & Fox, 2013) is a blended learning theory with the aforementioned significant elements of autonomous learning focused on writing skill. Students can access the resources on the content management systems, such as Edmodo, Canvas, Schoology, Moodle or Open class (Rakhmawati, 2020). Moreover, effective mix of media and methods under the umbrella of blended learning offers an adaptation to individual differences and a wide variety of start, stop, and learning times.

MALL has been recently promoted to cut down the barriers on language teaching and learning (Burston, 2021; Nielson, 2017; Vemula 2020). MALL can also enable students become more active, autonomous and collaborative because they contribute to their own learning path. Rewinding to Wedemeyer’s (1981) theory, the designers of language learning apps should reconsider the behind-the-scenes theory of their code. Meaningful learning and new learning anchored in the cognitive structure (not rote learning) are key elements to boost the effectiveness of these apps. Several research (Citrayasa, 2019; Malerba, 2015; Munday, 2017; Psychogyiou & Karasimos, 2019; Stockwell & Hubbard, 2013) present data from users who request apps centered on interest, with individualization of teaching and learning and supporting encouragement of critical thinking. Unfortunately, the astounding growth of new language learning apps (more than 70), glamorized by technology, amplifies the poor conceptual framework and the usual absence of identified essential elements and the 4-Square Map of distance education technology options (Gafurova 2021; Paris et al. 2021; Stock, McIsaac & Gunawardena, 2004).

3. Language learning platforms: Learning a language from distance

3.1. Previous research on the use of apps for language learning

Several research studies (Castaneda & Cho, 2016; Gafurova, 2021; Morgana, 2015; Vemula, 2020; Worden, 2021) into the use of language learning platforms and apps have presented positive results on language skill improvement, learner motivation and engagement; nevertheless, as Rosell-Aguilar (2018) points out, the researchers have mostly focused on rather small subject groups using pre-selected apps or platforms rather than on the choices made by the language users. The major impact factor of using language learning apps is the significant and prominent improvements in all language skills, grammar and vocabulary, and even in more advanced linguistic abilities (Castaneda & Cho, 2016; Morgana, 2015;). The effectiveness of LLPs has also been found comparable to face-to-face teaching at several CEFR level students. Rachels and Rockinson-Szapkiw (2017) found that there were no significant differences between the Spanish subjects that learned a language via Duolingo and the other group that participated with a traditional teaching approach. Similarly, Psychogyiou and Karasimos (2019) support that the vast majority of the participants was confident about the result, although there were some contrasting views which has produced a debate about the effectiveness of the Duolingo app. However, the users were pleased with learning new vocabulary and pronunciation compared to the typical coursebook approach.

The language learning apps not only promote autonomous language learning, but additionally they offer an alternative reality to the educators and students that want to use these apps in a classroom context. Chen and Kessler (2013) maintain that smartphones and tablets may enhance learner autonomy and ubiquitous learning in informal settings. Unfortunately, only a few studies (Steel, 2012; Stockwell & Hubbard, 2013) present results from LLPs and MALL activities in a face-to-face classroom compared to how the users engage in mobile-assisted language learning outside-of-the-
box (a traditional or even a techy classroom). Moreover, Stockwell and Hubbard (2013) study the learner's use of LLPs outside the designer's guide and suggestions and observe that regularly the learners' personalised use of apps may be deviated from what the best practices or ideal usage of each application is considered. In Mason and Zhang’s (2017) find that the majority of Chinese users recognised the value of apps to support their language learning, but learners only used a proportion of the available functionalities (Rosell-Aguilar, 2018, p. 6). For an extensive presentation of previous research on the use of apps for language learning, check Rosell-Aguilar (2018) and Goodwin-Jones (2011) for early language mobile apps.

3.2. Selection criteria for language learning platforms

The ongoing and impressive battle between the learning language apps has raised the need to provide and categorize some concrete and straightforward selection criteria for LLPs. On the other hand, the users’ reviews and the learners’ community feedback generated enough arguments for the researchers to modify and shape a best practice guide for the LLPs designers and developers. Whilst it is arguable whether apps can be considered at this point as a single solution to language learning, they can effectively support learner autonomy and interest in learning a language (Goodwin-Jones, 2011; Worden, 2021). Apps can provide a good supplement for language learners who are enrolled in formal instruction, as well as a good starting point for beginner independent learners. They can also provide regular practice for language learners who are no longer formally studying a language but wish to keep practising it (Rosell-Aguilar, 2017; 2018). The number of language learning apps grows rapidly the last decade, and the increasing accessibility of the apps generates the necessity for the appropriate selection of the apps. This session provides several evaluation systems of language learning apps.

As Smith and Ragan (2004) pointed out, examining the content, task, and context of the instructional materials or programs was important. Thus, evaluating the quality of the content provided by apps is important. Given that the apps are used for learning, it is vital to assess the pedagogical coherence of language skills within the learning activities. Apps are software installed on mobile devices; therefore, evaluating the usability, customization, and sharing options provided by such apps is valuable. In summary, seven elements are identified to evaluate language-learning mobile apps: feedback and self-correction, customization, pedagogical coherence on language skills, content quality, usability, motivation, and sharing (Chen, 2016).

Heil, Wu and Lee (2016)’s criteria are languages supported and platforms availability, monetization, user input, assessment and instructional focus, implicit and explicit grammar instruction, corrective feedback, and user interaction –listening, speaking, reading and writing. Their study points out that is a common pattern on teaching language to isolate basic and advanced vocabulary words via LLPs rather than contextualized usage. They observe that some drilling-type techniques offer very little explanatory corrective feedback, and there is little adaptation to the needs of individual learners. Despite advances in language teaching that have stressed the importance of communicative competence in language learning, MALL technology is still primarily utilized for vocabulary instruction rather than fluency building.

Additionally, Son (2016) introduces a language learning app review form, which guides language teachers to critically evaluate pedagogical and technical aspects of language learning apps. His evaluation form contains fifteen selection and reviewing criteria: purpose, accuracy, usefulness, flexibility, authenticity, engagement, feedback, integration, support, price, reliability, presentation, organization, navigation and multimedia.
Despite the perplexity or simplicity of a straightforward, well-organized and clear selection grid, choosing the right language-learning software is a highly personal decision and strongly related to each user’s needs and requirements. Bulatovych (2017) suggests creating an app for effective language learning by making learning easy and engaging for beginners and challenging for advanced learners. She comments that several LLPs take a more serious approach to vocabulary teaching, giving workloads that only advanced or intermediate learners can cope with; therefore, these apps usually offer complex and challenging quizzes with an emphasis on reading and memorizing words and phrases. Moreover, she filtered plenty app reviews to propose specific categories / criteria, such as personalized approaches to learning vocabulary and grammar, listening comprehension, pronunciation checks, language practice, training and checking progress, motivation and support of learners.

The different perspectives of learners and designers are mainly focused on intentions and experiences (there is a significant gap between them based on research of Stracke et al., 2018 and Hei, Wu & Lee, 2016). For this reason, we try to build a concrete and compact proposal to overcome this gap and fulfill their needs. Our selection and evaluation model is mostly based on Heil, Wu and Lee (2016), but it is expanded and modified to cover a wider area of language learning apps and their features, both from a selection point of view and from an evaluation perspective. The user wants a program that is adequate for his/her language education level, whether he/she is a total beginner or an advanced learner in need of some brushing-up or vocabulary refresh. The user also has to make sure that the language he/she wants to study is available, well-supported and not in an early-access version with issues and bugs. What about making sure you can study while driving or travelling? Almost every user’s life is becoming ever faster, and this situation enables users to take advantage of an app on the fly. Therefore, the option to swing from active to passive learning procedure due to some situations (driving, travelling, working, writing, etc.) is crucial.

<table>
<thead>
<tr>
<th>Selection and evaluation criteria</th>
<th>Values and guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Languages</td>
<td>Number and variety of language courses (including beta courses and constructed language courses)</td>
</tr>
<tr>
<td>Platforms</td>
<td>Operating systems: Android, iOS, Windows, Linux</td>
</tr>
<tr>
<td>Registration and subscription</td>
<td>Free registration, trial version or premium account (monthly cost or lifelong cost)</td>
</tr>
<tr>
<td>User interaction</td>
<td>Input options, such as writing on keyboard/ touch board, speaking into mic, typical or advanced use of touch screen, using of mouse</td>
</tr>
<tr>
<td>Support and updates</td>
<td>Frequency of updates, technical support, adding new features, bug and fixes</td>
</tr>
<tr>
<td>Vocabulary teaching and assessment</td>
<td>Vocabulary in isolation (with flashcards, video selection, picture pairing/ matching), vocabulary in context (inside monologue, dialogue or a story), supporting dictionary/lexicon</td>
</tr>
<tr>
<td>Grammar teaching and assessment</td>
<td>Autonomous grammar section, grammar in isolation (examples, tables), grammar in context, absence of grammar instruction</td>
</tr>
<tr>
<td>Reading, Writing, Listening and Speaking (+Pronunciation)</td>
<td>Reading texts and comprehension tasks, word formation processes, sentence structure building, listening to phrases or passages and comprehension or transcription tasks, dialogues or speaking interaction</td>
</tr>
<tr>
<td>Corrective Feedback</td>
<td>Sound effects, visual feedback, textual corrections, textual explanations, no feedback</td>
</tr>
</tbody>
</table>
Gamification features (experience points, levels, leaderboards, badges, challenges, etc.), motivation messages (emails, messages, pop-up notifications), gamification absence

Placement test for language levelling, progress tests, no tests

Connect with social media friends, connect with other app users, forum, bulletin lists

Table 1. The language learning platform selection criteria

4. Choosing the proper language learning app

4.1. The Duolingo application

Duolingo exists since 2009 and was nominated as iPhone App of the Year 2013 and Google’s Best of the Best app for 2014; it offers 68 different language courses across 23 languages with more than 300 million registered active users. Duolingo provides written lessons, dictation, an interactive dictionary (plus Stories and Events for several languages). It has a gamified skill tree that users can progress through and a vocabulary section where learned words can be practiced (Huynh & Iida, 2017; Psychogyiou & Karasimos, 2019; Rachels & Rockinson-Szapkiw, 2018). Duolingo enhances the learning experience by increasing access to the target language and allowing learners to work at their own time, level, and pace, thus making learning more exciting and personalized (Nushi & Eqbali, 2017). Moreover, Munday (2017) studies how the translation-based platform offers different possibilities for Spanish learners, while its gamified features provide an entertaining way for students to learn a language. Rachels and Rockinson-Szapkiw (2018) provide evidence that students learning a language with a free app like Duolingo, which applies gamification principles, exhibit equivalent learning outcomes with students being taught in a traditional face-to-face setting; their research indicates that Duolingo can have a substantial impact on foreign language learning. Finally, Leão and Amorim’s (2016) study present that Duolingo may engage and aid users in foreign language learning, but up to a certain extent. That is, the participants perceive that the presence of a teacher is still necessary to guide them in their learning process, and to foster interaction and production skills. When it comes to teaching and learning with technology, the need for teacher presence also depends on the age of the learner(s) and personal needs (Joy, 2021; Irawan, Wilson & Sutrisno, 2020; Savanni, 2019; Todorov, 2021).

Duolingo is the best-case scenario for gamification implementation, since it borrows plenty features from gaming platforms. It includes short groups of word families, vocabulary and grammar, but no one can learn a language just by practicing some words and phrases through flashcards. The Stories option provides a reading-like activity with interesting topics, real-life situation, graded difficulty, and a new-alternative approach of cross-skill approach (reading and listening alternately). The user can learn multiple languages simultaneously. Nevertheless, it focuses more on English speakers and there is no common ground theory behind its course design (for several language courses, the basic vocabulary has strong issues about word selection and context and it is only for beginners and not meant for advanced learners). There isn’t any proper and suitable grammatical explanation and rules for sentence structure. Everyone knows the varieties can be productive but are insufficient on Duolingo. On the other hand, its playful and game-like interface renders Duolingo such a popular and worth-testing application.
4.2. Busuu

Busuu is a language learning app with an exponential growth of its popularity. This can be seen in the fact that more than 100 million people choose it to learn Spanish, pick up a few French phrases, try Japanese, or one of the other 12 different supported languages. Several studies (Citrayasa, 2019; Malerba, 2015; Nushi & Jenabzadeh, 2016; Rosell-Aguilar, 2018; Saona-Vallejos, 2018; Shibata, 2020; Winans, 2019;) have been conducted about its efficiency, usefulness and impact for almost all language skills. Busuu provides well-crafted learning tools such as flashcards, conversation models, correction by native speakers, and grammar extracts for its users, plus the community help center. Busuu offers exercises for the four major language skills through its concentrated and well-formed exercises. However, the huge downside of the free version is the small variety of exercises that it offers, which gets boring after a while (Nushi & Jenabzadeh, 2016). Shibata’s (2020, p. 202) review underlines that “most of the activities in Busuu, however, still seem to be based on the theoretical perspectives of behaviorism due to the noticeable number of audiolingual activities and repetitive mechanical drills” and suggests the need to add some opportunities to collaborate and discuss senses with other users to boost its effectiveness. On the other hand, Winans (2019) supports that interactivity and social connection with native speakers of the target language can be motivating, intriguing and challenging for users, since they provide real feedback about daily and realistic use of each language. Results of Malerba’s (2015) study revealed that learners with an autonomous attitude tended to have a more effective learning experience, to find interesting topics and to be able to combine social and pedagogical paths even if they are not properly supported. Additionally, most of the users have positive feedback on their counter the app users community where learners can share goals and be engaged in continuous collaborative and meaningful activities (Saona-Vallejos, 2018). Finally, it is impressive that the longer the learners have been using the Busuu, the more likely they are to pay for premium content (Rosell-Aguilar, 2018).

Busuu is one of the most recent entries in LLP domain; trying to overcome other apps cons, the tailored lessons and well-structured course are good perks. The monthly subscription provides access to advanced grammar lessons and tailored content, and the subscribers can even start...
speaking to other native speakers (sometimes their feedback is quite vague and problematic). However, the relatively small language selection, the mobile app (compared to web app), the oversimple interface and some tedious typing exercises are significant issues. Additionally, although Busuu’s AI assigns a strength level to words and phrases, showing how well you understand them, it seems that for some languages (e.g., Greek, Spanish) AI loops the same material repeatedly. Finally, by speaking or writing in a new language, the user can get feedback on how to improve from native speakers.

Image 2: Samples from Busuu showing a vocab revision and a typical tip during a new lesson session.

### 4.3. Memrise

Memrise is a language learning app founded in 2010. It has quickly grown and now it numbers more than 50 million people and contains rich, real-life language content. Memrise’s courses have real-life language. More importantly in terms of richness, they are packed with thousands of video clips of native speakers speaking in their native language. Several studies (Cedercreutz & Ropicki, 2018; Karjo & Andreani, 2018; Nuralisah & Kareviati, 2020; Reknasari & Putro, 2019; Taebenu & Katemba, 2021; Widyaningrum & Putro, 2019; Zhang, 2019) have been carried out to date. Nuralisah and Kareviati (2020) maintain that vocabulary learning using Memrise is effective for improving students’ language skills. The advantage of using this application is to motivate students to learn English in a fun way and make it easier for students to memorize words. This can also increase the sense of cooperation with each other. Additionally, Zhang (2019, p. 161) concludes that “although adopting the audiolingual method in nature, Memrise has the strength of promoting learners’ memorization of words and phrases. However, the lack of contextual elements, of oral production, and of the capability to help students carry out authentic interactive conversation in real-life scenarios limits Memrise’s ability to function as the sole source for second language learning.” According to Reknasari and Putro (2019), Memrise offers comprehensive vocabulary learning through reviewing in order to reinforce the students’ word retention. Finally, Karjo and Andreani (2018) believe that reviewing previously learned words before learning new words, becomes annoying and frustrating very soon for learners.

Memrise is a worthwhile app for learning languages, particularly for beginners who want to learn new characters and basic vocabulary. Building up beyond typical LLP features, it offers a really strong language study content for beginners, the vocabulary is merely
based on flash cards (quite heavy on the word knowledge acquisition) and it implements a functional customization. Nevertheless, the absence of uniformity is notable since the user-created content has no guarantee on quality. It is actually a vocabulary-centered and focused app with awkward, bothersome and poorly designed interface. Moreover, it is not meant for advanced level learners and there is a significant lack of access to extra courses in-app.

![Image 3: Samples from Memrise showing the overall overview of language progress and a vocabulary task.](image)

### 4.4. LingQ

LingQ was launched in late 2007 by Steve and Mark to develop better tools for language learners. The basis of the LingQ system is to enable thousands of learners from all over the world to enjoy language learning and successfully reach their goals. LingQ is a fun, simple way to learn languages from an authentic and interesting content; its main goal remains to help users by learning the language from the language itself, just as most children learn their own language. Along with over 1,000,000 LingQ member, LingQ offers 42 language courses. Winda (2019) notes that LingQ provides some listening activities to enhance more vocabulary acquisition instead of improving the skill itself. Despite the active community and the 14-years of existence, there is a significant lack of publications and reviews about LingQ.

LingQ's build-your-own learning program may appeal to people who are looking to add something new to their current language studies. The service is not suitable for anyone learning a language from the very beginning. LingQ has a structured learning path that is guaranteed to ensure that every session will be productive. Aside from just plainly answering questions from the LingQ courses, you will be challenged to master new words and focus on those that truly matter to you. The use of the vocabulary building tool, like multiple choice, cloze, flashcards, and dictation is perfect, especially if the user’s goal is to increase how many new words need to be learned. However, the overall learning path isn't well structured and there is no guarantee of proper tutoring. Additionally, user interface is cluttered and distracting, and unfortunately, this app is not recommended for learning a new language.
4.5. Rosetta stone

Rosetta Stone’s Dynamic Immersion method uses interactive and contextual language lessons blended with Extended Learning features. With a variety of speaking-focused lessons and features, instantaneous pronunciation feedback with TruAccent, a track record of getting people speaking any language confidently and an almost five-star rating in the App Store, the award-winning Rosetta Stone mobile app (PCMag Editors’ Choice 2019, Tabby Awards Winner 2019, Best Mobile App Awards: Best Designed App and Best Overall App 2019) is an excellent way to learn and speak new languages with subscription. Lafford, Lafford and Sykes (2007) evaluate Rosetta Stone among other LLPs, based on a features group of a previous SLA research that has been proved quite significant and relevant in the acquisition process, such as the relevant contextualization of language, opportunities for interaction, etc. Santos’ (2011) review assesses the Rosetta Stone app by noting its fairly significant lack of context in the materials and an inability to respond to spontaneous student speech. Santos sums up that Rosetta Stone’s interaction is a quite poor and downgrade version of real-life conversations and speech encounters. Moreover, DeWaard (2013) explores the use of Rosetta Stone as a replacement of classroom instruction and points out that the interface of this app is appealing but lacking in several areas. More precisely, DeWaard (2013) stresses out the failure of cultural authenticity and the major limitations of a flawed machine learning system. DeWaard (2013, p. 61) concludes that Rosetta Stone is “not a viable replacement of current instruction at the postsecondary level”. Finally, a major study by Lord (2015; 2016) presents data that confirm Rosetta Stone’s marketing claims of being superior to in-class learning, in spite of learners’ appreciation for the flexibility and usability of the program. In fact, the learner outcomes indicate that this app does not seem adept at helping learners develop crucial communicative strategies in the foreign language, although it is capable of presenting isolated, decontextualized language elements.

Rosetta Stone used to claim that you could learn a foreign language the same way a child does if you just bought their program. The company faced a lot of well-deserved criticism for this claim. The truly innovative and unique approach without explicit grammar is research-based; TruAccent™ voice recognition, live tutoring lessons and well-done audio stories are the strong points of this learning app. It can be a decent and sufficient tutor for non-beginners. However, this common and unchangeable approach to very different languages, the tedious repetition and several culturally irrelevant photos may actually ruin any positive effort by this app. Furthermore, the absence of explanations is a crucial issue, the multi-choice questions don’t emulate immersion, and never advances past a beginner’s comprehension.
5. Research context and methodology

5.1. Participants

In order to cross evaluate these five apps with the highest user’s score in Google Play and Apple Store, ten (10) participants have agreed to test each app for 4 to 6 hours and then give their opinion about their experience with these applications. They are native Greek speakers, and they have a proficiency in English (C1-C2 levels). The participants are 32-44 years old, 6 men and 4 women, and they come from a variety of professional backgrounds. Some participants have already used a LLP before, and they all tested the same language course in all the apps, which was either German or Italian. The participants have been asked to complete a post-questionnaire based on the aforementioned criteria (see 3.2).

5.2. Research methodology

This case research included an online survey via Google Forms that was developed based on other studies in the use of apps for language learning (Joy, 2021; Karjo & Andreani, 2018; Nuralisah & Kareviati, 2020; Psychogyiou & Karasimos, 2019; Padilha et al., 2020; Rosell-Aguilar, 2017; 2018; Saona-Vallejos, 2018; Todorov, 2021), but tailored to the specific features of these five LLPs. The survey consists of 15 items, most of which were multiple choice questions (more than one option available) with two open follow-up questions asking participants to write some advantages or disadvantages of these apps. Moreover, there was some short open-ended interviews (via Google Meet) with each participant to provide further detailed feedback and to justify their final winner language learning app. The responses to open questions were categorised following the thematic analysis process suggested by Braun and Clarke (Brown & Clarke, 2006; Mikros, 2011).
6. The effective of learning a language: the comparison mode

6.1. Data analysis and discussion

Most of the participants (70%) have already used a language learning app to acquaint themselves with at least one language; a further 20% were using an app to learn or brush up on three languages usually during a travel preparation (60%) or for business purposes (50%). Only a minority of the participants (20%) used an LLP that was not in the top-10 of users’ choices. However, they reconsidered their options based on a testing period of the five selected learning apps. All of them think that a daily use for one week is a sufficient time to review an app and decide whether to keep up with it or drop it.

![Graph 1. Language learning apps and language skill satisfaction](image)

Based on their evaluation choices for the language skills and linguistic items, Rosetta stone received the highest (mean) score of these apps (69%) with a very short distance (65%) from Duolingo. For vocabulary acquisition, Duolingo and Busuu provide a greater variety of new words (80%) by building a more wide-category group of words compared to other apps. Memrise was described as the most sufficient app (80%) to cover grammatical phenomena, to teach the rules and to provide some straightforward guidelines of morphosyntactic features.

The participants still seem to be more active in the learning process; Nevertheless, some apps like Busuu and LingQ have a course design to engage and keep more active during the learning sessions. Moreover, the overall view of language skills points out the knowledge of what was taught by the Rosetta stone is closer to a teacher’s effort. Participants are initially taught with basic grammar, typical phrases and sentences and they fill in the answers from the questions given. They are very interactive, responsive, and like the media provided. Mobile assisted language learning apps such as Duolingo, Memrise and Rosetta stone indeed help users who want to learn a new language. However, learning a foreign language needs more than just memorizing and translating words or phrases. Learning a language means learning to master the four skills of language, i.e. speaking, reading, listening and writing; the above chart represents a fuzzy tense towards Rosetta Stone dominance, but the participants’ choices may differ based on the desired target skill. Yet, these skills
are taught incompletely, since the learning materials are provided in-game based activities without any further explanation. An essential advantage of mobile learning is that it is collaborative and thus encouraging for learners due to the possibility of exchanging knowledge, skills, creating communities, tailored-lessons, and experience through interaction (Czerska-Andrzejewska, 2016). The portability of the mobile device does always bring positive effect, i.e. to encourage learning or revise a specific course.

**Graph 2. Language learning apps and design evaluation**

### 6.2. Limitations and further research

Since no previous research into language learning from a group of users testing the same apps exists, it is not possible to comment on the return rate and its comparability to previous studies. The data collected for this study is self-reported, and therefore subject to the limitations that such research methods have (Paulhus & Vazire, 2007). The questionnaire did not manage to capture their detailed likes and dislikes, although the interviews collected data about features that dissuaded them from participating, and how these could have been improved. Additionally, it would be ideal to expand the number of testing learning apps (the top-10 of ratings) and the participants, so we can actually have a more holistic and clearer picture of this cross-platform overview of language learning.

### 7. Concluding remarks

The purpose of this article was to explore the effectiveness of five popular language learning application in learning a foreign language and developing language skills. Most of the participants prefer distance learning for professional and educational purposes. They stated some of the advantages and disadvantages of using these language learning apps; they claimed that distance learning is an easy, free and practical learning process, and therefore, they preferred Duolingo and Rosetta Stone as the best option. It is important that the learners had the opportunity to adjust the process according to their learning curve.

The vast majority of participants were confident about the result. However, when it comes to the development of language skills, there are some contrasting views which question the effectiveness of these applications. On one hand, the users were pleased
with learning new vocabulary and pronunciation with gamification. On the other hand, some users suggested that the way grammar is “taught” on most applications needs improvement or some apps overfocused on vocabulary. Cross-evaluation with more and new LLP apps is needed and comparative evaluation with other language learning courses, since not all the courses are so extensive. It is important that there are applications such as Duolingo, Memrise, Rosetta stone and other language apps to support learners in the learning process. Evaluating and constantly developing new ways of learning will only have positive results having in mind the learners’ needs.

References


Holmberg, B. (1986). *Growth and Structure of Distance Education.* London: Croom Helm.


Rosell-Aguilar, F. (2009). In the decade or so since smartphones were introduced, these and subsequent. Στο L. Lomicka, & G. Lord (Επιμ.), The next generation: Social networking and online collaboration in foreign language learning (σσ. 13-34). San Marcos, TX: Calico.


**Athanasios Karasimos** (karasimos.athanasios@ac.eap.gr) is a graduate of the Department of Philology, University of Patras. He holds two European Masters in Speech and Language Processing (one of them at the University of Edinburgh) and his doctoral dissertation is in Computational Morphology. He participated in several research projects on Modern Greek dialects, corpora, aphasic speech, digital humanities, and the training of English language teachers. He was a postdoctoral research fellow funded by IKY. He worked as an Adjunct Lecturer at HOU, AUTH, and NKUA teaching Educational Technology, Research Methodology, Computational Linguistics, and Corpus Linguistics. He is a researcher in the Center for the Greek Language and in the national infrastructure for Digital Humanities DARIAH-GR / DYAS (Academy of Athens). His research interests focus on computational linguistics and machine learning, the use of corpora, education technology, and integrating video and board games into language teaching and learning.
Using poetry to foster Critical Thinking and Metacognition in a Primary School EFL context

Aspasia DELILIGKA & Christine CALFOGLOU

Critical thinking and metacognition are two concepts that have played an increasingly important role in education and have been the focus of a substantial body of research related to the skill of Reading (Mohseni, Seifoori & Ahangari, 2020; Navarro, 2021). Poetry, on the other hand, though much discussed and analyzed as a genre, has not been adequately explored as a tool used to foster the development of critical thinking skills and metacognitive reading strategies among Young Learners. The present article attempts to show that the teaching of poetry can have a positive effect not only on Young Learners’ critical thinking and use of metacognitive reading strategies, but on their reading comprehension as well. This is demonstrated through the presentation of a poetry project implemented with 16 students, another 16 being assigned to a control group, in a primary education EFL context, over a three-month period and through analyzing learners’ performance and attitudes prior to and after implementation. The results reveal that there is indeed a statistically confirmed advantage in the experimental group learners’ critical thinking, use of metacognitive reading strategies, as well as reading comprehension.

Key Words: Critical Thinking, Metacognitive reading strategies, Reading Comprehension, Poetry, Young Learners

1. Introduction

Critical Thinking (CT) is a process that involves reflective and purposeful judgment when dealing with a problem and increases the chances of reaching a logical conclusion (Dwyer et al., 2014). The importance of its instruction has been underscored by various researchers and education councils (Association of American Colleges and Universities, 2005; Australian Council for Educational Research, 2002; Butler et al., 2012; Halpern, 2003; Ku, 2009), as it allows individuals to cope with complex information and become more adept at problem-solving and decision-making in the real world. CT is also described as a metacognitive
process (Dwyer et al., 2014; Magno, 2010), as it heavily relies on self-regulation, which is also an essential component of Metacognition (MC), the process of thinking about thinking (Flavell, 1979).

Poetry is a literary genre that encourages self-exploration and self-expression but has often been considered too difficult for young EFL readers (Nichols et al., 2018). A considerable amount of research, however, has shown that teaching it can be beneficial for young EFL learners’ reading skills (Khansir, 2012; Norris, 2011; Rasinski et al., 2015) because poetry is enjoyable, easily accessible, memorized and recited. Furthermore, it is non-threatening, due to its usually limited length and the fact that it allows for multiple interpretations (Calfoglou, 2004/2019).

The present article focuses on an area of research that has received very little attention so far, namely, whether the two interconnected concepts of CT and MC can be promoted through the teaching of poetry within a Primary School EFL context and whether this development can have an impact on learners’ reading comprehension. The article is laid out as follows: Section 2 deals with the theoretical context of the present discussion in relation to the concepts of CT, MC and poetry. Section 3 presents the methodology employed in the research and the intervention designed and section 4 discusses the data obtained. Section 5 offers some concluding remarks.

2. The concepts of Critical Thinking, Metacognition and Poetry in relation to Young Learners’ Reading

2.1. Critical Thinking and Metacognition

CT is a rather elusive concept, which has received multiple interpretations (Cottrell, 2017; Davies, 2015; Facione, 2013). It has been generically defined as the objective analysis and evaluation of an issue in order to form a judgment and its origins can be traced in the teaching of Socrates, who established the importance of asking questions that aim to explore meaning beyond appearances (Atabaki et al., 2015). In the 20th century, the idea of CT can be found in the writings of John Dewey, Jean Piaget, Lev Vygotsky, Paulo Freire, Peter Facione, Richard Paul. The 1980s witnessed a revival of interest in inquiry and CT in the United States. Fundamental questions regarding the definition of CT and its component skills, the most effective teaching methodology and how it could be assessed called for an answer.

The theoretical underpinnings of CT can be traced in not just one but three distinct academic disciplines: philosophy, psychology and education (Almeida & Franco, 2011). The philosophical approach focuses on the qualities of an ideal critical thinker and standards of thought that relate to the application of rules of logic. More specifically, Paul (1992, p. 9) refers to “perfections of thought” and Facione (2013) describes the ideal critical thinker as someone who possesses open-mindedness, flexibility, willingness to suspend judgment and to consider various points of view. The psychological approach examines the types of actions critical thinkers perform and defines CT as a list of skills. Halpern (1998, p. 450), for example, defines CT as “the use of those cognitive skills or strategies that increase the probability of a desirable outcome” and Willingham (2007, p. 8) as “seeing both sides of an issue, being open to new evidence that disconfirms your ideas, reasoning dispassionately, demanding that claims be backed by evidence, deducing and inferring conclusions from available facts, solving problems, and so forth”. Finally, the educational approach has heavily relied on the
work of Benjamin Bloom and his associates (1956). His taxonomy of cognitive skills - with knowledge, comprehension, application at the bottom and analysis, synthesis, evaluation at the top (cf. Anderson et al., 2001) - has long provided educators with a valuable framework for teaching higher order thinking skills, synonymous for many with CT (Kennedy et al., 1991).

Therefore, it seems that providing a concrete definition of CT for the purposes of the present article cannot rely on one single approach but should rather exploit elements from all three previously discussed perspectives. One of the definitions that endorse a holistic view of CT has been provided by Matthew Lipman, a philosopher and educator\(^1\), who argues that “critical thinking is skillful, responsible thinking that facilitates good judgment because it (1) relies upon criteria, (2) is self-correcting and (3) is sensitive to context” (1988, p.39). Lipman explains that any claim or opinion is vulnerable unless it is supported by reasons, namely criteria of a reliable kind, which are relevant to specific domains. They can form the basis of a comparison and can be formal or informal. CT is also self-correcting in the sense that it tries to identify its own weaknesses and repair faulty procedures. One final characteristic of CT is that it adapts to specific contexts and takes various parameters into account. It is also assumed in the present work that, in accordance with the psychological approach, CT is clearly evident in certain skills or abilities. Researchers from all three schools discussed earlier usually agree on the following abilities (Lai, 2011), which have also been selected as the focus of this research:

- analyzing evidence
- making inferences
- evaluating
- making decisions

From an educational perspective, these skills seem to correspond to Bloom’s higher order thinking skills of analysis, synthesis and evaluation.

Another form of reflective thinking can be traced within the realm of MC, which has become popular in educational and cognitive development theories within the last fifty years. Perhaps the most well-known definition of MC is “thinking about thinking” (Flavell, 1979, p. 906). Etymologically speaking, the prefix meta- implies a higher or second level of cognition, knowledge about knowledge. John Flavell\(^2\) first used the term in 1976, when he referred to MC as “the active monitoring and consequent regulation” of information processing activities, “usually in service of some concrete goal or objective” (p. 232). He, therefore, implied that MC is both intentional and conscious thinking that can either precede or follow any cognitive activity.

Metacognitive strategies are used to monitor cognitive progress. Their main aim is to ensure that a cognitive goal (for example, understanding reading material) has been achieved. Possessing good metacognitive skills and awareness means that a person can successfully oversee one’s own learning process, plan and monitor ongoing cognitive activities and compare cognitive outcomes against internal or external criteria (Cartwright et al., 2017; Flavell et al., 2002; Hargrove & Nietfield, 2015).

---

\(^1\)Lipman is considered to be the founder of Philosophy for Children, an educational movement that proposes enhancing children’s learning through inquiry and the exploration of ideas.

\(^2\)Flavell was an American developmental psychologist who was influenced by Piaget.
There seems to be a close connection between MC and CT, although researchers fail to agree which concept includes or precedes the other. For example, Flavell (1979) and Martinez (2006) agree that CT is inherent in MC. Flavell argues that “critical appraisal of message source, quality of appeal, and probable consequences needed to cope with these inputs sensibly” can lead to “wise and thoughtful life decisions” (p. 910), while Martinez considers CT to be one of three types of MC, the other two being metamemory and problem solving. Kuhn (2000) maintains that CT is equivalent to MC. Van Gelder (2005) and Willingham (2007), on the other hand, view MC as a component of CT, when they describe the ability to use appropriate strategies at the right time as an important CT skill. A critical thinker is, therefore, someone who uses metacognitive strategies to control one’s thinking processes. Halpern (1998) presents MC and CT together in a model. She defines MC as the ability to use knowledge to direct and improve thinking skills. It is, therefore, safe to say that the two concepts seem to complement or support one another, as they both aim towards a higher quality form of thinking.

As the main focus of the present article is the reading skill, it is also necessary to discuss the concept of Critical Reading (CR), which, of course, presupposes the application of CT. Freire (1983, p. 5) sees “the actual act of reading literary texts...as part of a wider process of human development and growth” and his claim that “reading the world precedes reading the word” underscores the importance of perceiving the intrinsic relationship between text and context. For Freire (ibid, p. 11), “reading always involves critical perception, interpretation and re-writing what is read”. CR in this light seems to be a frame of mind readers adopt in their attempt to explain the world and the word, rather than a set of strategies or skills. It differs from traditional reading in that it allows a wide range of interpretations rather than focusing on definite meanings (Wallace, 2001).

Several researchers have suggested that one of the most effective ways to achieve CR and ultimately enhanced reading comprehension is through the instruction of metacognitive reading strategies (Afflerbach et al., 2015; Duffy, 2002; Kamgar & Jadidi, 2016; Morshedian et al., 2016; Pressley, 2002; Tracey & Morrow, 2006). Pressley and Gaskins (2006) use the term ‘constructively responsive reading’ to refer to metacognitive reading and describe what good readers do before, during and after reading. In the first stage, good readers identify a purpose for reading and preview the text by noticing certain features, such as its title and length. During reading, they may skim or slow down, reread, take notes or make mental notes, update and relate ideas, constantly monitor and repair comprehension, draw inferences, construct mental images and evaluate. After reading, good readers think back to what they have read and revise it, while continuing to ask questions.

This description is similar to Urquhart & Weir’s (2016). According to them, planning corresponds to the pre-reading stage and heavily relies on prediction, monitoring is what readers do while reading by constantly asking themselves questions about the characters, the author’s message, their own level of understanding and, finally, evaluating is the post-reading stage, in which readers express their opinion about the text, the message and what helped them to understand.

The present article focuses on the development of the following metacognitive reading strategies within the stages of planning, monitoring and evaluating, which seem to have

---

3This term generally refers to awareness of one’s memory and of how it works.
increased relevance to aiding Young Learners (YL) towards becoming more proficient readers (Ahmadi et al., 2013; Chatzipanteli et al., 2014; Temur et al., 2010; Whitebread et al., 2010):

<table>
<thead>
<tr>
<th>PLANNING</th>
<th>MONITORING</th>
<th>EVALUATING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activating background knowledge</td>
<td>Self-questioning</td>
<td>Connecting ideas (cause-effect)</td>
</tr>
<tr>
<td>Predicting</td>
<td></td>
<td>Determining importance</td>
</tr>
</tbody>
</table>

Table 1. Metacognitive Reading Strategies selected

This categorization, however, is not a strict one, in the sense that some of these strategies can be used in more than one stage and are potentially omnipresent (Veenman, 2015).

In the case of YL specifically, it seems that, unlike earlier research, which considered children incapable of more complex cognitive processes, there is fertile ground for an early start, which will ensure better initiation to CT. Recent research suggests that there is indeed room for teaching CT in primary school, as younger children have been observed to think critically (Florea & Hurjui, 2015; Gelerstein et al., 2016; Heyman & Legare, 2005; Silva, 2008; Willingham, 2007). Similarly, recent work has demonstrated that younger students do exhibit metacognitive thinking abilities (Bares, 2011; Conklin, 2012; McLeod, 1997; Schneider, 2008; Schraw and Moshman, 1995; Whitebread et al., 2009) and can, therefore, receive and benefit from metacognitive strategy instruction.

2.2. The case for poetry

The present article argues that both CT and metacognitive instruction can be effected through the teaching of poetry, despite the fact that it has a reputation for being a rather inaccessible as well as deviant genre (Adamson, 2016; Khansir, 2012), which “departs from linguistic norms” (Lazar, 1993, p. 104). First of all, in relation to YL, there seem to be numerous advantages to using poetry. The content of poetry may offer young readers the opportunity to approach serious and difficult topics in a light-hearted way, as the length, rhythm and rhyme of a poem may encourage the reader to interact and play with it (Ara, 2009). Furthermore, the relative ease of mastering the reading of a short poem gives YL a sense of achievement and self-efficacy that is essential for ensuring further progress (Rasinski et al., 2015; Rasinski, 2017). Young readers’ affective engagement with poetry is also very important, as it facilitates learning (Hidi & Harackiewicz, 2000). According to Carter (2013, p. 4), “reading a poem should arouse feelings and opinions in a child and lead to the need to express and share them”. This need for expression, in turn, may promote linguistic competence and enhance comprehension (Creely, 2018).

Poetry can also be used to improve young EFL students’ fluency, word recognition and vocabulary (Vardell et al., 2006) and to promote phonemic awareness (the ability to identify and manipulate the sounds of language) and phonics (the ability to translate the written form of words into their oral representation) (Nichols et al., 2018). Unlike other literary forms, most poetry is meant to be recited in front of an audience. In order for a poem to be meaningful, multiple rehearsals need to take place, leading to fluency improvement (Rasinski, 2014). Moreover, the form of most poems is quite friendly to young readers, as there is ample blank space, while some even possess a particular shape, thus appealing to the senses and paving the way for better comprehension (Rasinski et al., 2016). Last but not least, young readers’ encounter with English poetry provides them with a unique
opportunity to explore cultural elements and ideologies that will not only facilitate comprehension but will also lead to intercultural awareness (Thaler, 2016).

In terms of CR, the form of most short poems encourages children to learn them by heart, just as they learn the lyrics of their favorite songs, which allows for the discovery of multiple meanings through repetition (Medwell et al., 2014). According to Hess (2003, p.20), “entering a literary text, under the guidance of appropriate teaching, brings about the kind of participation almost no other text can produce … when we read, understand, and interpret a poem we learn language through the expansion of our experience with a larger human reality”.

As regards CT and MC, the ambiguity of meaning encountered in most poems encourages reflection and discussion. Discussing the multiplicity of meanings and choosing one that fits personal preferences or that meets specific requirements imposed by a task is an exercise in problem solving, an essential component of CT (Dwyer et al., 2014; McPeck, 2016). Forming connections and relationships among parts of a poem, as well as among other poems on similar themes, seems to enhance learning and CT (Halpern, 1998; Fowler, 2019), along with metacognitive reading strategies (Eva-Wood, 2008). The latter, as we saw earlier, are activated when readers monitor the state of their understanding of a text. The challenge involved in making sense of a poem entails that readers will make the most of strategies such as activating their background knowledge and predicting (by using clues like pictures and titles), asking themselves appropriate questions (Blau, 2014) as they read and reread the poem, connecting ideas within a poem (like cause and effect) and determining their importance.

In addition, the use of unconventional language or the unusual treatment of a topic adds an element of surprise in the reading process and invites readers to engage in CT as they try to resolve feelings of imbalance (Halonen, 1995). Poetic language constitutes a major incentive for using inferencing and prediction, as readers usually need to resort to background knowledge and context in order to reach an interpretation of the poem (Cushing, 2018). Metaphors, for instance, demand inferential thinking and critical analysis, which, in turn, play a crucial role in processing experiences in a complex multicultural society (Hankin, 2017; Pistol, 2018). Being critical thinkers in such a society entails the ability to see and experience different perspectives on various issues and this is what reading poetry generously offers, without neglecting the issue of aesthetic appeal (Powell, 2021). It is this connection to CT, MC and reading comprehension that will be explored in some detail in the present work.

3. The intervention

3.1 The method

The research questions addressed in the present work are founded on the hypothesis that poetry teaching can have beneficial effects on the development of young learners’ CT, metacognitive reading skills and L2 reading skills generally and are as follows:

1. Can the teaching of poetry enhance Young Learners’ Critical Thinking and, more specifically, analyzing evidence, making inferences, evaluating and making decisions?
2. Can the teaching of poetry enhance Young Learners’ use of metacognitive reading strategies, and, more specifically, activating background knowledge, predicting, self-questioning, connecting cause with effect and determining importance?

3. Does the development of Critical Thinking and metacognitive reading strategies through poetry have an overall positive effect on Young Learners’ reading comprehension?

In order to answer the questions above, the present study pursues action research, whereby the researcher exposes experimental group learners to the new programme and compares their performance with that of a control group, whose members receive standard instruction (Mertler, 2016). Actually, both groups were given instruction on the use of the specific strategies targeted but, unlike the experimental group, which practised these strategies within a poetry context, the control group learners dealt with ordinary coursebook reading input.

The specific study followed a mixed methods approach, as it employed both quantitative and qualitative methods of data collection so as to triangulate results and achieve greater validity (Dornyei, 2007). Quantitative research tools involved a needs analysis questionnaire, a pre- and post-intervention reading comprehension test, a pre- and post-intervention CR test and a pre- and post-intervention metacognitive reading strategies questionnaire. The Statistical Package for Social Sciences (SPSS 21.0) was used for the data analysis. More specifically, regarding all three research questions, paired samples t-tests were used to determine whether there were statistically significant differences, between performance at the beginning and at the end of the intervention, within the control and the experimental group. Statistical significance was set at the 0.05 level. Qualitative data collection, on the other hand, involved a teacher's observation checklist, which was completed at three different points during the intervention.

The subjects of this research were 32 students divided into two groups of 16 in the sixth grade of a primary school in New Artaki, Evia. They had been learning English at school for 3 years, with three 45-minute lessons per week. All students were monolingual, except one, who had recently come from India. Their competence level ranged from A1 to B1 (Council of Europe, CEFR, 2001, 2018).

A pre-instruction reading test was administered to both groups to determine reading ability levels and make a comparison with post-reading test performance possible. The intervention spanned a period of three months.

3.2 The process

The intervention involved three stages: the pre-instruction, the instruction and the post-instruction stage. In the pre-instruction stage, the experimental group students answered a needs analysis questionnaire designed to explore their reasons for learning English, their familiarity with poetry and their preferences in relation to poetry themes. The results obtained from the questionnaire provided the teacher with insight into what kind of themes and tasks would be more suitable for the specific learners. This was used as a compass in the choice of poems and was meant to make the poetry input provided more appealing. In addition, both groups had to answer a metacognitive reading questionnaire (Appendix A), which focused on the use of the metacognitive strategies selected. Finally, they all took a CR test (see Appendix B for sample tasks), designed to explore the use of the CT skills selected.

---

4 The reading section of a sample Cambridge A2 Flyers exam.
The instruction stage consisted of two distinct phases. In the first phase, both the control and the experimental group received four 90-minute lessons of explicit strategy instruction, designed in accordance with the Gradual Release of Responsibility model, which was originally proposed in Pearson and Gallagher (1983). In the first lesson, that of ‘modeled instruction’, the teacher explained the strategies while thinking aloud and the students listened without participating. The lesson started with an introduction to the skills and strategies selected. The teacher used the think-aloud technique to show learners how to use the specific strategies, while learners were provided with the text (an article) the teacher read without the commentary so that they could focus on listening and completing a checklist. The checklist contained a few statements that referred to the strategies targeted and students had to decide whether the teacher actually used the specific strategies or not. For instance, the first item on the checklist asked students whether the teacher tried to infer content from the title of the text.

The second lesson, the one of ‘shared instruction’, was designed to encourage the learners’ participation in the teaching process. It therefore involved a much shorter and simpler text (a letter of complaint). The tasks were also less complex than those in other lessons and group work was specifically required. The lesson opened with a task aimed at reminding students of the strategies being taught in the previous lesson and went on with tasks that required students to practice the specific strategies. For instance, the students were given the opening of the text (‘Dear…’) and were encouraged to predict its content.

In the third lesson, that of ‘guided practice’, students practiced the strategies collaboratively and received guidance. Two texts (film reviews) on the same topic but of different levels of difficulty were employed. Providing students with multiple texts intended to facilitate comparison and connection of ideas. Learners were also provided with ‘a manual’ that meant to familiarize them with the idea of working on their own. In essence, this was a task sheet which provided very explicit instructions on how to think when approaching a text and guided learners through the process step by step so as not to intimidate them. For instance, in the planning stage learners were encouraged to focus on the titles and pictures accompanying the texts so as to predict the content. In the monitoring stage, the tasks required multiple readings of the texts before learners decided on differences and similarities between them. In the evaluating stage, learners were asked to find evidence that supported their evaluation of the writers’ treatment of the subject.

There was also a self-evaluating component introduced in the final stage, which was included in all the lessons that followed. It aimed to encourage learners’ self-awareness in terms of using the strategies selected and to improve future performance. In the final lesson, that of independent practice, students applied the strategies independently and received feedback at the end. A biographical text was used and the tasksheet followed a pattern like the one used in lesson 3. The level of difficulty was intentionally lower to encourage students’ independent practice at this point. There was also a feedback component at the end, identical with that used in the previous lesson, aiming to help learners realize which strategies were used in each task and whether their original answers should be revised. All four lessons relied on group work and teacher, as well as peer, scaffolding. The materials involved authentic, non-literary texts in the belief that, once

---

5 Lipman’s (1988) definition of CT (see section 2.1) entails the use of comparison and connection based on specific criteria in order to support any claim or opinion.
learned, these skills can be used across contexts and with various genres (see Lipman’s (1988) definition of CT, section 2).

In the second phase, the experimental group received 8 lessons of poetry-based strategy instruction (see Appendix C for sample tasks), whereas the control group continued with standard coursebook instruction and were encouraged to practice the reading strategies taught when approaching course book texts. The poems (Appendix D) in the experimental group lessons of the second phase were organized in pairs of thematic affinity and in accordance with the students’ preferences, revealed through the needs analysis questionnaire. They were selected following a set of general criteria (adapted from Hadaway & Young, 2010), namely topic familiarity, language level, textual support, and level of cultural fit. More specifically, the learners’ own choice of themes (as recorded in the needs analysis questionnaire) formed the basis for selecting sixteen poems, divided into eight thematically related pairs, to ensure that there was a certain degree of topic familiarity. The language level of the poems was also carefully considered, and students were provided with glossaries whenever this was thought necessary. An effort was made not to inundate students with too many abstract terms and new vocabulary was introduced in meaningful, contextualized ways. For instance, most of the poems were accompanied by pictures and there was extensive use of inferencing and activation of the students’ background knowledge in the planning stage. Textual support was yet another criterion considered and most of the poems selected were either accompanied by visuals or used wording which allowed students to infer the meaning of new vocabulary. Moreover, in relation to the level of cultural fit, the poems presented aspects of other cultures that were already quite familiar to the students.

At the post-instruction stage, all the learners had a reading comprehension test administered to them. This test was similar to the diagnostic reading test in the pre-instruction stage and aimed to record any significant improvement in the learners’ overall reading performance. They also responded to the metacognitive reading questionnaire (also used at the pre-instruction stage) and took a different version of the CR test, both of which aimed to reveal the extent to which students had benefited from strategy instruction and whether the poetry lessons had resulted in further improvement for the experimental group.

The teacher’s observation checklist was also completed for the experimental group at three different points during the study: at the beginning of Phase 2, halfway through Phase 2 and at the end of Phase 2.

4. The data

4.1 Poetry and CT

The project targeted the development of specific CT strategies, namely analyzing evidence, making inferences, evaluating, and making decisions, and had hypothesized a CT advantage for the poetry instruction learner group. This hypothesis seems to have been borne out by the data collected through the pre- and post-intervention CR test and analyzed using paired

---

6 Different versions of Reading and CR tests were selected for the pre- and the post-intervention stage because presenting learners with the same text input might compromise results in a number of ways.
sample t-tests. Thus, as can be seen in Table 2 below, the experimental group improved substantially between the pre- and the post-instruction stage: t (15) = -5.239, p = .000 (M=65.31, SD=32.89 at the pre-stage vs. M=83.75, SD=22.69 at the post-stage), while the control group showed no statistically significant differences between the pre- and the post-intervention performance: t (15)= -1.856, p= .083 (M=64.06, SD=29.62 vs. M=65.44, SD=28.16):

<table>
<thead>
<tr>
<th>Critical Thinking</th>
<th>Stage</th>
<th>M(%)</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>pre-intervention</td>
<td>64.06</td>
<td>29.62</td>
<td>-1.856</td>
<td>.083</td>
</tr>
<tr>
<td></td>
<td>post-intervention</td>
<td>65.44</td>
<td>28.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental group</td>
<td>pre-intervention</td>
<td>65.31</td>
<td>32.89</td>
<td>-5.239</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>post-intervention</td>
<td>83.75</td>
<td>22.69</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Pre- and post-intervention differences in Critical Thinking within each group (paired samples t-test)

As far as the qualitative data are concerned, a comparison between the various stages of the observation suggests that the experimental group made considerable progress from the beginning to the end of Phase 2. For instance, in Lesson 1, it was observed that students rarely made any inferences related to the poet’s intended message or evaluated poems in relation to specific criteria. Some students said that they could see no point in trying to infer the poet’s message beforehand. As instruction carried on, however, there was significant improvement in the frequency of use of these skills. In Lesson 4, students were observed to sometimes be able to infer the poet’s message and to evaluate the poems. They spent considerable time discussing these decisions in the planning and the evaluating stage. Finally, in Lesson 8, on completion of the intervention, most students were often able to perform the specific skills. For example, on receiving the worksheet, they would immediately look for clues that might help them reach certain decisions concerning the poems.

It might therefore be safe to argue that the control group students, who had received only explicit instruction in the CT skills selected, using non-literary texts, and who had continued with standard coursebook instruction, did not benefit as much as the experimental group, who had received poetry-based in addition to explicit instruction. It seems that the poems helped them to achieve the kind of thinking that Lipman (1988) defined as skillful, self-correcting, and sensitive to context (see section 2). Interaction with poetry can, therefore, be said to promote critical perception and interpretation of what is read, which, according to Freire (1983), as we saw in section 2, is the essence of CR. Learners were expected to re-write what the poet had originally written and were encouraged to approach meaning from any angle they found most suitable. This freedom may have contributed to their reaching a higher level of CT than the control group students, who were exposed to non-literary texts in Phase 1. By the end of the poetry lessons, some students remarked that they had never thought poetry would be so engaging or that it would help them reflect on their own lives: “I used to think poetry is boring but now I think it’s different”, “I never thought poetry is so interesting”, “After these lessons I try to write my own poems in English”.

These findings also seem to support the idea that poetry possesses certain characteristics which make it ideal for the teaching of CT. Poetic meaning is ambiguous and encourages reflection and problem-solving, essential to CT (Dwyer et al., 2014; McPeck, 2016). Reading
Poetry encourages the formation of connections not only between parts of a poem but also with other poems on similar themes, that is, hypertextually (Halpern, 1998). Readers may also employ as many strategies as possible to meet the challenge of making meaning out of a poem and heavily rely on inferencing (Hankin, 2017; Pistol, 2018). For instance, the learners involved in this study had to guess the meaning of unknown words based on context in all the poetry lessons. Apart from inferencing, the students also relied on the analysis of evidence, such as titles, pictures and form, in order to predict the meaning of the poems. For example, they predicted the meaning of the ‘little tree’ poem (Lesson 5, Phase 2), just by reading the title. In addition, they constantly engaged in evaluation and decision-making processes, when they were asked to decide whether they liked a poem or not and to justify their choices. For instance, the poems in lesson 6 (‘This is Just to Say’ by W.C. Williams and ‘The Little Boy and Old Man’ by Shel Silverstein) initiated a fervent discussion leading to the conclusion that family relations should be like those portrayed in the poems (open, sincere and sympathetic).

Finally, the results also suggest that the mixed approach to the instruction of CT skills, where teachers combine explicit, stand-alone CT instruction with applying these skills in a specific subject (poetry in this case) and which has been advocated by various researchers (Behar-Horenstein & Niu, 2011; Edwards, 2017; Ennis, 2013, 2018; Facione, 2013), might be more effective than others. Moreover, as the participants in the specific context had no previous engagement with CT, it seems that explicit instruction coupled with poetry-specific instruction had maximum impact on them.

4.2 Poetry and MC

The intervention targeted the development of specific metacognitive reading strategies, namely activating background knowledge, predicting, self-questioning, connecting cause with effect and determining importance and the quantitative data (Table 3) collected through the pre- and the post-intervention metacognitive reading strategy questionnaire show that there were no statistically significant differences in metacognitive reading strategy use between the control group’s pre-intervention and post-intervention performance: \( t(15)=1.775, p=.096 \) (M=49.36, SD=12.58 at the pre-stage vs. M=48.75, SD=12.33 at the post-stage). By contrast, there was a significant difference in the scores for metacognitive reading strategy use between the pre- and the post-intervention stage for the experimental group: \( t(15)=-7.383, p=.000 \) (M=46.75, SD=13.10 at the pre-stage vs. M=61.37, SD=10.77 at the post-stage). In other words, experimental group learners achieved a significantly higher score after the intervention:

<table>
<thead>
<tr>
<th>Metacognitive Reading Strategy Use</th>
<th>Stage</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>pre-intervention</td>
<td>49.36</td>
<td>12.58</td>
<td>1.775</td>
<td>.096</td>
</tr>
<tr>
<td></td>
<td>post-intervention</td>
<td>48.75</td>
<td>12.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental group</td>
<td>pre-intervention</td>
<td>46.75</td>
<td>13.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>post-intervention</td>
<td>61.37</td>
<td>10.77</td>
<td>-7.383</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table 3. Pre- and post-intervention differences in Metacognitive Reading Strategy Use within each group (paired samples t-test)
As regards the observation checklist, the findings obtained reveal that there is indeed a connection between exposure to poetry and improved metacognitive reading strategy use. For instance, in Lesson 1, it was observed that students rarely had pictures in their mind before reading a poem or reread it to clarify certain points, whereas by the time they had reached Lesson 8, they were often able to create mental images before reading a poem and they always reread it to make things clearer. In addition, in Lesson 1, it was observed that they could sometimes rank poems according to how important they believed they were in relation to specific criteria and could justify their choices, while by the end of Lesson 8, they were often able to do so.

These results indicate that there may be a connection between systematic exposure to poetry and enhanced metacognitive reading strategy use, as the students in the experimental group, who had received poetry-based instruction in specific metacognitive reading strategies on top of explicit instruction using non-poetic texts, seem to have benefitted more than the control group students, who had not been exposed to poetry. Such a conclusion is consistent with other relevant research into the effects of teaching poetry on MC. For instance, Threlfall (2013) considered how poetry can be used to aid adult students towards a process of self-guided reflection. Cloonan et al. (2011) showed that students can increase their metacognitive awareness through the teaching of poetry. Giovanelli (2017) used poetry reading to develop students’ MC. Simecek and Rumbold (2016) suggested that poetry response journals can lead to students’ enhanced engagement with subject matter as well as self-monitoring. The present study could therefore be added to those above while also establishing a beneficial effect for YL as well.

4.3 Poetry and Reading Comprehension

The study also examined the overall potential effect of the development of CT and metacognitive reading through poetry on Young Learners’ Reading Comprehension. For this purpose, students in both groups took a reading test at the pre-instruction and then again at the post-instruction stage and the data were subjected to statistical analysis. It was hypothesized that any improvement in reading comprehension would be generally reflected in the difference in students’ scores in the reading tests before and after the intervention. In support of this hypothesis, paired samples t-tests (Table 4) revealed that there was indeed significant improvement for the poetry group (t (15)= -5.233, p= .000, M=49.13, SD=29.19 at the pre-stage vs. M=62.25, SD=24.71 at the post-stage), whereas for the control group there was no significant change (t (15)=1.600, p= .130, M=51.06, SD=29.23 at the pre-stage vs. M=53.25, SD=27.92 at the post-stage). An eloquent example of the change brought about in the experimental group is that of a weak student who managed to climb from a score of 19% before the intervention to 47% after the intervention:

<table>
<thead>
<tr>
<th>Reading comprehension</th>
<th>Stage</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>pre-intervention</td>
<td>51.06</td>
<td>29.23</td>
<td>-1.600</td>
<td>.130</td>
</tr>
<tr>
<td></td>
<td>post-intervention</td>
<td>53.25</td>
<td>27.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental group</td>
<td>pre-intervention</td>
<td>49.13</td>
<td>29.19</td>
<td>-5.233</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>post-intervention</td>
<td>62.25</td>
<td>24.71</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 4. Pre- and post-intervention differences in Reading Comprehension within each group (paired samples t-test)*
These findings suggest that the development of specific CT skills and metacognitive reading strategies through poetry may indeed have an overall positive effect on students’ reading comprehension performance, since the experimental group students, who had shown the greatest improvement in CT skills and metacognitive reading strategy use (see Tables 2 and 3), also seemed to benefit the most in terms of reading comprehension. This conclusion agrees with the findings of earlier related research into the potential effects of teaching poetry on reading comprehension. Rasinski et al. (2016), for example, suggested that repeated practice in poetry reading might be very effective in improving struggling readers’ performance. Norris (2011) successfully used poetry instruction to improve her learners’ reading comprehension. Khansir (2012) analyzed the benefits of incorporating poetry in reading instruction. Schillinger et al. (2010) designed and taught an intensive poetry course to secondary students, which led to improved reading performance. What is of interest in the present study is that it contains evidence that may point towards concurrent CT skills, metacognitive reading strategies and general reading comprehension improvement through poetry among YL.

5. Concluding remarks

The statistical data presented above point to a strong relation between poetry instruction and enhanced performance in the CT skills selected. By the end of the intervention, the poetry group learners had become better critical thinkers and readers according to Lipman’s (1988) definition of CT, in the sense that they used specific criteria, introduced to them through the tasks, to reach reasonable and responsible conclusions, they went through a process of constant self-correction and they learned how to pay attention to different contexts. The data also seem to indicate that poetry instruction facilitated learning and applying the metacognitive reading strategies selected and that it encouraged students to think about the way they think, which is consistent with Flavell’s (1979) definition of metacognitive thinking referred to earlier. The nature of poetry seems to have fostered the process of discovering multiple layers of meaning in every reading. Finally, in relation to the overall effect of the development of CT skills and metacognitive strategies through poetry on Young Learners’ reading comprehension, the data show that improved CT and metacognitive performance was accompanied by enhanced reading comprehension. By the end of the instruction process, the experimental group learners were able to find their way through a text with considerable ease and were no longer intimidated by longer, unfamiliar texts. They had learned how to rely on their CT skills and metacognitive reading strategies in order to explore possible meanings and reach sound decisions.

Future research could focus on larger samples of primary school students in other grades as well and examine whether these results can be replicated. The experimental lessons designed for the specific intervention could be used as a poetry module incorporated in the coursebook. In addition, the length of the intervention could be extended, and participants could be tested at some point after the intervention to see whether the effect is a long-lasting one. The effect of poetry could also be explored among older students or in comparison with other literary genres. Alternatively, future studies could further explore whether the specific CT skills and metacognitive reading strategies, once acquired, can be transferred to other genres as well or whether other instructional frameworks can also contribute to the development of CT and metacognitive reading strategies. Finally, it would
be interesting if the development of other strategies through poetry instruction could be examined.

References


Appendix A
METACOGNITIVE READING STRATEGIES QUESTIONNAIRE (PRE- AND POST-INTERVENTION)
Read each statement carefully. Think about what you do before, while and after reading. Put a tick in the box that is true for you.

<table>
<thead>
<tr>
<th>METACOGNITIVE READING STRATEGY</th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I make connections between what I know and what I will read about.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. I ask myself questions to make sure I have really understood something.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I make connections between different ideas in a text.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I understand what the author’s main message is.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. I read the text again and again to understand some points.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. I have some pictures on my mind before reading a text.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I find connections between something that has happened in the text and why it has happened.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. I use clues such as title, subtitle and pictures to predict what the text is about.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. I understand what is and what is not important in a text.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. I use context clues to predict the meaning of words I don’t know.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B
SAMPLE TASKS FROM THE CRITICAL READING TESTS

making inferences
A) You are going to read two texts. Read their titles and try to guess what they are about. Look at the Glossary for the meaning of some words.
TEXT A: “Spiders Cover Greek Coast In 1,000-Foot Web For Mating”
This text is probably about ________________________________________.
TEXT B: “This Farm Collects Spider Webs for Art”
This text is probably about ________________________________________.
The first text is accompanied by the photo of a huge thick spider web covering everything on the coast.
Can you guess what photo might accompany the second text? Describe it in very few words: ________________________________________.

analyzing evidence
B) Now read the two texts carefully. Try to think of at least one similarity and one difference between them.
One similarity is that ________________________________________.
One difference is that ________________________________________.

evaluating
C) Can you understand whether the writers of these texts like spiders or not? Explain what makes you think so:
The first writer ________________________________________ because ________________________________________.
The second writer ________________________________________ because ________________________________________.
Which of the two texts was easier for you to understand? Why?
Which of the two texts was more interesting to you? Why?

making decisions
D) Which of the two texts would you choose if you wanted to surprise a friend? Explain:
I would choose ________________________________________ because ________________________________________.
Which of the two texts would you choose if you wanted to encourage someone to look for more information? Explain:
I would choose ________________________________________ because ________________________________________.
Appendix C:
SAMPLE TASKS FROM THE POETRY LESSONS

Planning
A. You are going to read two poems by two different poets. The first poem is called ‘Uphill’ and the second one ‘The Road Not Taken’. They share a common theme. Can you guess what they are both about? Discuss with your partners and write your answers:

B. Choose one of the following themes as the most likely. Justify your choice:
   a) love  b) seasons  c) travelling d) life and death e) friendship

C. Think about some things that we need to know when we start travelling. Circle some of the possible things that may worry a traveller:
   a) the weather  b) the route  c) accommodation (where to stay) d) means of transport e) expenses (cost) f) things to pack  g) duration of the journey (how long it will be) h) destination (where you want to go)

Monitoring
A. The first poem follows a question-answer pattern. There are 8 questions followed by an answer. Here is a list of the questions the poet asks:

1. Will there be beds for me and all who seek?
2. Will the day’s journey take the whole long day?
3. Does the road wind up-hill all the way?
4. May not the darkness hide it from my face?
5. Shall I meet other wayfarers at night?
6. But is there for the night a resting place?
7. Then must I knock, or call when just in sight?
8. Shall I find comfort, travel-sore and weak?

Are there any similarities between them and your answers to Task C (Planning stage) above?

B. Read through the answers given and try to put the questions back to their original place so as to reconstruct the poem:

C. Read through the second poem and try to find what this poet’s main worry is:

D. Read the poem once again and see if you can answer the following questions:

   Were the two roads the same or different? Which words or phrases make you think so?

Does the poet regret (feel sorry for) his decision? Which words or phrases make you think so?

E. Look back at your answer to Task B in the Planning stage. Do you think there is a further, symbolic theme hidden in these two poems?

Evaluating
A. Listen to the two poems being recited as you read them again.
   a) https://www.youtube.com/watch?v=yLnG6ONl9Bw
   b) https://www.youtube.com/watch?v=KUaQgRjluK

Which of the two do you prefer and why?

B. Which of the two poets do you think was inspired by their Christian beliefs to write their poems? What makes you think so?

C. Which of the two poems do you find more suitable to inspire travelling? Why?
D. How well did I read and understand?
a) Perfectly  b) Very well  c) Quite well  d) Well  c) Not so well

What strategies worked well for me?

What strategies did not work for me?

The strategies in order of usefulness (number 1 being the most useful to me in understanding the texts):

1. Analyzing evidence
2. Making inferences/Predicting
3. Evaluating
4. Making decisions
5. Activating background knowledge
6. Self-questioning
7. Connecting ideas (cause-effect)
8. Determining importance

What should I do next time? Do I need some help for next time?

Appendix D
THE POEMS USED IN PHASE 2
Lesson 1: ‘Uphill’ by Christina Rossetti and ‘The Road Not Taken’ by Robert Frost
Lesson 2: ‘The Tyger’ and ‘The Lamb’ by William Blake
Lesson 3: ‘I Wandered Lonely as a Cloud’ (also known as ‘Daffodils’) by William Wordsworth and ‘In a Station of the Metro’ by Ezra Pound
Lesson 4: ‘Hug a’War’ by Shel Silverstein and ‘A Poison Tree’ by William Wordsworth
Lesson 5: [little tree] by E. E. Cummings and ‘Easter Day’ by Oscar Wilde
Lesson 6: ‘This is Just to Say’ by W. C. Williams and ‘The Little Boy and Old Man’ by Shel Silverstein
Lesson 7: ‘Ozymandias’ by Percy Bysshe Shelley and ‘Ozymandias’ by Horace Smith
Lesson 8: ‘Home is so Sad’ by Philip Larkin and ‘The Shortest and Sweetest of Songs’ by George MacDonald

Aspasia Deliligka (aspa72deli@yahoo.gr) holds a B.A. in English Language and Literature from the National and Kapodistrian University of Athens and an M.Ed. in TESOL from the Hellenic Open University (2019) with distinction. She has been working as an EFL teacher in the public sector for the last 26 years. Her interests include teaching English Literature to Young Learners and intercultural education.

Christine Calfoglou (xkalfog@yahoo.gr | kalfoglou.christina@ac.eap.gr) was a University of Athens MA in Translation and Translation Theory scholar and obtained a PhD in Applied Linguistics and Language Acquisition from the University of Athens with distinction. She is currently teaching on the M.Ed. TESOL programme of the Hellenic Open University, for the Literacy module of which she has written a volume. Her research interests and published work include distance learning methodology, L2 writing instruction, grammar, the semiotics of language and art, poetry translation, translation theory.
Developing soft skills in adult learners of English in online courses: The case of Skyeng online school and Skyes digital platform

Alexey KONOBEIEV, Violetta USANOVA & Olesya GILMUTDINOVA

The paper analyses the importance of soft skills for adult learners, provides examples of soft skills valued by employers, examples of online courses aimed at developing soft skills, and focuses on the system of developing soft skills and resources, offered by Skyeng online school and Skyes digital platform for universities, used in several countries. The system offered is structured according to learner needs and course types and has been used by over one hundred thousand learners.

Key words: soft skills, online courses, EFL, ESL

1. Introduction

The importance of soft skills for work and life has been increasingly recognised in many countries. While there is no single exhaustive list of soft skills (though there are varying definitions of soft skills), they become increasingly recognised by employers as a key to employees’ success in the workplace. Therefore, it is important to understand what the content of education should be, how soft skills can be taught, and consider already existing ways of creating a framework for developing soft skills.

Collins English dictionary defines soft skills as “Desirable qualities for certain forms of employment that do not depend on acquired knowledge: they include common sense, the ability to deal with people, and a positive flexible attitude”. A more practice-oriented definition is offered by (Kenton, 2021), who defines soft skills as “character traits and interpersonal skills that characterize a person’s relationships with other people. In the workplace, soft skills are considered to be a complement to hard skills, which refer to a person’s knowledge and occupational skills.” Soft skills are skills which are desirable in all professions; these include critical thinking, problem solving, public speaking, professional writing, teamwork, digital literacy, leadership, professional attitude, work ethic, career management and intercultural fluency. This is in contrast to hard skills, which are specific to individual professions (Claxton, Costa & Kallick, 2016; Lippman et al., 2015).
1.1. Need for soft skills

Soft skills are regarded as very important in many countries, as recent research demonstrates. In the USA a survey, carried out in 2016 by the HR company Adecco Staffing, showed that 44% of executives (out of 500 senior executives surveyed) said that a lack of soft skills was the biggest proficiency gap they saw in the workforce. A report by Bloomberg in 2018 reveals that:

If recent graduates are not well-prepared for their new jobs, it is not because their hard skills are deficient. Some 90 percent of corporate respondents and 88 percent of academics surveyed said new recruits have the hard skills, such as computer literacy and written communication, to do their jobs successfully. But both groups, however, were far less satisfied with new employees’ soft skills. Nearly four in 10 corporations and almost half of academic institutions said new hires lack the soft skills they need to perform at a high level (“Building Tomorrow’s Talent: Collaboration Can Close Emerging Skills Gap”, 2018, p.2).

The survey included 200 senior-level respondents, 100 from businesses and 100 from academia, and showed that 34% of corporate respondents and 44% of respondents from academic institutions believe that new recruits are prepared with hard skills but lack soft skills to be effective. The Global Trends Report-2019 by LinkedIn involved 5165 businesses in 35 countries, and 92% of respondents said that soft skills were as important as hard skills, while 30% of employers stated that it is more important to employ people based on their soft skills than hard skills. What’s more, 89% of unsuccessful employees lack soft skills. It would appear that soft skills are becoming a key factor in employability and success at work. The Australian Chamber of Commerce and Industry in its 2019 Review of Australian Qualifications Framework Submission to the AQF Expert Panel states that soft skills should be a requirement for every qualification. This strongly correlates with the statement that “In the next ten years, soft-skill intensive occupations are expected to account for 63% of all jobs in Australia” made by AIM Business School CEO Ben Foote in Sky News business programme “The Ladder” in 2018. This demand makes soft skills an important part of learning content in professional education (Myers, 2021).

However, perception of the importance of soft skills varies from country to country. For example, in Russia soft skills are recognised as important, but digital skills (which are hard skills) are viewed as slightly more important than soft skills by Russian employers. ANCOR, an international staffing group, surveyed 373 CEOs and 4013 employees, who said that digital skills are among the very important skills that employees lack, and that the importance of digital skills is nearing that of soft skills (ANCOR press-release, 2021). Another survey, carried out in 2020 by OPORA Russia, included 526 employers, 63% of which stated that soft skills are key skills at work. This means that soft skills are recognised as desirable and may be part of teaching content in Russia as well as in other countries. Stek (2021) presents an educational soft skills experiment with IEM graduates, and it provides evidence that soft skills learning can effectively be introduced in existing courses. The graduates self-rated their competence levels of 36 soft skills before and after the course that provided soft skills workshops and a case study. In the first survey, ”strategic thinking” ranked low and could be improved the most in the second survey.

Because of the rising importance of educational soft skills, the necessity to teach soft skills has become a major concern for educators and employers worldwide (Crowley, 2019). Because soft skills are poorly defined, teaching them is more challenging, compared to classical skills. For this reason, the
first step consists of understanding how to evaluate them, so that educators can track student progress.

As for teaching, evaluating soft skills is harder than technical skills. “Quizzes or exams cannot accurately measure interpersonal and leadership skills” (Zhang, 2012, p. 159). Group projects seem to be a good way to develop soft skills, but evaluating them still represents a major obstacle. Researchers consider peer evaluation a good compromise between working in groups and an objective evaluation. The research conducted on this topic reported both positive and negative results although with few participants “is an initial step in designing and validating a peer assessment scale” (Zhang, 2012, p. 161).

As Dell’ Aquilla et al. (2017) write,

soft skills are important to everyone in every context, as it entails the involvement of personal aspects within any relationship. Soft skills are important to students, as they are linked to job performances and career development; they are crucial for employees who need to manage their interactions and emotions in order to interact effectively with customers and get engaged with the workplace missions; for management and leadership skills, as they help lead teams towards common and shared goals, accomplish organisational missions and support organisations in their future directions and visions.

1.2. Soft skills as content of specialised courses

Online courses that have the words “soft skills” in their description are highly popular. For example, Coursera offers soft skills courses that have as many as 1-3.1 million students per course (e.g. the courses Leading People and Teams from University of Michigan with 330,000 students, or Learning How to Learn: Powerful mental tools to help you master tough subjects from McMaster University). Other examples of soft skills courses available on Coursera include: Teamwork skills: Communicating Effectively in Groups (offered by University of Colorado Boulder, 150 000 students), Creative Thinking: Techniques and Tools for Success (offered by Imperial college, London, with 190 000 students).

As we see, such courses are not job-specific, but they deal with universal skills. It is important to determine the soft skills that may be used as the content of soft-skills oriented courses.

Soft skills are more difficult to teach and to learn than hard skills, because hard skills are related to information and job skills, and therefore to specific knowledge and easily measurable performance outcomes, while soft skills are dependent on traits of character and strategies of behaviour and are difficult to measure. The Australian Chamber of Commerce and Industry (2019) states: “It is very difficult to align soft skills with qualifications and to allocate them to competence levels”.

This means that for different businesses and positions the importance of particular soft skills may vary, which will affect the contents of the soft-skills oriented courses.

Various employees list different skills as important for their businesses, including communication, adaptability, leadership, persuasion, organization, teamwork, collaboration, problem-solving, time management, interpersonal skills, emotional intelligence, stress management, creativity, active
listening, empathy, public speaking, social skills, critical thinking. The World Economic Forum provides a list of 35 soft skills in its Future of Jobs report and a list of top 15 skills in 2025 includes 10 soft skills and only 5 hard skills (The Future of Jobs Report 2020, p. 36). The article Soft skills — skills of the 21st century: what do employers value most of all in Russia and in the Perm region (2019), published by Higher School of Economics, Russia, lists the top 10 soft skills in the world and the top 10 soft skills in Russia.

<table>
<thead>
<tr>
<th>Top 10 soft skills in the world</th>
<th>Top 10 soft skills in Russia</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ability to solve complex issues</td>
<td>1. Active learning</td>
</tr>
<tr>
<td>2. Critical thinking</td>
<td>2. Coordination</td>
</tr>
<tr>
<td>3. Creativity</td>
<td>3. Negotiating</td>
</tr>
<tr>
<td>4. Leadership</td>
<td>4. Teaching people</td>
</tr>
<tr>
<td>5. Coordination</td>
<td>5. Information literacy</td>
</tr>
<tr>
<td>7. Decision making</td>
<td>7. Oral communication</td>
</tr>
<tr>
<td>8. Customer focus</td>
<td>8. Ability to solve complex issues</td>
</tr>
<tr>
<td>10. Cognitive flexibility</td>
<td>10. Time management</td>
</tr>
</tbody>
</table>

In comparison, the Boostrs’ (2020, p.3) study states that there is a core of eight fundamental soft skills which are essential to most jobs. These are communication (written and oral), time management, decision making and judgment, team spirit, adaptability, analytical skills and autonomy. Beyond these fundamental soft skills, there are other, specific, soft skills which tend to be required in different jobs and functions. For example, for the Sales function, two of the most sought-after soft skills are “ambition” and “handling conflicts”. The number of soft skills required varies with the job function corresponding to the position recruited for: as an example, the Human Resources function requires the greatest number of soft skills (25) among all corporate functions. We also find that the soft skills required depend on the seniority of the position. As an example, senior management positions require more autonomy, initiative, people management, self-confidence, and strategic thinking.

This means that although there may be varying needs for soft skills depending on what kind of future job a course is aimed at, it is still possible to determine the core soft skills that may be used as the content focus of most courses. Based on comparison of the reports listed, we believe that the core skills should include the following, such as time management, planning and coordination, decision making, adaptability, ability to solve complex issues, teamwork, leadership, critical thinking, interpersonal skills, negotiating. Other skills may be added to help students reach their job-specific goals.

2. Incorporating development of soft skills in courses of the English language: a case of Skyeng courses

A separate soft skills course will require time and effort (separate soft skills courses on Coursera last, on average, 40 weeks per skill), and since a lot of soft skills are interpersonal, such courses need to
include a significant amount of communication. Therefore, it is possible to incorporate development of several soft skills into a course that focuses or heavily relies on communication. Courses of foreign languages are particularly well-suited for this purpose, as the experience of Skyeng, one of the largest online schools in Europe, demonstrates.

Skyeng is a large online school with divisions that develop and teach courses from preschool age and school-age students (Skysmart), to adult learners and offers courses on a university level. These courses are used by universities in Russia, Belarus, Indonesia, Kazakhstan, Mexico, Morocco, Pakistan, Peru, Thailand, Tunisia, Turkey, Vietnam, in total by more than 250 institutions, 2059 teachers and more than 37000 students as of November 2021.

Skyeng began in 2012 as a language school, offering English as a foreign language courses, but since 2020 it added courses for universities (both general English and English for Specific Purposes offered by Skyes for Universities platform), courses in other subjects for school students and in 2021 the school developed a platform and materials for 15 primary and secondary school subjects (Skysmart Class) and nanodegrees courses for those wanting to acquire new qualifications in computer programming, management and other subjects (Skypro university). As of November 2021, English courses alone are taught to over 110 000 students by about 12 000 teachers.

Skyeng develops soft skills through EFL and ESP courses for adult learners through a system of lessons and units.

The general stages of developing soft skills follow the same PPP (presentation, production and practice) model as a regular language course does:

1. Presentation: raising awareness of a soft skill and its uses through a communicative situation.
2. Production: teaching to use the skill through controlled exercises or case studies.
3. Practice: practicing the skill through exercises and situations that require it throughout the course.

A more detailed description of the methodology used in developing Skyeng courses is provided by Smyshlyak in “ESL/EFL online lesson development (based on Skyeng practices and techniques)”, (Smyshlyak, 2020), who demonstrates how General English lessons are created to cater for the learners’ needs and learning goals. Within one course, the choice of learning tools can vary from lesson to lesson, while remaining within the same lesson framework, depending not only the goals of the lessons, but on the goals of the students, with the lesson structure adapted to better get the content of the lesson across to the students. This means that although the PPP model prevails in the General English courses, it can be altered to accommodate particular goals and form of learning, like projects or case studies.

Stoyanova (2021) describes the principles of development of online courses of English for Specific Purposes and shows how the content of courses is selected according to the results of numerous customer development interviews, as well as field studies, which enable course developers to design courses that provide adult learners with the training that addresses their future employers’ needs and covers the fundamentals of the specific professional areas so that the courses are broad enough to prepare students for further professional development without focusing on one narrow job-specific
Such methodological balance of content is viewed as a tool of personalisation of the courses.

Konobeev et al. (2020) describe the model of personalised learning used by Skyeng to offer personalisation on several levels: 1) the lesson, where certain exercises can be skipped or supplemented with other resources, 2) the entire course, where students can choose between fast track (covering the essentials of the programme only) and full track (covering each topic in detail), and 3) the level of the entire school where students can use a wide variety of extra resources, self-study courses and training apps to supplement their learning. Valkova (2020) mentions that students learning English with Skyeng courses become motivated and expand their vocabulary even when they use supplementary courses only.

It appears that Skyeng has developed a flexible model of creating personalised courses for adult learners that makes it possible for material writers to include different kinds of content into the lessons, and for students to decide on how much of the particular content they want to study as long as they meet the basic course requirements. This model seems convenient because it incorporates the development of soft skills in lessons and activities aimed at learning English, and at the same time makes soft skill practice a recurrent pattern in the course. However, the amount of soft skills depends on the type of course offered.

The Skyeng system of soft skills development through EFL/ESP courses is based on learners' needs. In General English courses there are lessons that include soft skills as secondary aims, while the main aim is to help students reach a certain level of proficiency in the English language. In the *New General English course*, soft skills are an integral part of every unit in most lessons. Exercises to develop soft skills play a supporting role in such lessons as students’ goals in General English courses are less job-related than in professional courses. Therefore, the soft skills taught in such courses are more interpersonal skills, emotional intelligence and skills needed to communicate in cross-cultural settings in very broad contexts such as working in an international office and being aware of cultural features of co-workers. However, job-related soft skills are also taught through *New General Course* lessons. These skills are selected in such a way that they, too, would be useful in a variety of contexts, including everyday life. For example, students are taught to plan their daily schedules effectively and are provided with scaffolding to boost their planning skills. The following example (Image 1) from an elementary level lesson makes students consider what productivity means for them as well as what elements make up a productive routine; thus raising their awareness of the skill:
Create your lists to become more productive. Explain your points. The wordlist and questions in the cards might help you

If you are not ready to talk about yourself, you can create lists for someone you'd like to help become more productive: your friend, acquaintance, colleague or a film character.

- What would you like to achieve in the next month?
- What activities would you like to incorporate into your daily life?

Image 1. An exercise to help elementary students to focus on the soft skill of productivity

Then elementary level students are led to consider their own productivity:

<table>
<thead>
<tr>
<th>To-feel list</th>
<th>To-be list</th>
<th>To-do list for the next month</th>
</tr>
</thead>
</table>

To make a transition from presentation to production, students compare how attitudes to productivity have changed over the past century and how these ideas compare to the ones that they may have had or have started to develop. After comparing and discussing routines of people from the past as well as modern people, they think about such questions as “What part of your routine is productive? What is not productive? How can you make it more productive?”. In the next lesson, they discuss and practise time management and organisation skills with the focus on making their routines more productive; thus putting the skills listed into a very personal perspective. Once again, this is done as a secondary aim to developing communication skills, (namely, learning how to make suggestions) which is a way to practise the soft skill in a language exercise:

Image 2. Information and discussion questions to help elementary students focus on the soft skill of productivity

Kim Kardashian
I usually wake up very early, at 6 o’clock. I get dressed, make my bed, brush my hair, and at 6.15, I go running — it gives me a lot of energy for the day. When I get home, I take a shower. At 8 o’clock, I wake my children up, and we have breakfast together. At 12.00, I have a couple of meetings. At 18.00, I hang out with my friends. At 21.00, I go to bed. When I can’t fall asleep, I read some books or posts on Instagram.

You can also discuss:

- Whose schedule (Charles Darwin's, Sigmund Freud's, or Kim Kardashian's) is the best? Why?
- Who is more productive — a modern person or a person of the past? Why?
In the next lessons, students move on to working with interpersonal skills when discussing aspects of working in multinational companies. Students see from the experience of the lesson’s character that working in multinational settings may be very beneficial (e.g., learning interesting information from around the world), but that they need to make an effort sometimes to bridge the cultural gap; thus becoming aware of the skill through exercises that further teach them how to practise it through communication. Examples of such lessons and exercises in the *New General English course* on different levels are numerous.

Since ESP courses are mostly meant for adult learners who already have a job or seek one. Soft skills in such courses become primary goals along with communication skills development, and there are entire units that heavily focus on the development of work-related soft skills. While in the *General English courses* students may remain unaware that they are actively acquiring certain soft skills (unless they read lesson descriptions), in ESP courses there are units dedicated to teaching particular soft skills.
For example, the *New Business English course* has, among others, units dedicated to team building, leadership, and project management.

![Table of contents for New Business English Pre-Intermediate course](image)

**Image 4.** The lessons in a “soft skills” unit in the New Business Pre-Intermediate course by Skyeng

The Project management unit in the *New Business English Pre-Intermediate course* centres on developing time management and organisation skills. Other units in the course help to develop negotiation skills, critical thinking, teamwork, collaboration and problem-solving skills. Most skills from the list are present in the *New Business English course* and are developed through case studies and role plays. Case studies help students to become aware of a lack of soft skills as a possible cause of failures in work and communication. By trying to solve real-life problems through case studies and focusing on the soft skills that are key to the solutions students become aware of their importance and learn to apply soft skills to a variety of professional situations. As we have mentioned earlier, in ESP courses the focus on soft skills is at least equally important as that on language learning.

Another course rich in soft skills, is the *Just Speak Course*. This course is available on two levels: upper-intermediate and advanced. The main goal of this course is to prepare students for a long stay and work in the USA (upper-intermediate level) and the UK (advanced level), so in a way it is a cross between an EFL and an ESP course. The course covers a wide range of everyday and work situations, it has a storyline where students trail the life of the main character, but then they roleplay a variety of situations and learn not only to speak English in these situations but also to deal with people, develop persuasion skills, critical thinking, social skills, public speaking, emotional intelligence, and many others that can be a valuable addition to the skills taught in the New Business English course, or an important preparation for life in a multicultural environment.

In another lesson of the *Just Speak Course*, Upper-Intermediate students learn about workplace culture. They think about their own expectations and habits; thus raising their awareness of teamwork as a skill:
And after learning about - and practising - interpersonal skills in the office, they move on to more difficult issues - ethical dilemmas that can be important in building working relationships with colleagues and adhering to rules as well. They face dilemmas and then learn about working rules and expectations. Not only do such exercises provide excellent basis for communication, but they also help to develop emotional intelligence, interpersonal skills, and collaboration skills.

Teaching English to university students also has its own features and specifics. On the one hand, students need a good course in General English to prepare them for taking up a course in English for their future job. On the other hand, first-year, and sometimes second-year students have a particular need for lessons that would teach them how to study at universities. The ability to study consists not only of certain learning strategies and a knowledge of how the university operates, but it also requires such skills as planning, time management, teamwork, collaboration, problem-solving, interpersonal skills, emotional intelligence, stress management, creativity, active listening, empathy, public speaking, social skills, critical thinking - in short, many of the skills listed by businesses as highly desirable for potential employees. Therefore, it is important to include teaching these skills in university-level courses.

The Skyes for Universities digital platform contains entire courses, special trainers and sets of interactive exercises, most of which have been developed for university students. The core Skyes courses are general English courses with a focus on academic skills, supplemented by ESP courses that students need for their future jobs. On top of these courses, sets of training exercises (for example, vocabulary trainers, grammar trainers, pronunciation trainers) and short courses are available for the students to choose from to help them solve particular problems and close gaps in their school education.

A large part of Skylike courses is dedicated to teaching soft skills in a way that is relevant to university and college students. In the Skylike A2 course (Alexiou et al, 2020), the key skill is planning, when students learn to plan their time and work in a variety of settings and contexts. Students learn how to use different planning tools, including digital calendars like Google calendar, they learn how to schedule their time to prepare for exams, they learn and practice creation of SMART goals (Specific,
Measurable, Achievable, Relevant, Time-bound) and put this new skill to practice in their own university setting:

**Match the words with the questions**

| Measurable | How do I know that it is achieved (numbers, measures)? |
| Relevant   | How much time do I have/need? |
| Achievable | What can I do to achieve it? |
| Time-based | What good things will happen when I achieve it? |
| Specific   | What exactly do I want to achieve? |

**Choose a goal or write your own. Make it SMART**

1. I want to go on a nice trip this year
2. I want to make a nice surprise for my friend
3. I want to win in a sports competition

**My own goal:**

**Image 6:** tasks to help to create SMART goals. Skylike A2

Even though most of Skyes course content focuses on academic contexts, soft skills are practised in non-academic courses as well, as they are needed not only for work purposes, but in the daily life as well:

**Image 7:** (left) Applying planning skill to traveling, Skylike A2 and (right) limiting conditions in a planning task in Skylike A2
To make such planning practice realistic and help students learn to make their own life plans, exercises include certain limitations, like limited funds, limited time and several other conditions.

As it has been mentioned earlier, Skyes courses consist of units dedicated to developing such soft skills for academics, as Planning, Exam preparation, Networking at the university, Cross-cultural collaboration to study, Critical thinking and others. In the rest of the units, soft skills are incorporated in the lessons, but do not control the contents of the entire units.

The Skyeng system of developing soft skills in adult learners through teaching English may be represented in the following table:

<table>
<thead>
<tr>
<th>Type of courses</th>
<th>Types of soft skills</th>
<th>How soft skills are taught</th>
</tr>
</thead>
<tbody>
<tr>
<td>General English</td>
<td>Interpersonal, emotional intelligence, organisational, time management as applied to everyday life</td>
<td>As secondary goals in lessons</td>
</tr>
<tr>
<td>ESP (professional + Business English)</td>
<td>Time management, leadership, collaboration, persuasion etc with focus on work</td>
<td>As primary goals in entire units and lessons</td>
</tr>
<tr>
<td>English for university students</td>
<td>Planning, exam preparation, networking at the university, cross-cultural collaboration to study, critical thinking for academic purposes</td>
<td>As part of nearly every communication exercise</td>
</tr>
</tbody>
</table>

Table 1. Skyeng system of developing soft skills through EFL/ESP courses

It is important to note that students may choose to supplement courses with lessons or units from other courses. For example, a New General English lesson, where soft skills are secondary, can be supplemented by an entire unit from the New Business English course and thus continue acquiring soft skills as long as he/she deems fit.

Such a system allows to develop soft skills in adult learners in accordance with their interests and needs and without shifting the focus of studies away from the students’ goals.

3. Pedagogical implications

The model of developing soft skills through online EFL/ESL courses for adult learners, used by Skyeng online school and Skyes digital platform for universities, shows that soft skills can be taught as part and parcel of more general English-language courses. Soft skills for teaching can be selected, based on research in job markets. This research shows that, although the sets of soft skills that are valued in a particular country, can be more or less unified, they can differ greatly in particular industries, with different companies valuing different subsets of soft skills. Therefore, when developing courses of English for adult learners, it is important to take into account not only the broad content of international industries, but also expectations and requirements of a particular industry in the country
where this course is going to be implemented. Also, it is important to use a course design that will allow students to focus more on the skills that they feel they need and less on the ones that they do not expect to benefit from, while still meeting the basic requirements of the programme.

The PPP model offers a convenient framework to implement the development of soft skills in such courses on different levels. At the same time, other methods and techniques can be used, like CLIL to integrate professional content and language learning, and to better meet the goals of adult learners. Many soft skills like leadership, coordination, customer focus, negotiating and ability to solve complex issues, are more connected to job situations. Therefore, CLIL and ESP courses seem to be particularly well-suited for the task.

Soft skills may require an extensive use of such tasks as debates, critical analysis and evaluation of information, planning and time management activities, case studies with real-life cases when goals were not achieved because of a lack or insufficient application of soft skills. Cross-cultural communication cases can be particularly useful in building strong interpersonal skills as well as persuasion and leadership skills. The wide range of soft skills makes it necessary to use a wide range of techniques in a soft-skills inclusive course.

It is important to further research the comparative effectiveness of different levels of teaching soft skills (lesson level vs unit level) and to select content for ESP courses for different professions regarding development of soft skills in adult learners of English. Also, the next steps in research may include categorisation of soft skills into those that can be best taught through General English courses and those that can be demonstrated through job-related situations. Such research will help to select teaching content more carefully and to match it to the requirements of particular industries and learners. A study to compare the effectiveness of different strategies of teaching soft skills is also needed to select more precisely the teaching tools to be used in such courses. For this research the experience and teaching model of Skyeng courses can be useful.

4. Concluding remarks

As shown above, soft skills are becoming an important component of teaching content for adult learners, as they are required in many jobs. Even though different jobs require different skills, it is still possible to determine core soft skills that can be used as a basis for soft skills-oriented courses or courses. Online courses aimed at development of separate soft skills are offered by various institutions and some of such courses are highly popular. However, these courses require time, while development of soft skills is based on communication. Therefore, it is possible to incorporate development of soft skills in EFL and ESP online courses for adult learners, as Skyeng experience demonstrates.

The PPP model allows a smooth implementation of soft skills in EFL and ESP online courses. Depending on the learners’ needs, soft skills can be taught on a lesson level, when they become part and parcel of exercises and activities of a lesson, as well as on a unit level, when entire units are dedicated to development of particular soft skills.
References


**Alexey Konobeiev** (alexey_konobeiev@mail.ru) is the Academic Director in the Content department at Skyeng online school. He holds a PhD degree in Applied Linguistics (teaching English as a Foreign Language). His main interests include online teaching and learning, teaching writing, testing and discourse analysis.

**Violetta Usanova** (usanova.violetta@skyeng.ru) is the Head of Skyes - digital platform for universities. Her main interests include EdTch,higher education organisation and management, online learning and continuous professional development.

**Olesya Gilmudtina** (l.gilmudtina@skyeng.ru) is the Head of Educational Content Lab at Skyeng. Her main interests include online course development and design, teaching grammar, and evidence-based education.