AMERICAN UNIVERSITY OF ARMENIA

College of Humanities and Social Sciences

The Impact of Multiple-Choice and Short-Answer Questions on EFL Learners' Test Performance

A thesis submitted in

partial fulfillment of the requirements for the degree

Master of Arts in Teaching English as a Foreign Language

By

Mariam Ghulinyan

Rubina Gasparyan, Adviser

Irena Galikyan, Reader

Yerevan, Armenia

2021

ACKNOWLEDGEMENTS

I would like to greatly acknowledge everyone who supported and assisted me throughout the research.

Firstly, I would like to express my gratitude to my supervisor, Professor Rubina Gasparyan, whose support was irreplaceable during these months. Your guidelines and feedback regarding all aspects of my thesis helped me to achieve more insightful results and bring the research to a higher level.

I would like also to thank all my professors who guided me in MA TEFL program for the valuable knowledge they provided. They equipped me with the necessary tools that contributed to make the right decisions and successfully complete the study.

In addition, I would like to acknowledge the teachers from the English Afterschool program who worked with me as a part of my research. I am very thankful for their support and willingness to help with my study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The potential influence of multiple-choice and short-answer question formats on students' test performance has been a subject for researches for a long time. It is established that EFL learners show different performance and have different attitudes towards MCQ and SAQ items. This study aims to determine how the MCQ and SAQ question types can affect students' test performance. Specifically, the study investigates whether the learners' proficiency has an impact on their performance of MC and SA question formats, if the constant practice of SAQ items can result in better performance of MC format and overall language test, and what are the students' perceptions towards SAQ and MCQ formats.

For this purpose, 19 students from two Low-Intermediate Proficiency language level groups participated in the research. The findings showed that students' performance on both MCQ and SAQ formats does not differ much. Moreover, the learners' test performance is mainly affected by their language proficiency. In addition, students give preference to both MC and SA question types, although mentioning that in most cases the MC tasks are easier to complete.

The results of the current study suggest that the implementation of short-answer question format in language tests can not only positively affect the validity and reliability of the assessment but also provide the teachers with insightful feedback about learners' course achievements. On this basis, it is preferable to mix both MCQ and SAQ formats when assessing small groups of EFL learners.

Keywords: formative assessment, assessment in CLT, online assessment, multiple-choice questions (MCQs), short-answer questions (SAQs), question formats, student perceptions

Chapter One: Introduction

Assessment is a crucial part of the language learning process. It intends to measure learners' language knowledge and abilities, as well as to get general feedback on their learning achievements. The choice of the question format highly impacts the assessment and students' overall performance on the test. As Bachman (1990) highlights, it is important to investigate the potential influence of question formats on students' test performance. In order to create a valid assessment tool, the test developers need to use the data about the interaction between question types to design assessment tools that can be considered as a credible and reliable measure of language abilities. Consequently, among the most commonly used test items are multiple-choice (MCQ) and short-answer (SAQ) questions. The comparison of these two test formats has always been a ground for debates between language professionals on the subject of their reliability in studies. Besides the debatable aspect, the main question still remains the choice of a suitable format of test items that can meet the objectives of the course. Regardless of some similarities between the skills measured by MCQ and SAQ items, they provide information about different features of assessment (Budiyono, 2018). According to various studies, the blended approach with an increased ratio of SAQs can result in higher reliability of the grades (Faremi, 2016; Mozaffari et al., 2017; Budiyono, 2018).

The advantages of both MCQs and SAQs make them very valuable for teachers and test creators to use them as a part of formative assessment. These two formats are used for the measurement of low cognitive level abilities (e.g., remembering, understanding and applying) (Chan, 2009). Still, there are several disadvantages of MCQs that need to be considered during language assessment. By forcing the test takers to choose the single correct answer, the MCQs prevent revealing students' in-depth knowledge which then can result in a surface learning

approach (Mullen & Schultz, 2012; Tychonievich, 2012). Unlike MCQs, the SAQs as a type of constructed-response items, provide opportunities to evaluate learners' higher-order thinking skills (e.g., analyzing, evaluating, creating) (Tankersley, 2007). Besides, communicative language teaching (CLT) also aims to teach the skills necessary to students for real-life interactions, which means that the essential aspect of the assessment is the authenticity of test items. The MCQs fail to precisely replicate the learning outcomes and language competence of the students (Canagarajah, 2006). Whereas the constructed-response items are more applicable to reflect on the students' communicative abilities and provide the teachers with valid and reliable data about learners' level of achievement.

To further investigate the relationship between student performance and test item types and their impact on test results, research is conducted in an English Afterschool program in Yerevan, Armenia. The participants are two Low Intermediate proficiency level groups. The study aims to draw correlations between MCQs and SAQs to find out the differences in their effect on learners' test performance. By giving the learners similar tasks with different question types, makes it possible to recognize to what extent the answers of the students vary regarding the item format. In addition, the research tries to find out whether familiarity with the SAQ format can be an advantage when answering MCQ type of items, and also discover test takers' preferences of the test format. Furthermore, the findings provide some information about the practicality of SAQs as a part of not only grammar and vocabulary tasks ("The Language Use" section of the test), but also when assessing the reading and listening comprehension.

To explore the impact of multiple-choice and short-answer question formats on participants' test performance, the current study tried to find out the answers to the following research questions:

1. Does student test performance on MCQ and SAQ formats change depending on their proficiency?

2. How does the systematic treatment with the SAQ format affect students' test performance?

3. What are students' perceptions towards MC and SA question formats?

Chapter Two: Literature Review

2.1. Item Formats in Formative Assessment

Formative assessment is an evaluation process that aims to monitor students' learning process. Specifically, it provides teachers with general feedback about the changes in student language performance, as well as impacts on ensuing assessment practices (Dunn & Mulvenon, 2009). Moreover, the evaluation process reflects the main objectives of the course and is repeated over a period of time to measure learners' progress (Rademakers et al., 2005). The assessment process needs to meet certain conditions to have high productivity and precisely represent the students' achievements (Stiggins, 2005). Consequently, to attain positive changes in the learning process, it is essential to develop a sound assessment tool.

The validity and reliability of the assessment are linked to the testing format. There are a number of studies claiming that the format of the test has a huge impact on students' test performance. Item formats affect not only test results, but also measure different language skills and provide information about various aspects of students' linguistic competence. For this reason, the construction of valid test items is significant to diminish the possible negative effect of the testing format on test takers' performance to get consistent results about their learning process (Bachman, 1990; Kobayashi, 2002; Solano-Flores & Trumbull, 2003; Cheng et al., 2007). The item formats interfere with the construct and the diversity of the language skills measured in the test. Moreover, each of the item formats requires the test takers to apply certain abilities to answer the questions (David, 2007). Hence the scores received by the students may be altered by their performance and do not represent their linguistic competence well enough. As a result, test developers need to pay particular attention to assessment item formats and their

interaction with learners' performance to develop more valid and reliable tests to measure language skills.

One of the most common item formats used in the assessment is the multiple-choice question type (MCQ). MC items are widely used in second language testing as they cover a broad range of knowledge and skills (Bleske-Rechek et al., 2007; Currie & Chiramanee, 2010; Tychonievich, 2012). This format is considered an objective, easy-to-check, and valid testing method which allows the teachers to broadly use it as a part of achievement tests. MCQ items require the respondents to choose the single-best correct answer from the given choices. The right answer is called the key and the rest of the options are the distractors (Faremi, 2016). The options can vary from two to four depending on the question type. MCQ items are mostly used for checking acquired facts and factual recall information. On the other hand, the opponents of the MCQ format argue that this question type is inapt for the assessment of higher-order thinking skills and may contaminate the results of the test.

Another broadly used testing format is the short-answer questions (SAQ) item type. It is a type of constructed-response questions or open-ended questions that requires the test takers to apply their knowledge, language skills, and critical thinking abilities in authentic tasks (Tankersley, 2007). In this item type, the students are requested to produce their own responses without providing any hints. SA questions are used in a variety of formats ranging from filling the missing word to writing one or two sentences (Chan, 2009). Generally, they are used to evaluate learners' low cognitive level skills and in-depth understanding of concepts. Short-answer items are relatively easier to develop, however, the scoring can require multiple raters (Budiyono, 2018).

2.2. Assessment in Communicative Language Teaching

Communicative language teaching (CLT) is a teaching technique that emphasizes meaningful negotiations and interactions between students in the language learning process (Wu, 2008). For the development of communicative competence, CLT encourages using authentic, real-life activities which allow the learners to practice their language in different contexts.

The development of a suitable assessment tool for checking language proficiency can be very challenging for EFL teachers (Adair-Hauck et al., 2006). The tests need to meet not only the objectives of the course but also provide valid and reliable results about students' language competence. The materials used during the assessment need to be authentic to imitate real-life situations where the students will be able to apply their problem-solving skills (Huang, 2016). As an attempt to answer authentic questions, learners analyze, synthesize and evaluate the information that leads them to deeper level thinking (Tankersley, 2007). By facing real-world problems, they process their knowledge in complex situations and demonstrate their intellectual independence. Many language tests consist of MC item types that measure only lower-cognitive level skills. Therefore, they do not entirely reflect students' language proficiency and contradict the key concepts of CLT (Huang, 2016).

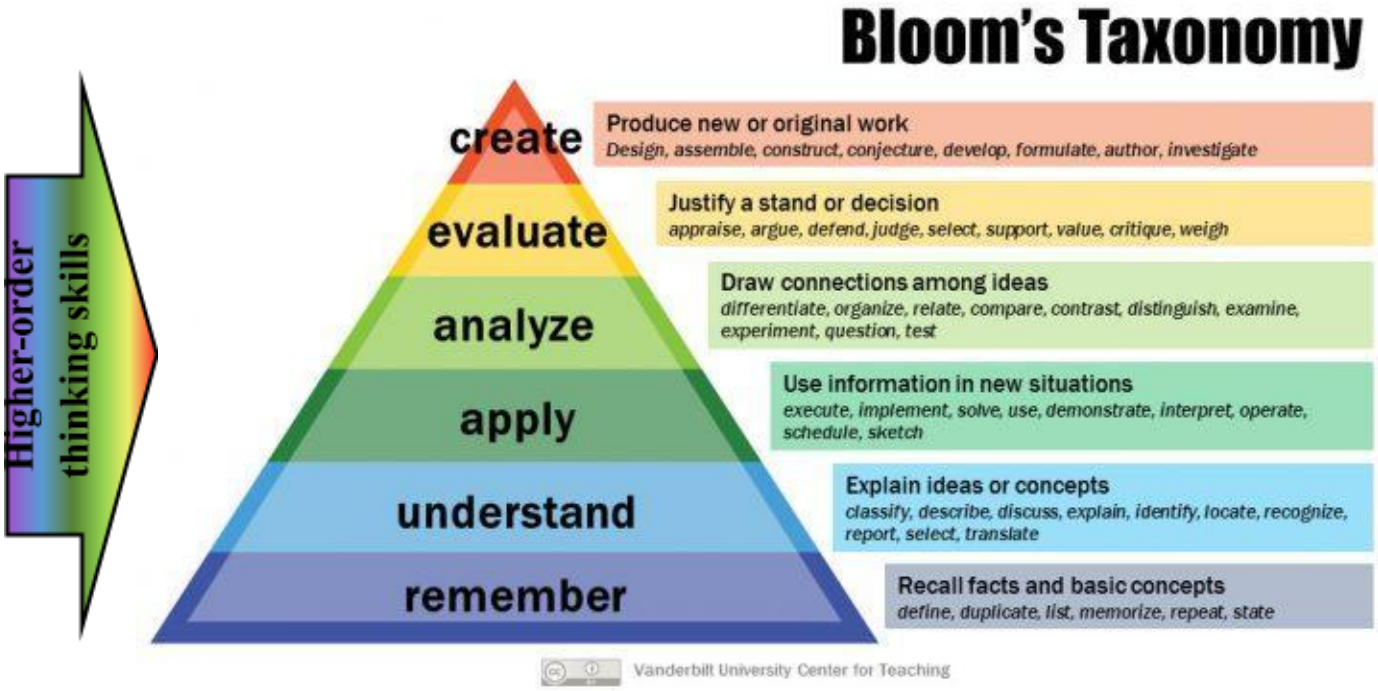
To demonstrate their complete language knowledge, constructed-response items such as open-ended or short-answer questions are more applicable for performance-oriented tasks. Alderson (2000) states that SAQs can assess not only factual reading skills but also students' inferential and evaluative reading abilities which results in a higher construct validity of this format. The research conducted among the Korean high school students revealed that the SA question format provides the students with an opportunity to develop essential abilities in foreign language learning (Kang, 2005). The integration of more SAQ activities in language classrooms

can especially benefit the low proficiency students since the constant practice of this question format can result in better test performance.

2.3. Assessment of Higher-Order Thinking Skills

Higher-order thinking skills are essential in language learning, especially to acquire new concepts. This process results in a meta-cognitive level of thinking, which leads the students to deeper understanding and learning of material rather than simple memorization of the facts (surface learning) (Alfred et al; 1998). At the meta-cognitive level, the students not only understand the concepts but also make judgments, evaluate through self-reflection, link them with previously obtained knowledge (Granville et al, 2004; Tankersley, 2007). In that sense, higher-order thinking skills emphasize the ability to go beyond basic skills and demonstrate knowledge in a context. Blooms' taxonomy can be very useful for this purpose since it allows to recognize the level of understanding of the concepts (Figure 1).

Figure 1. *Blooms' Taxonomy*



Note: by P. Armstrong (2010). Vanderbilt University Center for Teaching

Multiple-choice questions are commonly used in language assessment due to the accuracy of the results, though MCQs usually require recall of factual information (Granville et al, 2004; Simkin & Kuechler, 2005; Palmer & Devitt, 2007). Providing the students with the correct answer can result in one-right answer thinking. This approach prevents the students from demonstrating their higher-order thinking skills and leads to surface learning (Mullen & Schultz, 2012). Consequently, when the plausibility of the distractors is low and the correct answer is easy to guess, the test takers tend to discard those answers by using lower-level reasoning skills (Rupp, et al, 2006; Tychonievich, 2012; Kerkman & Johnson, 2014). The negative effect of MC testing can be the creation of false beliefs outside the classroom as this format lacks authenticity (Roedeger & Marsh, 2005). These false beliefs can include the approximation of language models which may appear for the learners to be correct, though in reality the correct forms are never sufficiently embedded. However, some researchers believe that it is possible to create MC

items that measure higher thinking skills (Hancock, 1994; Budiyono, 2018; Simkin & Kuechler, 2005), but it can be very time-consuming and require skilled test writers to create valid distractors.

Another drawback of the MC format is the cueing effect that can hint the examinees towards the correct answer of the question. It gives an advantage over the constructed-response format of getting approximately 11% higher scores on the test (Granville et al., 2004). Test-wiseness is one of the strategies widely used among students. This strategy helps test takers to eliminate the alternative options that are poorly constructed or can be considered as distractors because of grammatical impreciseness, option similarity, or item giveaway (Cheng et al., 2007; Cohen 2012; Masoumi & Sadeghi, 2020). All these factors help the students to recognize the correct answer and get a point without general understanding and even memorization of the concepts. As a consequence, the learners who lack sufficient knowledge can successfully select one of the alternative answers (Weimer, 2015; McKenna, 2019). According to the study done at the Western Sydney University, there is always a probability that 5% of the students can pass the minimum required score without any knowledge about the subject (Ibbett & Wheldon, 2016). On the other hand, the skilled test-takers can easily recognize the key answer by simply examining the wording of the options and receive a high grade (Tychonievich, 2012; Funk & Dickson, 2011). This can contaminate the scores and result in low accuracy of the MC test format.

Unlike MC items, which facilitate the respondents with options and provide an opportunity for guessing or elimination of distractors, constructed-response items such as open-ended questions and short-answer questions, require the students to express their thoughts by themselves (Srivastava et al, 2004; Tychonievich, 2012). SA questions measure a deeper understanding of the material and due to the absence of the guessing factor, these items allow the

students to construct their answers (Parmenter, 2009; Weimer, 2015; Budiyono, 2018; McKenna, 2019). SAQ items let the teachers examine students' higher-order thinking skills. In the case of SA format, learners need to apply, synthesize and evaluate the knowledge in order to develop a meaningful response (Tankersley, 2007). This task-focused approach enables the test takers to face authentic situations and replicates their knowledge more accurately. Students generally spend more time preparing for SA format tests and the questions require long-term retention (Pinckard et al., 2009; Sam et al., 2018). Thus, all these features make this item format more appealing for assessment.

2.4. The Construction of MCQ and SAQ Items

The construction of good multiple-choice test items can be very challenging. In order to produce a valid question, several test writers should be included in the process who have considerable knowledge and skills to exclude possible flaws (Granville et al, 2004; Azevedo et al., 2010; Budiyono, 2018). It is very time-consuming to create questions that assess not only simple recall and memorization of the facts, but also the higher-order thinking skills of the students (Simkin & Kuechler, 2005). From the other perspective, the number of MC questions is directly proportional to the reliability of the test. For these reasons, researchers argue about the number of options included in the question, since they affect not only item difficulty but also the level of reliability. According to several studies minor difference was found between three, four and five option multiple-choice test results (Baghaei & Amrahi, 2011; Dehnad et al, 2014). Moreover, a smaller number of options resulted in higher reliability as the test constructors used less incredible and flawed distractors.

On the contrary, it is relatively easy to create short answer questions (Rademakers et al., 2005; Weimer 2015). They can provide reliable results on the students' progress, and the time and cost of marking the responses are very practical for the limited number of test-takers. The teachers who have difficulties constructing MC items can benefit from short-answer questions since they can better suit the course objectives.

2.5. The Scoring of MCQ and SAQ Items

The MC questions are less time-consuming and effortless for assessment. They can be marked manually by the teacher with the help of an answer key or can be scored automatically (Granville et al, 2004; Weimer, 2015; Budiyono, 2018). Several scoring systems can be used when evaluating the MCQ item format. The most common scoring rule used with this format is *Number Correct* scoring which adds to the total score of each correct answer but does not subtract points for the wrong responses. The disadvantage of this scoring method is affected by the guessing factor of the MC items which restrains to fully provide the total overview of language knowledge (Kastner & Stangla, 2011). In comparison with the MC scoring system, if the rater does not reward the students for partial knowledge in SAQ items, then the evaluation systems of both formats can be considered similar in the case of Number Correct scoring.

The alternative for Number Correct scoring is *Negative marking* of MC questions. This scoring method penalizes the respondents for each incorrect answer to eliminate the guessing effect (Campbell, 2015). Researchers think that the negative effect of guessing is superior to negative marking. The standard formula for this making method is 1/(n-1), where n represents the number of options. As an illustration to this case, in McKenna's study (2019) when students received a 0.33-point penalty for each wrong answer in MC format, they gained slightly lower

grades in constructed-response (CR) item tests in comparison to MCQ format. As a result, the number of students passing the minimum grade was almost equal in multiple-choice and constructed-response test format types. On the other hand, in case of Number Correct scoring, the number of students getting over the pass mark was much higher than in CR format.

The scoring of SA questions is usually done manually as the student responses can differ from the correct answer in structure or formulation of thoughts. The evaluation of this format can be expensive, as multiple raters are required for the high reliability of the scores (Budiyono, 2018; Yaneva et al., 2020). Depending on the item type, questions can be scored by a rubric or a prescored sample paper called *anchor paper* which helps to identify the allowable range of answers for each score (Tankersley, 2007). The anchor paper guides the multiple raters to establish a unified grade for the answer, increasing the reliability and objectivity of the score. Therefore, according to the criteria established in the rubrics, the students can gain partial or full credits depending on the accuracy of their response.

To increase the reliability of results and overcome the possible grading challenges regarding SA question scoring, it is preferable to incorporate a grading rubric including the reference answer and all possible variations (Marvaniya et al., 2018). The rubrics incorporate the answers in different grade categories, and cluster students' answers according to provided categories. In order to overlap each answer with the provided reference answer, several techniques are applied, taking into account lexical and grammatical features. The grading of SAQs can be very challenging since it is impossible to generalize to all correct answers and moreover, sometimes credits are provided to partially correct answers (Marvaniya et al., 2018). The scoring rubrics create a model that helps to effectively cluster and rank students' responses

according to different variations in their answers. As a consequence, the grading of SAQs with the help of rubrics results in the increase of objectivity and reliability of the scores.

Since the SAQ items are thought to be harder than MC questions, test creators usually place them after the multiple-choice items. The reason for this arrangement is that students need more concentration and put more effort while answering constructed-response items. As the answer derives from their knowledge and thoughts, it is important to improve students' skills to answer this type of questions (Tankersley, 2007). The reason is that the quality of their response impacts the number of points they get for the question and can change the total score. In the study conducted by Funk and Dickson (2011) the test performance of 50 college students on multiple-choice and short-answer question formats was compared. As a result of three tests conducted with MCQ and SAQ items, the students showed better performance on the MCQ format. In cases, where the test-takers were unable to answer the SAQ items, they were able to provide answers to similar MC questions. Thus, the authors suggest that the multiple-choice items require recognition of correct answers, whereas in SAQ items the learners need to understand and interpret the information.

2.6. Validity and Reliability of Item Formats

Validity and reliability are two important features to be considered when conducting a language test. Both are dimensions of psychometrics that measure the knowledge and abilities of the students (Faremi, 2016). The validity of the test item represents its ability to accurately evaluate what it is supposed to measure. On the other hand, the reliability of the test shows the consistency of the received scores.

The format of the test can affect its validity and reliability. Therefore, test creators pay attention to the item types included in the assessment. Productive assessment items have the ability to discriminate between the poor and the stronger students (Granville et al, 2004). Some researchers claim that MCQ items negatively affect the reliability of the assessment tool and can lead to biased results (Rademakers et al., 2005; Moore, 2014). The effort of creating valid and reliable MC tests can be very time-consuming and challenging for the teachers. Hence, it will be more reasonable to spend that time on writing SA items if assessing a limited number of students (McKenna, 2019). This is based on the fact that the SA questions are relatively short and may require filling in a word or a couple of sentences, which can be more reliable and effortless to score.

2.7. Students' Performance on the Test Format

Different studies have investigated the impact of the test format on students' performance. The majority of researchers suggest that there is a considerable difference between the test formats, and both multiple-choice and short-answer test items influence the test takers' final scores. According to the majority of the studies, students tend to perform better on MC format tests rather than on the SA format (Bleske-Rechek et al., 2007; In'nami & Koizumi, 2009; Currie & Chiramanee, 2010; Sangwin & Jones, 2017; Faremi, 2016; Budiyono, 2018; Liao, 2018; Masoumi & Sadeghi, 2020). For instance, students who have difficulties answering SA questions are able to respond to the same type of questions in MC format (Funk & Dickson, 2011). Meanwhile, there are only a few cases when students outperformed in constructed-response tests (Bleske-Rechek et al., 2007).

Masoumi and Sadeghi (2020) have conducted a study to investigate the role of gender in multiple-choice and constructed-response vocabulary assessment. The participants of the study were 243 Pre-Intermediate high-school EFL students (132 males and 11 females) in Iran. The gender analysis of the participants suggested that females perform better in the SA format test than males (Masoumi & Sadeghi, 2020). Thus, while developing a language tests several factors such as performance on the format or gender performance, should be taken into account.

There are a number of factors that can result in poor performance on the SA test format. One of them is test takers' anxiety, which can be a reason for the insufficient time that is provided to answer the constructed-response items (Hussey et al., 2010; Budiyono, 2018). Students require some time to get familiar with the item format, comprehend the question, and production of their answers. Conversely, in MC format learners are provided with multiple options, and those additional cues help them to answer the questions (Cheng & Gao, 2002; Famularo, 2007; Mozaffari et al., 2017).

Another reason seems to be that MC and SA items measure different language skills which results in the item difficulty, making the constructed-response items more challenging for the students to respond (Currie & Chiramanee, 2010; Cohen 2012). While MC questions ask for factual detail information, SA questions demand analysis and evaluation of general concepts (Simkin & Kuechler, 2005). Due to this, students who are trained on SA item type perform better on MCQ format (Funk & Dickson, 2011). However, it should be considered that not all MC questions are easier in comparison to SA questions.

Finally, the lack of experience in the SAQ format can highly impact performance during the exam. Students experience difficulties when trying to summarize or paraphrase main ideas or concepts (Budiyono, 2018). Although in the classroom they are frequently interacting with SA

item format, it becomes a challenge as a test taker to respond to these kinds of questions during language assessment. The reason can be students' perception of the test format (Liao, 2018). This is a cause of learners' consideration of MC questions to be easier than SA questions and their performance might be influenced by their beliefs.

It can be concluded that the instructors have poor benefit from just including either MC or SA question formats in language assessment (Granville et al, 2004; Bleske-Rechek et al., 2007; Mozaffari et al., 2017; Budiyono, 2018; Masoumi & Sadeghi, 2020). The combination of the two formats will allow language tests to evaluate all the essential skills matching course objectives. Both multiple-choice and short-answer question items have their advantages and disadvantages. Thus, test administrators need to spend much time on the construction of more high-quality MCQ items and SAQ scoring rubrics to get valid and reliable results on students' language progress.

To sum up, it is interesting to mention that similar pattens have been noted with the studies done in other scientific fields. For instance, in studies conducted in the medical field the findings suggest that in comparison with the multiple-choice questions, the SAQs provide more validity to the test items and incorporate the authenticity of the exams (Behizadeh & Egelhard, 2014; Sam et al., 2019). In the technology course, the students perform better on MCQ items rather than on constructed response questions, since the cueing effect of the MC items helps the test takers to easily recognize the distractors (Grunert et al., 2013). There are a number of other studies from psychology, philosophy, computer sciences, and other fields, that compared students' test performance on MCQ and SAQ formats, and the results of those studies and their explanations are closely related to the findings in linguistics.

CHAPTER THREE: METHODOLOGY

3.1. Context

This quasi-experimental research was intended to find out answers to the following three research questions:

1. Does student test performance on MCQ and SAQ formats change depending on their proficiency?
2. How does the systematic treatment with the SAQ format affect students' test performance?
3. What are students' perceptions towards MC and SA question formats?

The study was conducted in an afterschool English program in Yerevan, Armenia. The program aims to provide knowledge to students with the communicative language teaching approach. The purpose of this teaching method is to develop the ability of the learners to apply their language knowledge in a real-life context. Thus, all the materials used for teaching and assessment purposes need to have high authenticity. The coursebook used for the program is "English in Mind" published by Cambridge University Press in 2011, which mainly focuses on the enhancement of grammar and vocabulary knowledge through creative listening and reading topics. Besides the coursebook, additional materials are used to engage the students in communicative activities and integrate them into genuine language practice.

The classes are conducted twice a week and have a duration of 90 minutes. Considering the emergency state of the country as a result of COVID-19 pandemics, the classes are held online via Zoom online video conferencing platform. One of the main advantages of this online platform is that it allows the teachers to increase the student-student interaction by putting them in breakout rooms. It enables the teachers to create more student-centered lessons which is one of

the core aspects of CLT. Apart from Zoom, other online platforms are implemented in online classes: such as Edmodo, Google Docs, Viber, WhatsApp, and others. These online tools are mostly used for homework and supplementary assignments. As a part of the course, the students have to take two achievement tests once in four weeks which are provided to students via Google Forms or Google Docs. The tests are constructed based on the materials included in the syllabus and are divided into four main sections: listening, reading, language use (grammar and vocabulary), and writing.

The participants of the study were two groups of EFL learners from the afterschool English program. Both groups had a Low Intermediate proficiency level of English. In general, 19 students (14 male, 5 female) participated in the research with the age range from 12 to 16. During the study, the groups were participating as treatment and comparison groups.

3.2. Ethical Considerations

Before the beginning of the study, the researcher obtained an IRB certificate which permitted to work with human subjects as a part of the research. All the ethical considerations were followed according to standards provided in human subject certification guidelines. Both groups were provided with information about the design and purpose of the research. The subjects were informed that the received data is going to remain confidential and anonymous.

Since the research studied the assessment of students, the confidentiality of the results was a critical factor. Throughout the research, all the answers of the participants were uploaded to Moodle learning platform where each student had their login and password to enter their accounts. Once the students entered the accounts, they changed the password which made it impossible for others to enter their personal pages, except for the moderator of the page and the group teachers. No individual data were reported, as the participants received separate feedback

during the treatments and final tests, and their scores were kept confidential as well. As for the interviews, the recordings were saved in the researcher's personal computer and were never published or sent over the internet to any person related or not related to the study. Confidentiality was maintained also while scripting the interviews. In interview scripts, the names of the subjects were coded and no information was provided about the groups or scores of the participants.

The personal information of the research subjects was kept anonymous during the online survey. This condition was met in order to get more reliable information regarding the perceptions of the participants and reveal the actual advantages and drawbacks of the study. The survey did not require the participants to provide any data that could in any way identify their personality.

3.3. Data Collection Instruments

During the research, several instruments were used to examine each aspect of the students' performance and the impact of item format (see Figure 2). Firstly, to establish the students' primary test performance on MC and SA question formats and later to make comparisons on their progress regarding the test performance, the midterm test was administered to both treatment and comparison groups. It was the same test used by the program to assess students' progress after four weeks of instruction and consisted of listening, reading, language use (grammar and vocabulary), and writing sections (see Appendix A). The duration of the test was 100 minutes and the total amount of points was 40 (see Table 1). The test involved both SAQ and MCQ formats, however, they were not equally dispersed through the sections. For instance, the grammar and listening sections included a number of SAQ tasks, yet the reading and vocabulary sections consisted of mainly MCQ and dichotomy items.

	Listening	Reading	Language Use	Writing
Number of Tasks	2	1	5	1
Points	5	5	15	15
Item Formats	multiple-choice, short-answer	true/false	multiple-choice, short-answer, matching, ordering	essay

Table 1: *The structure of the midterm test*

The results of the midterm test were used to determine the initial proficiency of both treatment and comparison groups, regarding the scores obtained by the students on MCQ and SAQ tasks (Figure 2). The correlation of the midterm and the final test results helped to establish students' progress by the end of the course.

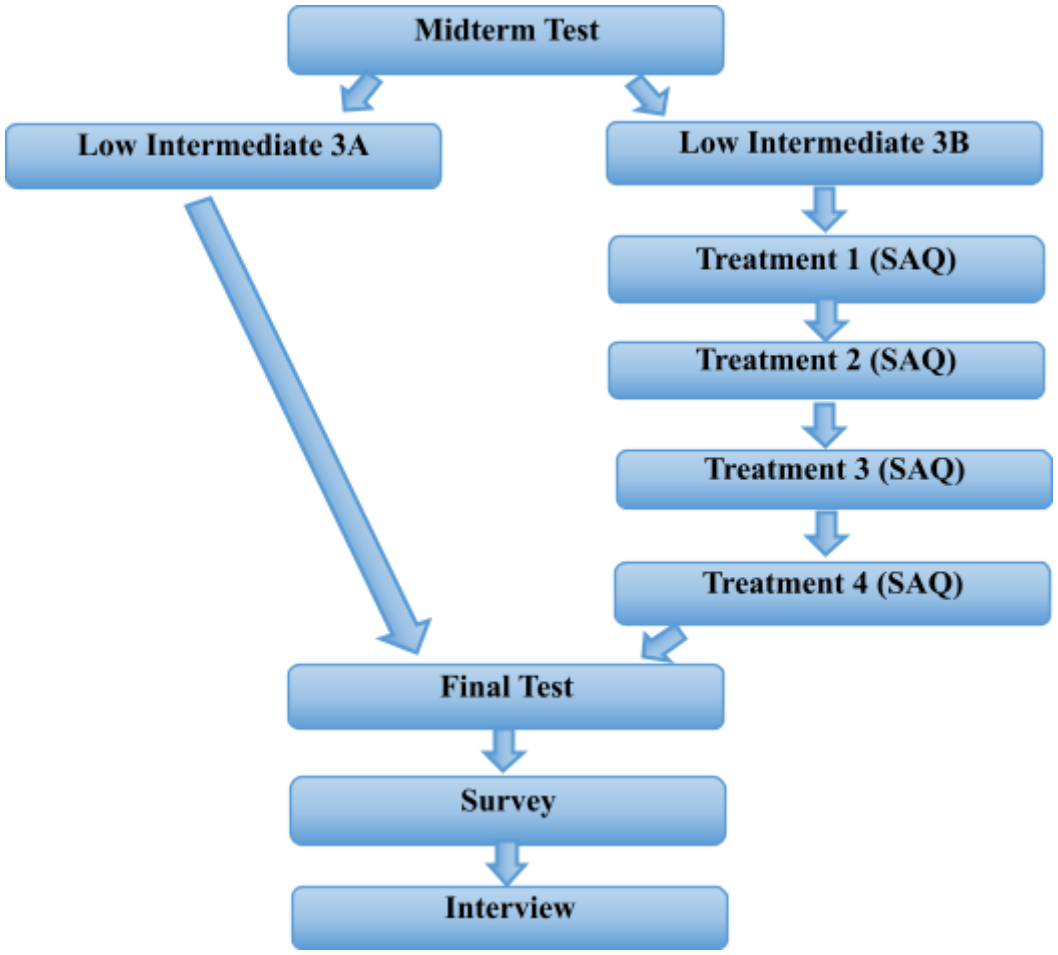


Figure 2: *Instruments used to investigate the influence of question formats on students' performance*

As it can be seen from Figure 2 the midterm test was followed by four treatment tasks within the following four weeks. Most of the assignments consisted of reading, listening, grammar, and vocabulary sections (see Table 2) (see Appendices B, C, D, and E). The tasks included in four treatments were designed based on the coursebook units four to six. All the items were in short-answer question format where the answers varied from one word to a sentence. The tasks included in the treatments required not only factual details but also an interpretation of the concepts, self-judgment, and analysis of the material. After each treatment, detailed feedback was provided to the students. The treatments were implemented to investigate the potential benefits of SAQ item format on students' test performance.

Table 2: *The construction of treatments*

		Grammar	Vocabulary	Reading	Listening
Treatment 1	**Number of Tasks**	2	1	X	X
	Points	5	5	X	X
	Item Format	short-answer	short-answer	X	X
Treatment 2	**Number of Tasks**	1	1	2	2
	Points	5	5	5	5
	Item Format	short-answer	short-answer	information transfer	information transfer, SA-cloze
Tr	**Number of Tasks**	1	1	1	1

eatment 3	**Points**	5	5	5	5
	Item Format	short-answer	SA-cloze	short-answer	short-answer
Treatment 4	**Number of Tasks**	1	1	1	1
	Points	5	5	5	5
	Item Format	SA-cloze	SA-cloze	short-answer	SA-cloze

At the end of the course, a final test was given to both treatment and comparison 3B groups (see Appendix F). This test intended to establish the difference in test performance of the treatment and comparison groups after four weeks of treatments and tried to reveal the influence of the constant practice of SAQ tasks. The test included four sections: listening, reading, language use (grammar and vocabulary), and writing (see Table 3). The first three sections consisted of an equal number of MCQ and SAQ item formats. Each MC question was followed by an SA question format. The duration of the test was 120 minutes. The total number of points was 40, where the maximum number of points in the first three sections for MCQ was 12.5 points, and the maximum number of points for SAQ was 12.5 points. Each listening task, reading text, vocabulary, and grammar exercise was followed by SA and MC question formats on the same topic (e.g., 5 MCQ items and 5 SAQ items checking the knowledge of the reported speech). The SA questions included fill-in a word, information transfer, sentence completion,

true/false statements justification exercises, and the MC format items included questions with three, or four options on the same topic.

Table 3: *The construction of the final test*

		Listening	Reading	Language Use
Number of Tasks	MCQ	2	1	6
	SAQ	2	1	6
Points	MCQ	2.5	2.5	7.5
	SAQ	2.5	2.5	7.5
Item format	MCQ	multiple-choice,	multiple-choice	multiple-choice, select the word
	SAQ	SA-cloze	short-answer	short-answer, SA-cloze

After finishing the final test, the students filled out a survey which helped to investigate students' perceptions towards MCQ and SAQ item formats. The survey aimed to examine the MC and SA question formats from the students' perspectives. In general, there were 14 questions (7 for each format) that were aimed to discover the preference of SAQ and MCQ item formats from different aspects. The statements helped to display the participants' perceptions regarding the difficulty of question types, how the SAQ and MCQ item formats assess their level of understanding of the course material, and if they were able to notice their language gaps with SA and MC question formats.

At the end of the research, an interview was conducted with four students from both treatment and comparison groups to find out the reasons for the preference of any question format. The interviewees were selected according to their scores obtained in the final test. Based on their performance on MCQ and SAQ formats one low proficiency, two mid proficiency, and one high proficiency students from two groups were chosen to participate in the interview. The

interview questions helped to discover in more detail the students' attitudes about the advantages and disadvantages of item formats and uncovered the possible difficulties that arose during the language test.

3.4. Data Collection Procedure

Two groups of Low Intermediate English proficiency were chosen at the beginning of the study. The convenience sampling method was implied to select the learners, as the number of accessible groups within the same proficiency level was limited. In the afterschool English program, the groups are reorganized after the completion of each level according to the learners' final and midterm test results of the previous level. Therefore, the students had almost the same level of proficiency. Both groups were using the same materials and had the same syllabus for the course.

The data collection process included the gathering of quantitative and qualitative data, received in different stages of the study. The first step was the introduction of Moodle learning platform and registration of the research subjects. For the assessment of participants, Moodle learning platform was chosen as it enables to analyze the data more thoroughly at an individual level for each of the students. Compared to Google Docs or Google Forms, Moodle allows to examine the responses in all details and has a variety of question formats to carefully assess the language knowledge. The platform establishes a detailed description of assessment results including from the amount of time the students spend to answer the questions, to statistical evaluation of group scores.

A week before the midterm test accounts were created for the teachers and students of both groups. The participants from both groups were enrolled in corresponding groups on Moodle. The logins and passwords were sent to all participants individually. The students were

introduced to Moodle in advance, to avoid any technical issues. They were presented to all formats included in their test, and detailed explanations were provided on how to write a test in Moodle by using different devices. Short tutorials were sent to teachers on how to grade and provide feedback to students in Moodle.

Overall, nineteen students from two groups took the midterm test. All the participants were using different devices such as mobile phones, computers, and tablets; however, it did not cause any issues during the test. Only one student had difficulties with entering the Moodle account, but the problem was solved within a short period. The test was administered by the teachers of the groups. During the test, most of the students were connected to an online Zoom conference and turned on their cameras, which allowed to minimize the possibility of cheating. Since Moodle had some technical issues with the assessment of short-answer questions, these items were double-checked and regraded by the group teachers.

Based on the results of the midterm test, the treatment and comparison groups were chosen. Initially, the group with a lower average result was planned to do the treatments during the following weeks. However, the group with a lower mean proficiency fell behind the syllabus, and a decision was made to do the treatments with the group with a higher mean proficiency level in order to not reduce the number of treatments.

The treatment tasks were created and placed in Moodle once a week after the completion of each chapter. In general, there was 60-70% participation of students in all four treatments. Treatment 1 and treatment 2 (see Appendices B and C) had a higher level of participation than treatment 3 and treatment 4 (see Appendices D and E). Overall, 7 out of 10 students did treatments 1 and 2, and 6 students did treatments 3 and 4. The answers of treatments were checked manually and the answers were compared to the sample correct answer. No rubric was

used for the assessment of four treatments. The main criterion in the reading and listening sections was the coherence of the answers and in the grammar section the correct use of grammatical forms (e.g., no points were taken for the incorrect use of articles when assessing conditionals).

The treatment and comparison groups took the same final test. During the final test, the students from both groups were connected to the same Zoom conference and most of them were with their cameras turned on, except the students who were taking the test on their mobile devices. To increase the objectivity of short-answer questions included in the test, they were checked by two raters. Since in SA format the answers can differ from the sample correct answer in structure and meaning, items were graded by two independent raters to reach high reliability of scores.

After the treatments and the final test, several students from the treatment and comparison groups were excluded from the research. The main reason for the attrition of the participants of the treatment group was poor participation during the four weeks of treatments (participated in one treatment or did not participate at all). As for the comparison group, the results of three students were not analyzed as part of the research. One of the students skipped answering all of the SAQ questions and provided random answers in MCQ format. The other two students from the comparison group were excluded for cheating since they have given completely identical answers on SAQs. The possibility for this to happen was quite small taking into account the fact that the students were required to make their own judgments. For these reasons, their results were eliminated from the later data analysis.

The final test was followed by a questionnaire. The participants filled out the survey about their perception of the MCQ and SAQ item formats included in their test. Eighteen

students out of nineteen from treatment and comparison groups answered survey questions. The students were provided with several statements about question formats included in the final test which they needed to evaluate with the help of the Likert Scale. For each answer, the participants gave from 1 to 5 points to each statement (1 point for completely disagree and 5 points for completely agree). These statements helped to find out the perceptions of students' towards SAQ and MCQ formats.

After getting the final test results of all the participants, an interview was conducted with several students from both groups to find out the reasons for the preference of any question format. The interviewees were selected by purposive sampling, relying on their scores of final tests. According to their test performance, two students were selected from both treatment and comparison groups. The interview was with each student individually in Zoom, and the answers of the students were recorded and scripted.

3.5. Data Analysis

At first, for the objectivity of scores on the SAQ items in the final test, a second rater was included in the data analysis. The raters graded the short-answer questions relying on the sample correct answer provided in advance. The final score of each SAQ item was defined by the average score given by rater 1 and rater 2. Inter-rater reliability was calculated, to measure the level of agreement between two independent raters. Besides, the mean percentage difference between two raters was measured for the tasks where different points were given by two raters. These statistical tools helped to define the degree of accuracy and reliability of the results.

To find out the impact of students' proficiency on their scores of MCQ and SAQ item formats several instruments were implemented. Firstly, to reveal the difference between the

scores obtained on the multiple-choice and short-answer question formats of the final test, the students' performance on these item types was examined individually. The total amount of points received for each format was compared to the student's final score and the ratio of SAQ and MCQ items was established.

Based on the results of the final test, the participants of treatment and comparison groups were divided into three subgroups relying on the total scores received in listening, reading, and language use sections. Although the writing section was included in the final test, the points for this part of the test were not taken into account during the data analysis. According to the final scores, three groups were created: low proficiency (0-12.99 points), mid proficiency (13-17.99 points), and high proficiency (18-25 points). For each of these three groups, the mean points received for SAQ and MCQ item formats were correlated to the proficiency group's mean final score. Pearson Correlation Coefficient (Pearson's r) was used to calculate the correlations between these variables (points received for MCQ and SAQ items and the total score). The results displayed how the scores received from MCQ and SAQ items affect students' test performance, and also how the relation between the variables changes with the increase of the total score.

In addition, the scores on MCQ and SAQ format obtained in the final test were correlated for each proficiency group (low, mid, and high). The analysis helped to find out the influence of the question formats on students' test performance, and if the correlation between the scores of two formats changes with the increase of learners' proficiency. Also, the mean difference of the scores on MCQ and SAQ formats was compared for each proficiency group to establish how the impact of MCQ and SAQ formats differs for each proficiency group (low, mid and high proficiency).

The initial proficiency of treatment and comparison groups was established by the calculation and comparison of mean scores of two groups received on the midterm test. In addition, to display the difference in test performance of the groups, the distribution of grades received on the midterm test for the treatment and comparison groups was compared. The analysis helped to reveal the difference in proficiency between the treatment and comparison groups. The mean of the scores on the final test of the groups and the distribution of students' grades was also calculated for the final test. The results helped to display the difference between the performance of the two groups on the final test and to understand whether the existing gap between the treatment and comparison groups changed after four weeks of treatments.

The difference in the performance of the treatment and comparison groups in MCQ and SAQ sections was analyzed in both midterm and final tests. The mean scores of the students on each question format were calculated for both groups. The difference between the mean scores of treatment and comparison groups on MCQ and SAQ formats helped to show the difference between the performance of the two groups on each question type. The comparison helped to explore the changes in group performance on MCQ and SAQ items in midterm and final tests and to reveal the effect of treatments with SAQ items on the final results of the treatment group.

In order to measure the influence of the treatments, and to discover whether the constant practice of SAQ items helped the treatment group to perform better during the final test, additional calculations were done. The analysis of four treatment tasks included only statistical calculations of the mean scores for each treatment. The mean scores were compared to see the changes in the progress of the treatment group during four weeks. The scores were also compared for four different sections included in the treatments. The mean scores of the students

of the treatment group for reading, listening, grammar, and vocabulary tasks were analyzed separately. This helped to discover the sections where the learners demonstrated greater progress.

The mean scores for SAQ and MCQ item types of treatment and comparison groups were also compared for reading, listening, grammar, and vocabulary sections in the final test. This detailed analysis tried to establish if there was any change in groups' performance on both question formats regarding each section of the test. The differences of mean scores on MC and SA items between the treatment and comparison groups were calculated. This helped to understand which sections caused the differences in the performance of two groups in MCQ and SAQ question formats.

Since the students mentioned their groups in the survey, the results for the treatment and comparison groups were separated and compared. The answers for each question were compared between the treatment and comparison groups, to see whether the treatments resulted in the difference in perception towards MC and SA question formats. The interviewees' answers were compared and linked to the findings of the survey. The qualitative data received from the interviews with the participants helped to reveal the reasons for preference of each format.

CHAPTER FOUR: RESULTS

4.1. Research question 1: Does student test performance on MCQ and SAQ formats change depending on their proficiency?

To answer the first research question, the results of the final test of the treatment and comparison groups were compared. Overall, scores of 13 students from both groups (treatment group – 7 students, comparison group - 6) were included in the analysis. The investigation of the results established the relationship between students' performance on MC and SA question formats and their proficiency (low, mid, and high) regarding the scores received on the final test.

To increase the objectivity and reliability of the SAQ format, the second-rater was included in the assessment of the final test. The inter-rater reliability was calculated between the two raters. They were provided with a sample correct answer to grade the SAQ items, however, in several cases, the scores of the raters varied (see Appendices H and I). Table 4 presents the inter-rater reliability for the treatment and comparison groups on the final test. There was found high inter-rater reliability for both groups: raters for the treatment group had 86.76% agreement of scores, and for the comparison group the agreement between two raters was 89.22%. In addition, the mean percentage difference was calculated for the grade mismatch cases between two raters (when the raters provided different grades on the same task), in order to examine their impact on students' final scores. The results of the mean percentage difference for the treatment

and comparison groups were accordingly 19.30% and 13.50%. This showed that even if the scores given by two raters were not always similar, the difference was quite small to negatively influence the results.

Table 4: *Inter-rater reliability of final test results for treatment and comparison groups*

	Treatment Group	Comparison Group
Inter-rater Reliability	86.76%	89.22%
Mean Percentage Difference of mismatch cases between two raters	19.30%	13.50%

The mean scores on the MCQ and SAQ formats of the final score were calculated for all the participants in the treatment and comparison groups. The findings helped to display the change between the scores of two formats regarding students' proficiency (low, mid, and high). Figure 3 represents the ratio of MC and SA question formats in the final score received by the students of treatment and comparison groups. The highest score is 25 points (not including the 15 points from the essay in the writing section), and the scores are displayed in increasing order. Along with the increment of the final test scores, the ratio of scores between the MC and SA question formats registered noticeable changes. For instance, the number of points received for MCQ items was bigger than the number of points for SAQ items for the students who scored less than 13 points (for low proficiency students). As the student's grade increases, the ratio between the scores in MCQ and SAQ formats becomes more balanced (for mid and high proficiency students).

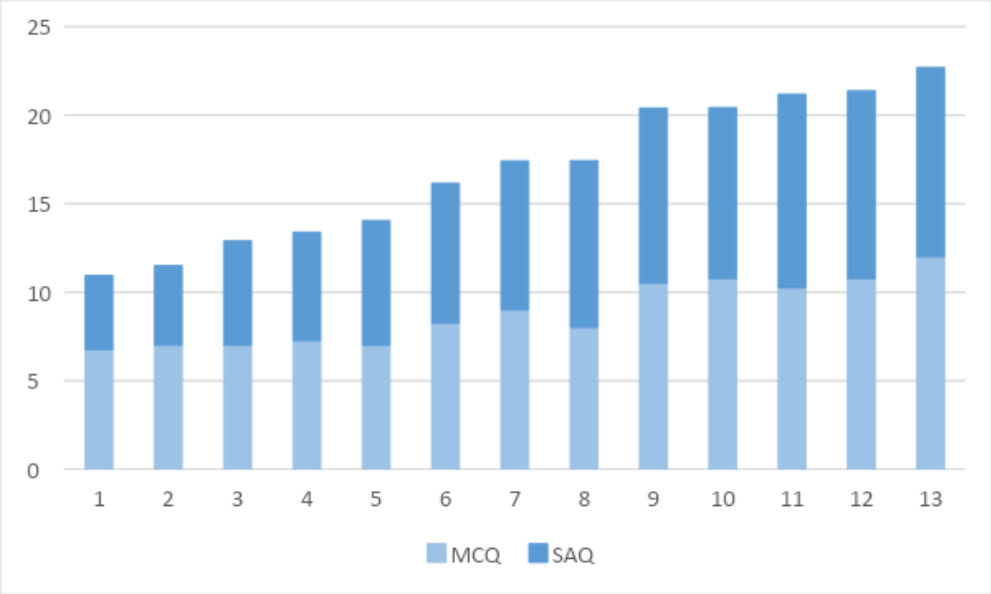


Figure 3: *Ratio of students' scores in MCQ and SAQ formats*

The internal reliability of the final test was measured with the split-half method. It helped to reveal how consistently the MCQ and SAQ item formats were performed within the final test. Since each multiple-choice task was followed by a short-answer task, the correlation between the two halves of the test was calculated by correlating the scores of all the participants received from the MCQ and SAQ formats. Pearson Correlation Coefficient (Pearson's r) was used to relate the two halves of the test, and the analysis showed a high correlation ($r = 0.89$) between the two sections. The results indicate that both MCQ and SAQ parts of the test equally contributed to measure the students' achievements on the final test.

To understand if students' performance can be affected by MCQ or SAQ formats, correlations were drawn among the scores received from MCQ and SAQ items and the total score on the final test. Furthermore, to find out how the interaction between these variables change based on students' proficiency, correlational analysis was done for three proficiency groups. The results were calculated for the students receiving low (0-12.99 points), mid (13-18.99 points), and high (19-25 points) total scores on the final test. According to the

correlational analysis for the low proficiency group, the correlation between the scores of MCQ items and total scores on the final test was r = 0.72 (r > 0.7) (Figure 4). This is a moderate positive correlation which shows that with the increase of the final grade the scores on the MCQ format have a tendency to rise, however, the change is not significant.

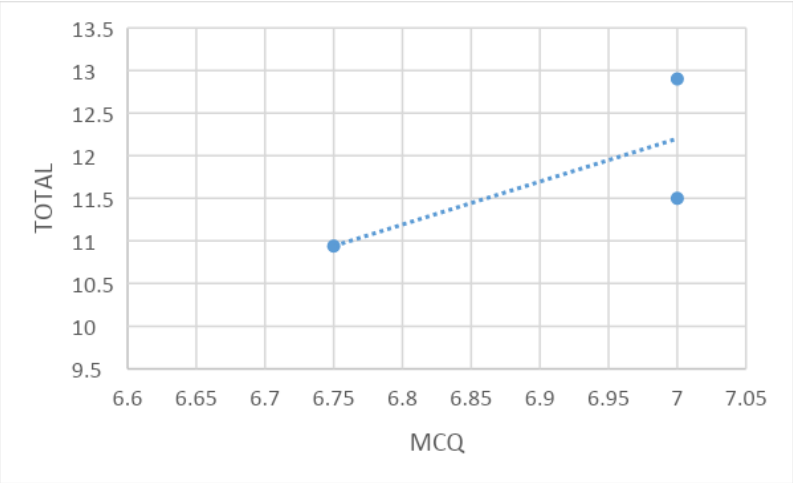


Figure 4: *Pearson's r for MCQ format and total scores of low proficiency students on final test*

The correlation coefficient for the scores on the SAQ items and final score was r = 0.99 (r > 0.9) (Figure 5). The results established a very high positive relation between these two variables. It shows that with the increase of the final score the ratio of correctly answered SAQ questions notably rises.

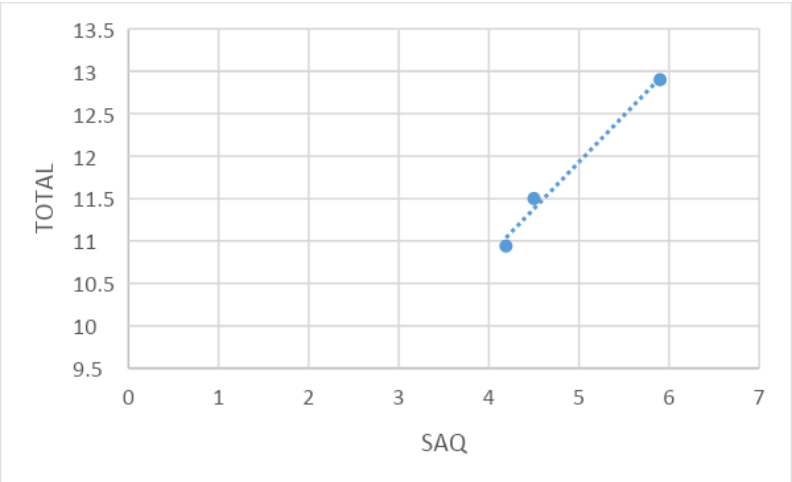


Figure 5: *Pearson's r for SAQ format and total scores of low proficiency students on final test*

The relation between the performance on MCQ and SAQ formats for the low proficiency students showed a moderate positive correlation (r = 0.64) (Figure 6), which reports that the high results on one of the formats generally result in a good performance on the other question type. Additional analysis of the mean performance of low proficiency students on the final test showed that they received higher scores on the MCQ items (55.36%), meanwhile, the mean results on the SAQ format were only 38.96%. There is a 16.40% difference in the results of both formats, which shows once again that the SAQ format has a big influence only on students with low proficiency, and for this reason, the students with lower scores scored lower on SAQ items, unlike MCQ tasks.

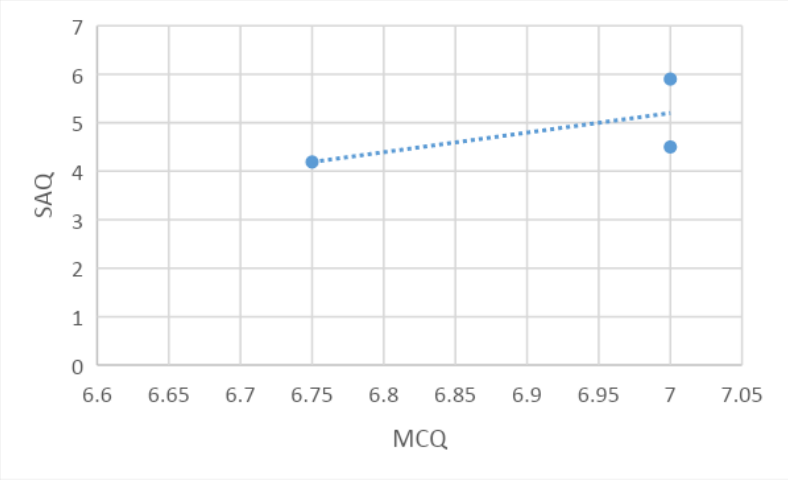


Figure 6: *Pearson's r for scores of low proficiency students on MCQ and SAQ formats on final test*

For mid-proficiency students, the Pearson's correlation between the scores in the MCQ and SAQ items and the final test scores was r = 0.86 and r = 0.94 (Figures 7 and Figure 8). These variables (scores on MCQ and SAQ formats and the total score) were highly correlated (r > 0.8 and r > 0.9), and indicate that the performance on both question formats becomes better with the increase of total grades. It is noticeable that the performance on SAQ format indicates progress.

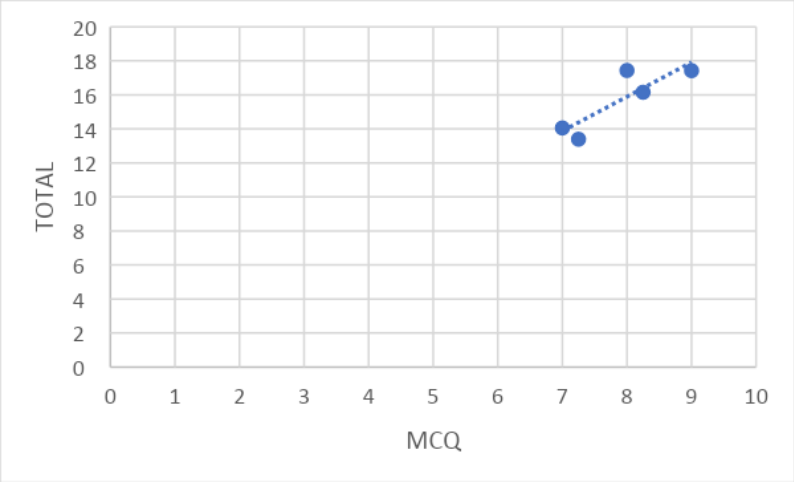


Figure 7: *Pearson's r for MCQ format and total scores of mid proficiency students on final test*

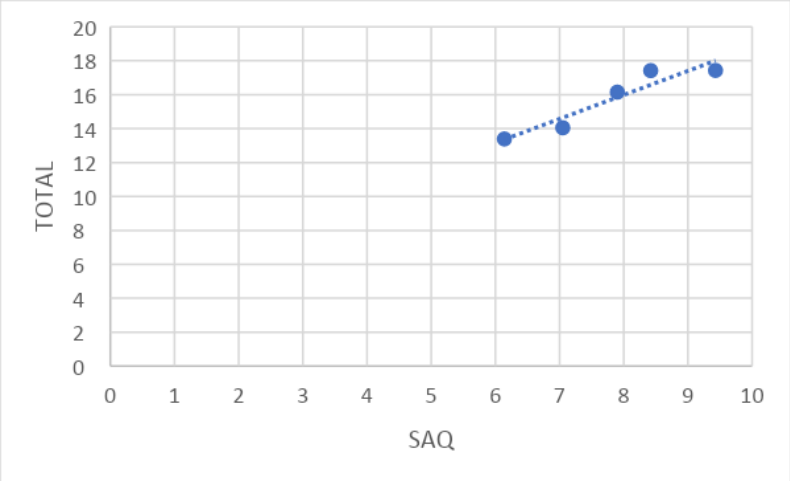


Figure 8: *Pearson's r for SAQ format and total scores of mid proficiency students on final test*

The comparison of the mean scores on the MCQ and SAQ formats of the final test also showed very little difference between the scores on these question types (only 0.9% difference of mean scores). The students performed almost equally on both formats. Even though all of the participants received higher results from the MCQ items, yet the big gap between the scores of the two formats decreased resulting in a moderate average correlation between the scores of MCQ and SAQ items (r = 0.65) (Figure 9).

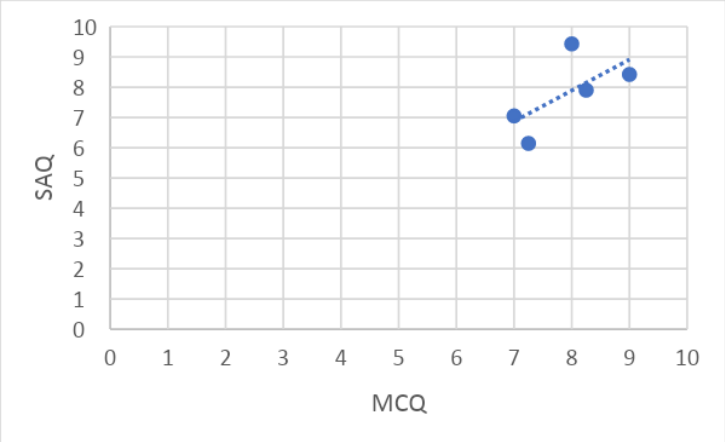


Figure 9: *Pearson's r for scores of mid proficiency students on MCQ and SAQ formats on final test*

As for the high-proficiency students, the correlational analysis established a high positive correlation between the scores on MCQ format of the final test and the final score (r = 0.81) (Figure 10).

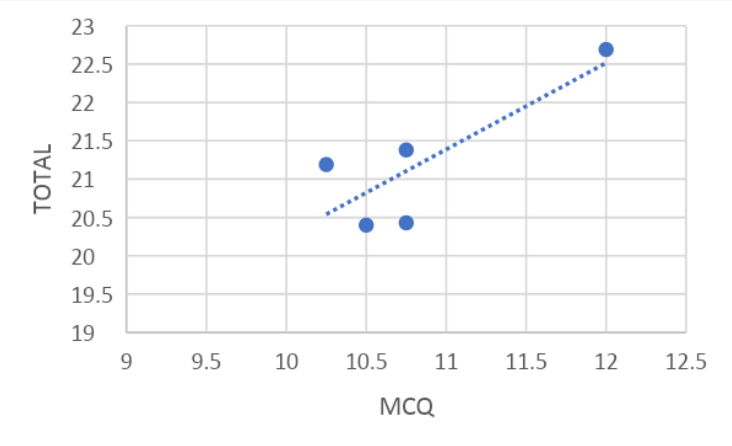


Figure 10: *Pearson's r for MCQ format and total scores of high proficiency students on final test*

We can see from Figure 11 that he scores on the MCQ format raised with the total score. For the SAQ format, the correlation with the total score was positive moderate r = 0.7, which also shows that the students tend to perform better on this question format with the increase in proficiency.

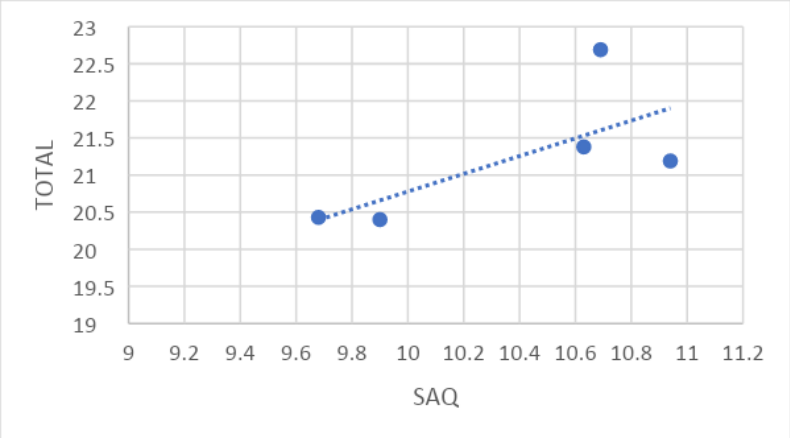


Figure 11: *Pearson's r for SAQ format and total scores of high proficiency students on final test*

However, the analysis revealed almost no correlation (r = 0.16) between the scores received from the MCQ and SAQ formats on the final test (Figure 12). This showed that the results differed a lot and several students even performed better on the SAQ items. As a result, with the increase of learners' proficiency, the performance on both MCQ and SAQ formats not only changed for the better but also in several cases the students overperformed in the SAQ format.

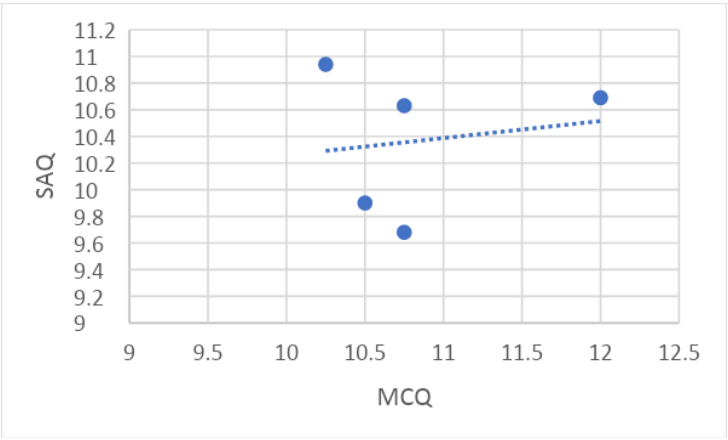


Figure 12: *Pearson's r for scores of high proficiency students on MCQ and SAQ formats on final test*

4.2. Research question 2: How does the systematic treatment with the SAQ format affect students' test performance?

For the analysis of the second research question, the students' performance in the midterm and final tests for treatment and comparison groups were compared. Figures 13 and 14 illustrate the distribution of grades for midterm and final tests. According to the results, there were minor differences in final scores between the treatment and comparison groups for the midterm test. More than half of the participants (66.7%) in the comparison group showed an average performance with final grades between 15 and 19 points out of 25 points. The remaining 33.3% of the students in the comparison group received high grades on the midterm test which were equally distributed between 18 and 25 points. As for the treatment group, 72% of students received high grades (above 19 points), whereas there were only a few students with average performance on the midterm test. On the whole, the results indicated that the treatment group fairly overperformed the comparison group in the midterm test.

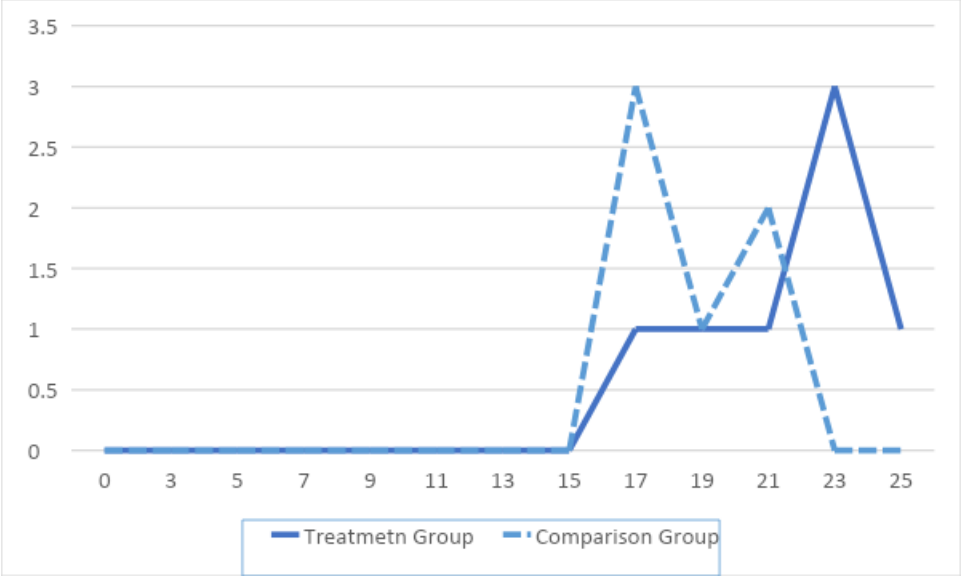


Figure 13: *Distribution of scores in midterm test for treatment and comparison groups*

However, the gap between the groups increased as a result of the final test. As it can be seen in Figure 14, none of the students in the comparison group scored more than 19 points. Half of the participants in the comparison group scored between 13 and 19 points, and the other half of the students showed poor performance (between 10 and 13 points). This resulted in the poor performance of the comparison group on the final test. On the contrary, 72% of the students in the treatment group received more than 20 points out of possible 25 points. In general, the treatment group demonstrated consistent performance in midterm and final tests, whereas the comparison group performed a lot better on the midterm test in comparison with the final test. This indicates a change in performance between the groups after four weeks of treatments.

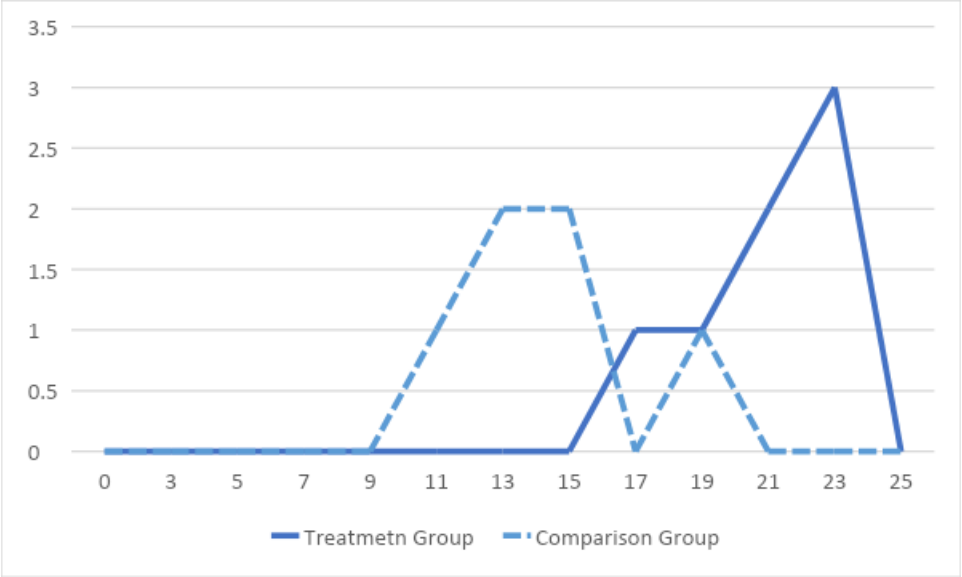


Figure 14: *Distribution of scores in final test for treatment and comparison groups*

Later in the data analysis are included the results of the students of the treatment group who participated in two and more treatment tasks. Figure 15 establishes the differences in performance on final and midterm tests in MC and SA question formats. In the midterm test, there was found only a 14.32% difference in total scores for the treatment and

comparison groups. As a result of the final test, the gap between the total scores of the treatment and comparison groups increased by 12%, resulting in a 26.32% deviation. The evaluation of students' performance on SAQ format, revealed that the difference between the performance of treatment and comparison groups after the final test increased by 6.98% compared to their results in the midterm test (20.56% difference in midterm test and 26.32% difference in final test). As for the MCQ items, there was a considerable change of 16.27% (8.05% in midterm test and 24.32% in final test). The results illustrate that the better performance of the treatment group was not only a result of high scores received in the SAQ format but also there was a huge difference in scores on the MC question format, which was not as significant in the midterm test.

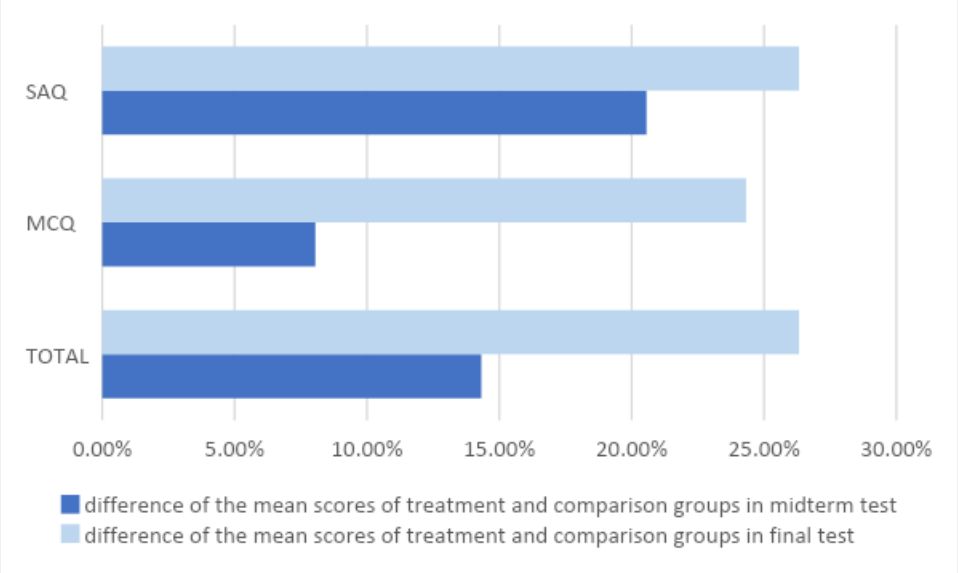


Figure 15: *The differences in test performance of treatment and comparison groups on midterm and final tests*

Further investigations were done to understand the reason for the increased gap in test performance between the treatment and comparison groups. The evaluation of student's performance on treatment tasks showed the learners' progress during four-week practice

(Figure 16). From the beginning of the study, the treatment group was performing quite well with an 80.23% average score on the first treatment tasks. There was a small percentage of decrement in grades for the second treatment, but the reason was the greater number of tasks included in Treatment 2. Overall, there was only a 4,87% increase in grades throughout the participation in four treatments.

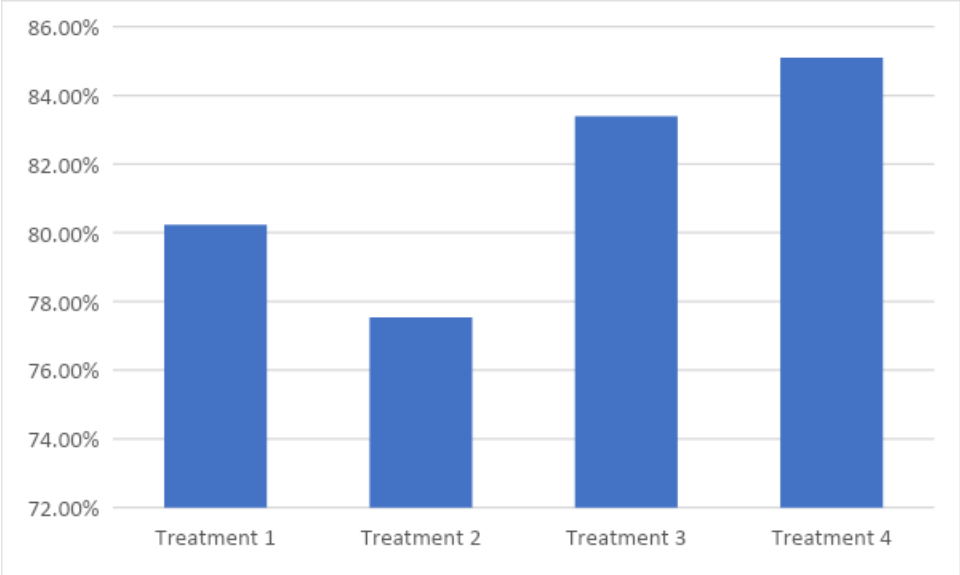


Figure 16: *The performance of treatment group on four treatment tasks*

The learners' performance in listening, reading, grammar, and vocabulary sections is illustrated in Figure 17. In all four treatments, the students received high grades in the grammar section (more than 87%). The first treatment included only grammar and vocabulary tasks and resulted in higher scores compared to the next treatment. In Treatment 2, the weaker performance in reading and listening sections affected the students' scores. However, later on, the grades in listening, reading, and vocabulary sections were within a similar range. In general, the students demonstrated small progress in the scores in all sections from Treatment 1 to Treatment 4.

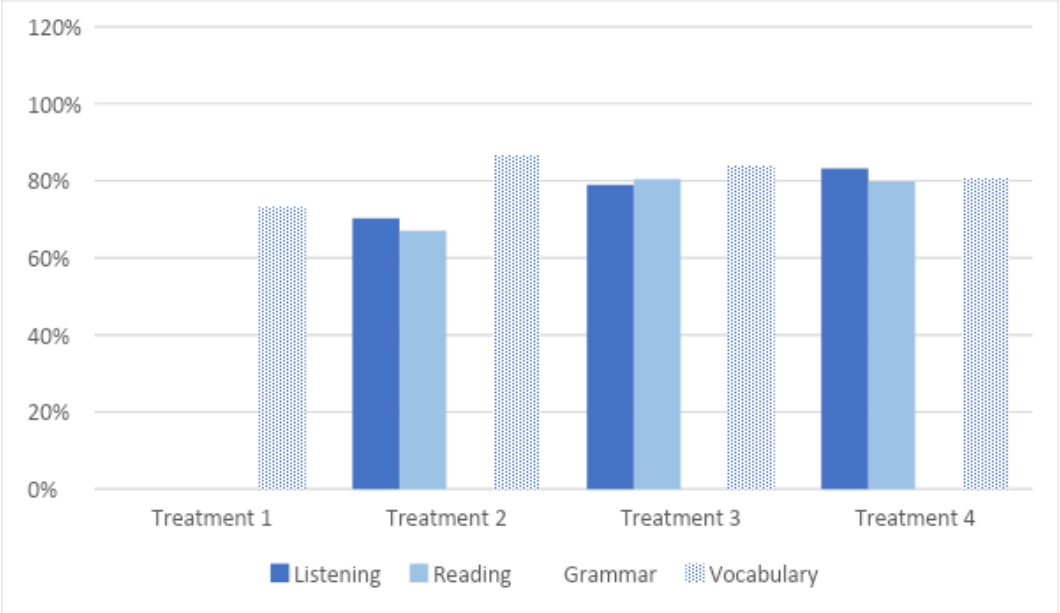


Figure 17: *The performance of treatment group in listening, reading, grammar and vocabulary sections.*

The comparison of scores in listening, reading, grammar, and vocabulary sections for the treatment and comparison groups is presented in Figures 18 and 19. Figure 18 compares the performance of the groups in SAQ format on the final test. The scores in all four sections for the treatment group in SA question format were quite similar. The participants answered correctly to 70-80% of the SA items. The comparison group showed a good performance only in the listening section of the final test where the difference was very small (6.4%) compared to the treatment group. However, the results in the other three sections were considerably lower. The lowest grades were received from the reading (31% difference) and vocabulary (57.6% difference) sections of the final test. These sections in the midterm test were conducted mainly of dichotomous (true/false) and MCQ item tasks.

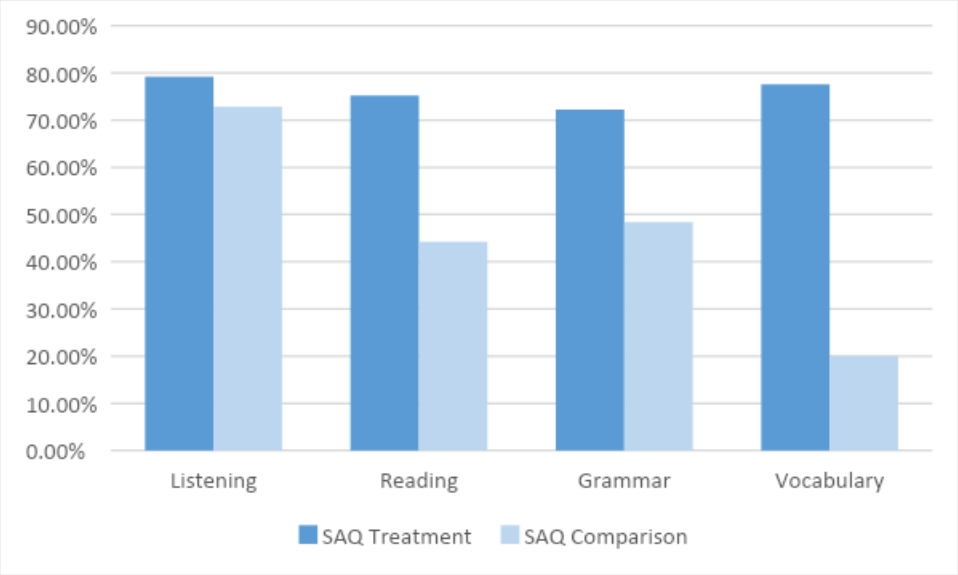


Figure 18: *The performance of the groups in SAQ format of the final test*

Figure 19 shows the difference in performance in the MC question format in the final test. Previously in data analysis, it was revealed that the difference in test performance of the treatment and comparison groups was significantly affected by the scores on the MCQ format in the final test. Even though in every section the treatment group showed better performance in the multiple-choice format, the comparison group also succeeded in more than 60% of the tasks in listening, grammar, and vocabulary sections. Yet, there was a significant difference between the groups regarding the reading section of SAQ format which caused the extreme diversity in grades. While the treatment group was able to answer 91% of the questions, the comparison group answered only 29.2% of the reading tasks. This resulted in a 62.4% difference in grades which was greater than for the short-answer format tasks (31%) in the reading section. As for the vocabulary section, the comparison group performed better in the MC question format compared to the performance on the SAQ format, and the difference between the mean scores of the treatment and comparison groups in the vocabulary section was just above 20%. The results indicate that the average performance of the comparison

group in the MCQ format of the final test was negatively affected mainly by the scores in the reading section.

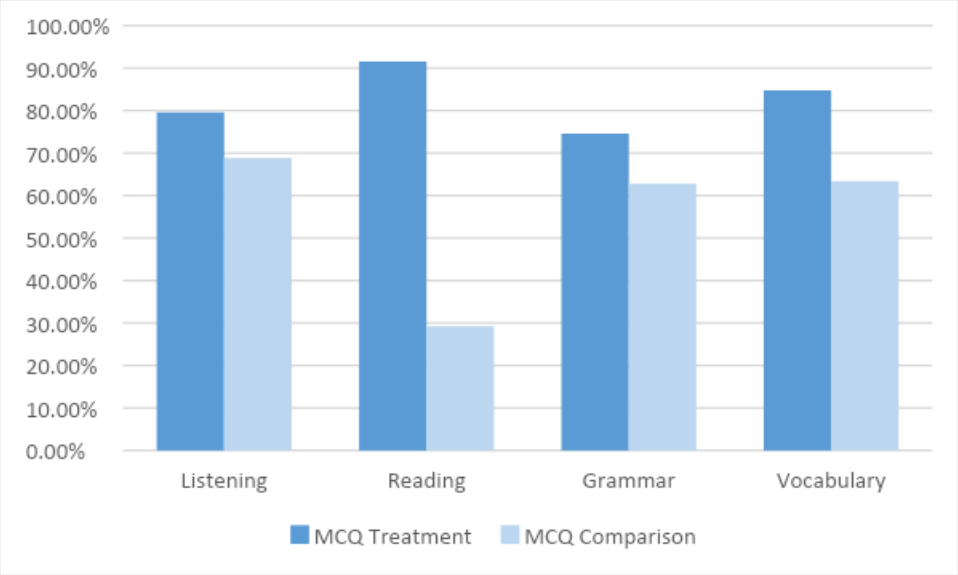


Figure 19: *The performance of the groups in MCQ format of the final test*

Another major factor that impacted the final score was the tendency of the students to skip the SAQ questions in the final test. The results of the students from the treatment and comparison groups were examined to establish the tendency of not attempting the tasks in the treatment and comparison groups. As displayed in Figure 20, the comparison group showed a high tendency for skipping the short-answer questions. In the treatment group, none of the students skipped the SAQ tasks, meanwhile, in the comparison group, 10% of the participants did not make an attempt to answer the tasks with the short-answer question format. It is important to mention that the most amount of not attempted tasks was from the reading and vocabulary sections. The ratio of not attempted tasks of the students from the comparison group was 40% from the reading section in SAQ format and 40% from the SAQ vocabulary section. This also resulted in lower grades on the reading and vocabulary sections on SAQ format in the final test.

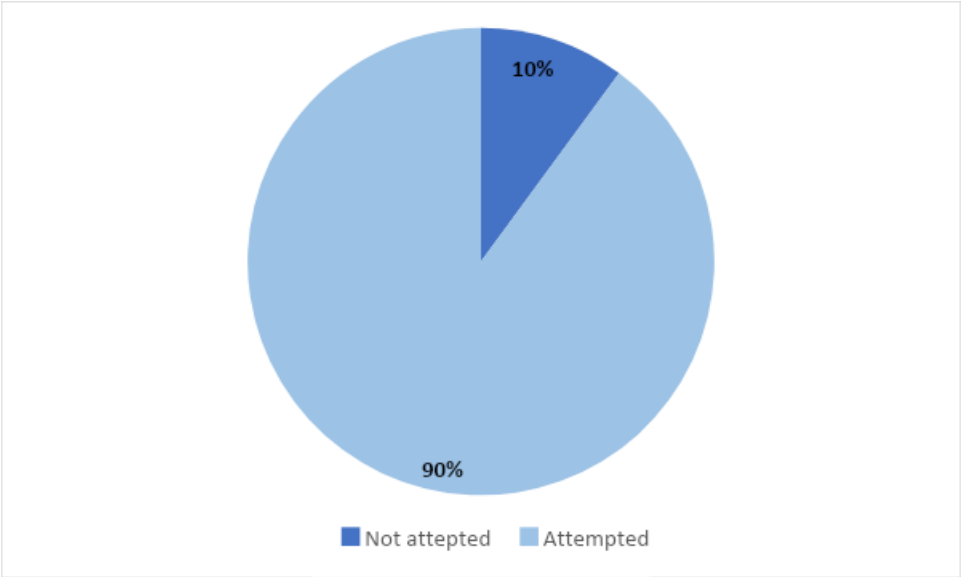


Figure 20: *The ratio of attempted and not attempted SAQ items in the final test for the comparison group*

4.3. Research question 3: What are the students' perceptions towards MC and SA question formats?

Eighteen students out of nineteen participated in the survey and 4 students from the treatment and comparison groups were interviewed to find out the reasons for preference of MCQ or SAQ formats. Before the analysis of the survey results, the internal consistency of the survey items was measured with Cronbach's α reliability coefficient. There was found an α = 0.70 internal consistency for the questions included in the survey. The findings indicate a high internal consistency, and the results of the survey are reliable.

Figure 21 represents the preference of MC and SA question formats as a part of the language test. The results are established with the mean rate of the points evaluated with the Likert Scale. Both treatment and comparison groups showed a high agreement on including MCQ format items in language tests. At the same time, the SAQ items were less preferred by the

students. Particularly, the participants from the comparison group were not sure whether they would like to have more SA questions in their achievement tests. Whereas the treatment group was more enthusiastic about the constructed-response items. The results of the interviews supported the outcomes of the survey and revealed that in the comparison group, the learners would prefer to include more multiple-choice tasks, meanwhile even the students from the treatment group who did not receive high grades in the final test had a preference for SAQ format.

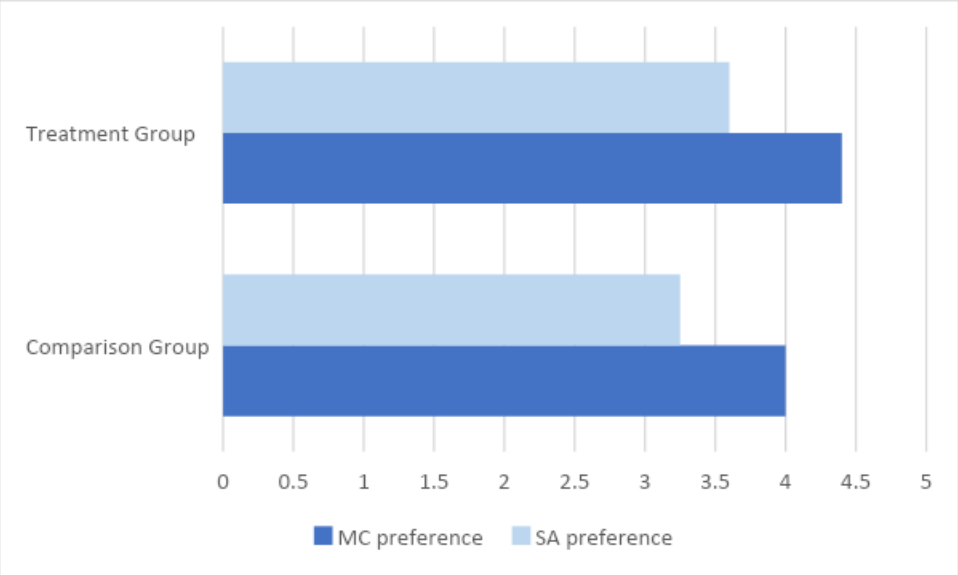


Figure 21: *I prefer answering multiple-choice (MC) / short-answer (SA) questions in my English language test.*

The simplicity of answering the MCQ and SAQ question formats and the amount of time allocated to answer these formats were linked together. As presented in Figure 22, the students found MC questions easier to answer than SA questions. The comparison group believed that it was faster to answer the MCQs but was more neutral to the SAQ format. The participants in the treatment group disagreed with the statements and found that neither format took much time to answer. The majority of the interviewees pointed out that the main reason for considering the

MCQ items easy-to-answer is that they required less time to respond. On the other hand, several students considered SAQ items easier. They found it faster to construct responses once they knew the correct answer rather than trying to find the best answer among the multiple options.

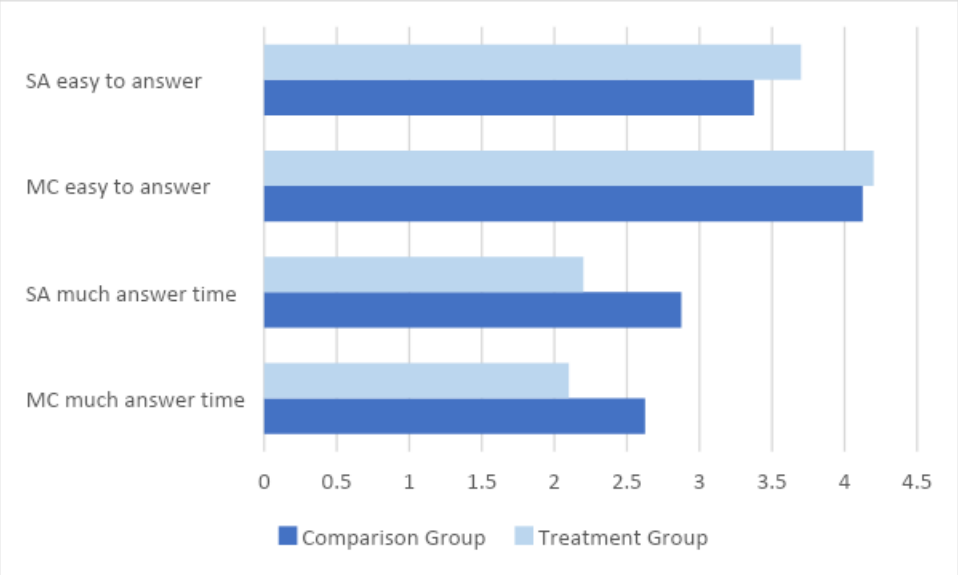


Figure 22: *Multiple-choice (MC) / short-answer (SA) questions take much time to answer.*

Multiple-choice (MC) / short-answer (SA) questions are easy to answer.

The survey results displayed that for the treatment group SAQs unlike MCQs checked the understanding of the course material better (Figure 23). Whereas the comparison group gave an advantage to the multiple-choice format. In addition, the students from both groups agreed that MCQ and SAQ formats allowed the test-takers to notice the gaps in their knowledge, though, the SAQ items were more preferable to reveal the gaps (see Figure 23). The interview indicated the same results and the participants accepted that MCQ and SAQ formats are equally appropriate for checking language knowledge. However, the students mentioned that it was challenging to answer the short-answer questions with poor knowledge about the material, which resulted in the number of not attempted tasks in this question format.

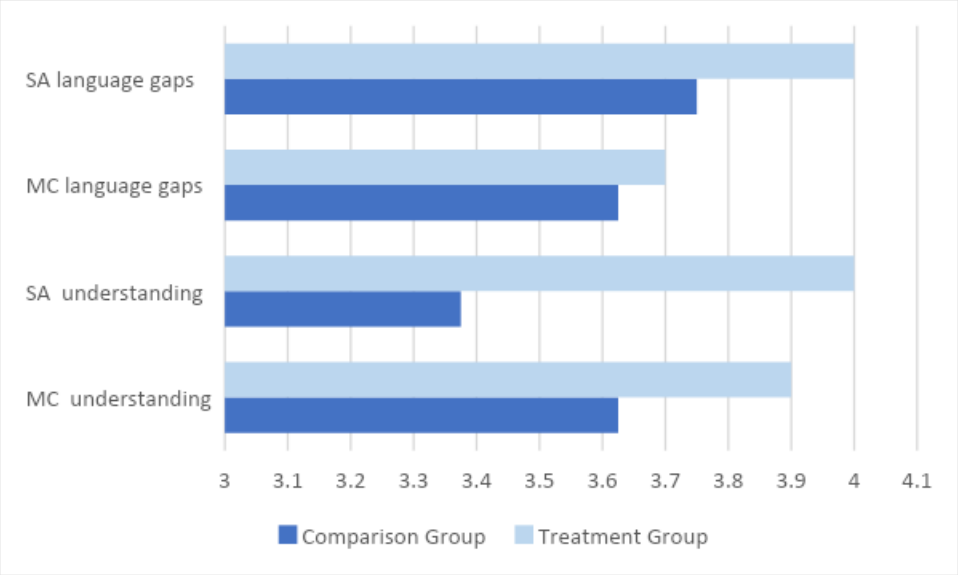


Figure 23: *Multiple-choice (MC) / short-answer (SA) check my understanding of the course material. Multiple-choice (MC) / short-answer (SA) provide me information about my language gaps.*

The groups responded quite differently to the preference of question format which evaluates better their language practice in a real-life context (Figure 24). The treatment group gave a considerable advantage to the SA question format, meanwhile, the comparison group favored the MCQ items. As for the better performance in the language test, most of the participants of the treatment and comparison groups agreed that MCQ questions help the students to show a better performance on tests (Figure 24). During the interview, the participants explained that by saying that the MCQ questions are easier to answer. The learners of both groups found the SAQ items were less helpful for this case, they believed that there was a lower possibility to gain high grades with this item format. The interview results showed that in students' opinions the SAQ items eliminate the guessing factor and make them more difficult to answer for learners who do not know the correct answer to the question.

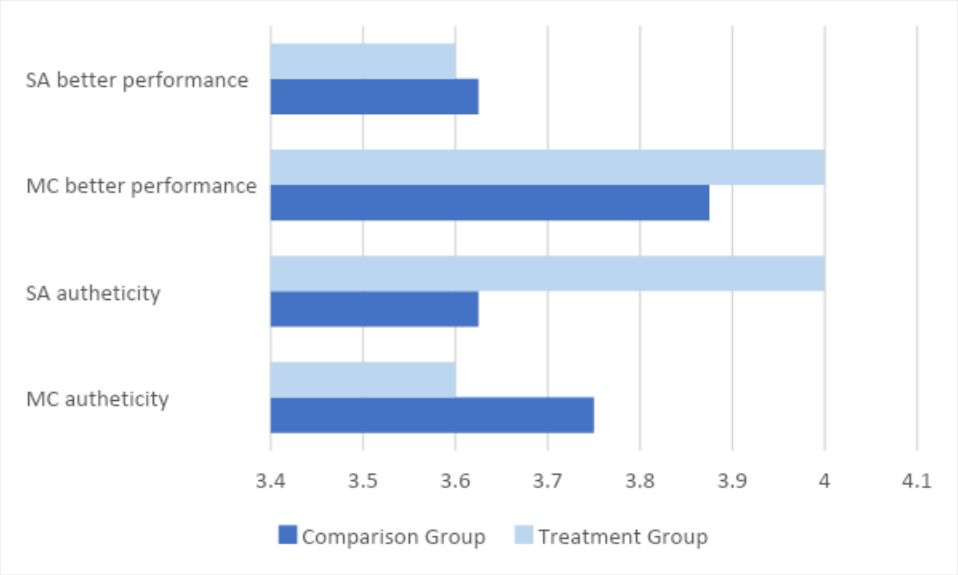


Figure 24: *Multiple-choice (MC) / short-answer (SA) show my language performance in real-life. Multiple-choice (MC) / short-answer (SA) questions help to perform better during the language test.*

During the interviews, the participants mentioned several advantages and disadvantages of MCQ and SAQ formats. The multiple-choice questions required less time to answer and this was the main advantage for all of the students. Besides, the MCQ format allowed them to evaluate and compare the options before choosing the correct answer. Among the drawbacks were mentioned the guessing factor, which enabled the students to answer the questions without any background knowledge. Moreover, the participants reported that they were restricted with provided options and their answers did not always match with the possible choices. Based on the learners' opinion, the main advantage of the SAQ format was the ability to express ideas and thoughts, also short-answer questions required more time to think before providing the answers. From the disadvantages of the short-answer format, the students pointed out that it was almost impossible to answer the questions without any knowledge about the material which resulted in skipping the questions rather than providing incorrect answers.

CHAPTER FIVE: DISCUSSION AND CONCLUSION

5.1. Discussion

The results of the current study are in line with identical studies where the tests with multiple-choice and short-answer question formats are compared. Particularly, the research draws parallels with what Masoumi and Saeghi (2020) discovered, where the topic of the study is similar to the current research.

The findings of this research report that in most of the cases the students perform consistently on MCQ and SAQ formats. Similar to previous studies (Hastedt & Sibberns, 2005; Bleske-Rechek et al., 2007; McKenna, 2019) the mean scores on the MCQ items were slightly higher than the scores on the SAQ format, however, a number of students performed equally well, or even outperformed in SAQ format. The vast majority of test-takers rarely received more than 20% higher on one format than on another, and the effect of the format was not notable. Hickson et al.'s study (2012) also confirms that the majority of the students perform equally on the MCQ and SAQ formats. On the other hand, other researchers suggest that the difference between learners' performance on MC and SA question formats is significant (In'nami & Koizumi, 2009; Budiyono, 2018; Masoumi & Saeghi, 2020). These findings revealed that the MC format is considerably easier for the students and their test performance can be highly impacted by the format of the question. However, the participants of these studies are bigger in number compared to the current research.

Correlations were found between the performance on multiple-choice and short-answer question formats. There was a strong relation between the scores received on both formats and students' final test scores. The good performance on SAQs resulted in a good performance in MCQs and vice versa. In their study, Bleske-Rechek et al. (2007) also state that there is a strong

correlation between learners' performance on SAQ and MCQ items. However, students' performance differs regarding their proficiency levels. In the current study, the scores of the students with lower grades are highly influenced by the performance on short-answer questions. The study established a large difference between their scores on MCQ and SAQ formats for lower proficiency students even within the same proficiency level. Meanwhile, there was found a minimal difference in scores on MC and SA questions for the learners who got higher grades on their language test. Yeneva et al. (2020) have also found a high correlation between students' performance on question format and their proficiency. As stated in the study conducted by Mozafari et al. (2017) learners with lower proficiency are less comfortable with the short-answer question format and this results in their poor performance on the test.

The results of the study confirm that the MCQ and SAQ question formats do not have a great influence on students' test performance. The learners' performance is greatly influenced by their proficiency. Both MC and SA question types help the teachers to reveal test-takers' level of understanding of the course material. The only difference between the two formats is that the SAQ items display a more detailed analysis of students' language knowledge and can result in a fair performance on the SAQ format, only when the participant's proficiency is low.

The results also demonstrated that the students who were completing the treatments on a weekly basis performed well on both the MCQ and SAQ formats of the final test. There was no considerable difference between the scores on multiple-choice and short-answer tasks. Several students even received higher grades on the SAQ format. Their good performance on both question formats could be a result of the constant practice of similar constructed-response tasks during four weeks. The familiarity with the task types could positively affect their performance on the SAQ format and give an advantage over the students of the comparison group. Yet, the

analysis of the treatment tasks revealed little progress (4.87%) throughout that time. The learners of the treatment group were already quite proficient and had minor difficulties with the reading and listening section only during the second treatment. It appears that the performance on tests is impacted by the constant practice of the SAQ format but this is not a critical factor, students' achievements on both question formats are mainly influenced by their language competency. However, the results do not provide clear evidence of how the practice of the SAQ format resulted in the good performance of MCQ items on the final test. These findings support the study of Yaneva et al. (2020) where the researchers state that the performance on question formats is highly dependent on students' proficiency.

The further analysis of listening, reading, grammar, and vocabulary sections established that the low scores were caused by students' poor performance especially on reading and vocabulary sections. Learners with little practice of SAQ format skipped the questions that required much time for the analysis and interpretation of the provided information. They preferred to give very short answers (one to three words) and avoided the tasks where explanations were needed. Budiyono (2018) suggested that short-answer questions are more demanding since in reading tasks the learners prefer dealing with factual details rather than concepts. Regarding the vocabulary tasks, the students' performance varied in MCQ and SAQ formats. Similar to Masoumi and Saeghi's (2020) study the learners performed remarkably better on multiple-choice tasks. It can be implied that it is easier for students to select the word from the provided options than to recall it from the context. Another reason for low grades in reading and vocabulary sections could be explained by the fact that the previous achievement tests included the short-answer questions only in listening and grammar sections, but no SAQ items

were included in the reading and vocabulary sections. This could become a challenging factor for the students during the test and could hugely affect their final scores.

The findings of the current study assume that the students who regularly practiced short-answer questions performed significantly better on the multiple-choice format. Whereas the learners from the comparison group were able to complete only 57% of the tasks in the MCQ format. The analysis of different sections established that in the listening, grammar, and vocabulary sections the performance of the comparison group was reasonably stable, whereas in the reading section the students received quite low grades. One possible explanation is that the MCQ format was not included in the reading section of their achievement tests, and the learners' reading comprehension was assessed with dichotomy (true/false) questions. However, they were familiar with the MC question format as similar tasks were completed during their classes. Another reason for the huge gap between the treatment and comparison groups on the MCQ reading section can be the constant practice of the SAQ items of the treatment group. Since the short-answer questions require a more thorough analysis of the text, the students were able to analyze the information in deeper levels, and possibly that helped them to easily exclude the distractors and find the correct answer. Despite that, further investigations are needed to find out how the constant practice of the SAQ format can influence the performance of reading comprehension in the MCQ format.

Another finding of the study is that during the research, the students answered to a greater number of MCQ items than in SA question format. A number of short-answer tasks were left not attempted by the students from the comparison group, whereas the multiple-choice questions received full answers for all tasks. For instance, in Grunert's (2013) research the students from a technology course left 25% of the constructed response tasks open and that resulted in lower

scores on the test. This can be due to the fact MCQ items provide the students with options and help them to make deductions to find the correct answer. However, sometimes the students find the correct answer without knowing it. In the current study in the final test students' knowledge on the same topic was assessed with different question formats, since all the tasks were designed of MCQ and SAQ items. Due to the fact that the learners answered the MC questions but skipped the SA questions on the same topic, we cannot exclude the possibility that those learners did not know the correct answer to the question and simply made an attempt to guess the answer in the MCQ format. In the study conducted by Sam et al. (2009) the students of medical faculty were using clues of the provided clinical information to eliminate the incorrect options. McKenna (2019) argues that although the distractors can provide cues to test-takers, yet sometimes they can mislead the learners. The researcher suggests providing partial credit for partially correct distractors.

On contrary, the SAQ format eliminates the guessing factor and requires the students to produce their answers. Constructed response items not only show the students' actual knowledge but also reveal their impairments in other linguistic aspects. Sam et al. (2009) and McKenna (2019) believe that the SAQ format highlights the areas with gaps in students' knowledge, and therefore is more suitable to measure students' competence. These features make the SAQs a more authentic assessment tool since it shows how the students use their knowledge in real-life contexts where they have no clues to choose the answer.

Based on several studies (Cohen, 2012; Masoumi & Saeghi, 2020) the reason for the difference in performance on MC and SA question formats is that the students use various skills and processes. The SAQ format requires the test-takers more processing, as they need to understand and produce the answer, meanwhile in MCQ items require only to comprehend and

choose the option. In a multiple-choice format, the participants can imply different test-wisiness strategies, which stands as a benefit for them to perform better during the test (Liao, 2018). McKenna's research (2019) indicated that it is possible to construct objectively scored questions in SA format similar to MC questions.

One more reason for the difference of scores in multiple-choice and short-answer question formats can be the inter-rater reliability. Even though in most of the cases the raters agreed on the scores of SAQ items, they differed in their self-judgments and as a result, the points provided for some of the tasks were not similar. However, the variance between the scores was not big enough to affect negatively students' scores on this format. For instance, in Liao's (2018) research the author encourages the examiners to participate in assessment training beforehand, in order to increase the reliability of scores provided for the SAQ items. In addition, it would be preferable for the raters to use a rubric for the assessment of short-answer responses, as the credits provided by the raters for partially correct answers very often differ. For this reason, the thoroughly conducted rubric can help to decrease the cases of score mismatch between the two raters and result in the increase of reliability of scores on SAQ format.

One of the main disadvantages of short-answer questions is thought to be their assessment time, as they are more time-consuming compared to the MCQ format. Yet, during the current research, they did not cause any issues regarding the marking time. The answers of the students did not differ much from pre-existing sample correct answers which made it easier for the raters to check students' responses. Although all the possible answers were registered in Moodle in advance, the platform did not always work perfectly for the SAQ format, and most of the answers were checked manually. Therefore, if the number of test-takers is not big, the time restriction is not a reason to reduce the number of SA questions in the assessment and affect the

reliability of the test. The findings support the study of Sam et al. (2019), who believe that the marking schemes reasonably reduce the scoring time. Moreover, with the practice of constructed response items, the learners gain experience how to answer them and give fewer incorrect responses, which also positively impacts marking time and costs.

According to the survey and interview results, the students have different attitudes towards MCQ and SAQ formats. The participants mainly prefer multiple-choice tasks, as they have a lower difficulty level and enable them to finish the test faster. On the other hand, the students also mentioned that in case they know the correct answer to the question, it is easier for them to produce their answer, rather than to eliminate the incorrect options. Similar to Liao's (2018) study the results of the philosophy test indicated that the options of MC question format provide clues that help them to choose the correct answer faster. However, if there are two credible answers, the MCQ format can hinder understanding and become more time-consuming.

Furthermore, the students stated that the short-answer question format not only checks their understanding of the course material but also helps them to reveal their knowledge gaps. Several students just left the SAQ tasks without attempts of accomplishing them, meanwhile, in MCQ format they would randomly choose one of the options. The reason for missing the SAQ tasks was that the students did not know the answer to the question, whereas in MCQ tasks they tried to eliminate the incorrect answers and find the correct ones. In addition, the students think that SAQs show their language performance in a real-life context; this means that they are more authentic than MCQs. According to Liao's (2018) research, the participants noticed many gaps in their vocabulary knowledge, word spelling, sentence structure, and overall writing skills. Constructed-response items provide the learners an opportunity to improve their writing skills.

Additionally, it is worth mentioning that, unlike the MC question format, SAQs provide the teachers with general feedback about students' course achievements and display the areas that cause difficulties for the students. Moreover, with the SA question format, it was possible to determine the cases of student cheating during the final test. As the online assessment format does not allow the teachers to control fully the testing process, and many students have problems with their devices, it is not feasible to hinder the test-takers from sharing the answers. Contrasted to the MCQ format, in short-answer tasks the majority of the students write the same answers without even paraphrasing (especially when they are required to produce sentences). This helps the teachers to detect cheating students, whereas in the MCQ format it is almost impossible.

To sum up, the findings suggest including more questions using the short-answer format to assess students' knowledge in reading, listening, grammar, and vocabulary sections, since they provide not only reliable results but also in case of small groups are quite easy to score. A similar conclusion was reached by a number of researchers (Granville et al, 2004; Bleske-Rechek et al., 2007; Mozaffari et al., 2017; Budiyono, 2018; Masoumi & Sadeghi, 2020) who believe that by mixing the MCQ and SAQ formats the assessment process can benefit more.

5.2. Limitations

There were several limitations and delimitations included in the study. One of the limitations was the number of participants. Although overall 19 learners participated in the research, the number was not enough to have generalizable results. In addition, the attrition of several students made it difficult to make conclusions about the impact of MCQ and SAQ items on students' performance regarding their proficiency level.

Further, the short duration of the study and consequently the limited time for the treatments did not allow to fully demonstrate the impact of short-answer question format on

students' test performance and also the performance of multiple-choice items. This perplexed the further considerations regarding the extent to which the SAQ format can particularly influence on better performance of MCQ tasks, and thus restricted to bring clear explanations about the degree of effectiveness of the treatments.

Another limitation was the development of a parallel test to monitor students' performance on MCQ and SAQ formats. By reason of the requirements of the program, the final test was longer in a format compared to the midterm test and accordingly included a greater number of tasks. This restriction created additional difficulties with calculations of the differences in students' test performance when comparing their overall progress in midterm and final tests.

5.3. Delimitations

The delimitations of the research were the context and proficiency level of the participants. Due to the fact that the research aimed to identify the level of influence of multiple-choice and short-answer question formats on students' test performance, the questions tried to cover a wide range of higher-order thinking skills. The skills that are necessary to answer these types of questions require the participants to have relatively high language proficiency. The context of the study was another essential factor to expose the authenticity of the tasks and understand the use of the acquired knowledge in real-life situations.

5.4. Conclusion

Overall, the current study has tried to reveal the impact of the multiple-choice and short-answer questions on students' test performance. The results of the research lead to the following conclusions.

Firstly, students' performance on the language test changes depending on the scores received from the MCQ or SAQ items formats. However, there is no considerable difference between the learners' scores on both question formats, as they show nearly identical performance on MCQ and SAQ items. Although generally, students perform better on the MCQ format, with the increase of proficiency the gap between MC and SA question types becomes minimal. If the students' language proficiency is low, the gap between the performance on MCQ and SAQ formats is quite noticeable since it is easier for them to answer the MC questions. The reason for this difference is mainly caused by the lack of students' linguistic knowledge in that specific aspect, and the constructed-response items enable to reveal the language gaps better than the MCQ items. The findings of the study indicate that students' test performance is more dependent on their language proficiency rather than on the question format.

Secondly, the constant practice of short-answer questions can positively affect students' test performance. The format of this question type allows the students to analyze the information more accurately, and pay attention to details, which helps them to find the correct answer without hints. They also show the learners' ability to apply their knowledge in a real-life context. The SAQs also help the learners to have fewer difficulties responding to the MC questions. As the students already know the correct answer and they know what they are looking for, it becomes easier for them to deal with the distractors. In addition, with the practice of the SAQ items, the

learners start spending less time answering them, which will positively affect their performance on the test.

Furthermore, for assessment purposes, the combination of both formats can provide the teachers with additional feedback about students' achievements on the course. In some cases when students give incorrect answers on the multiple-choice questions, it is hard to follow the reason for misleading factors that brought the respondents to the wrong conclusion. Meanwhile, the SAQs help the teachers to find out the language areas that need additional explanations and practice. Additionally, the scoring of the test, with combined format is not much time-consuming, as for the SAQs the students in most of the cases provide answers that are similar to the sample correct answer. If the teachers would like to decrease the assessment time, they can use online assessment platforms such as Moodle or Google Forms that are quite simple to use, as grading is done automatically and there are free trails provided.

Finally, students feel comfortable with MCQ and SAQ formats and they believe that both question types should be included in the language test. As an advantage of the multiple-choice format, the learners believe that this question type helps them to perform better on the language test. Many of them find the MC questions slightly easier and faster to answer compared to the SAQs. Yet, part of the learners prefers the short-answer question format. They explain their choice with the fact that short-answer questions are a more reliable measure for their language knowledge. In their opinion, the SAQs display their actual understanding of the course material and are more authentic than the MCQs. Despite that fact, there was found little difference between the students' perceptions towards MCQ and SAQ formats, and from their viewpoint, none of the formats can negatively affect their performance on the test.

The findings of the current research are not generalizable; however, the study can be replicated. It would be important for future research to investigate the impact of question formats on students' test performance including a larger number of participants and conducting the research within a longer period of time. It would be interesting to compare EFL learners' test performance regarding different proficiency levels (including Low Intermediate, Intermediate, Upper Intermediate, and Advanced levels). Also, further research is needed to confirm how can the constant practice of the SAQ format affect students' performance on MCQ tasks in different sections of the test (reading, listening, grammar and vocabulary). For this purpose, the study should include a greater number of treatments and also use different instruments apart from the pre-and post-test and the treatments.

REFERENCES

Adair‐Hauck, B., Glisan, E. W., Koda, K., Swender, E. B., & Sandrock, P. (2006). The Integrated Performance Assessment (IPA): Connecting assessment to instruction and learning. *Foreign Language Annals*, *39*(3), 359-382. https://doi.org/10.1111/j.1944-9720.2006.tb02894.x

Alderson, J. C. (2000). Assessing reading. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511732935

Alfred, L., Dison, L., & Hagemeier, L. (2000). Developing a sociology foundation course: an analysis of principles, methods and materials. *South African Journal of Higher Education*, *14*(3), 112-119.

Azevedo, J. M. M. L., Lopes, A. P., Babo, M. D. L., & Torres, C. (2010). Multiple-choice tests: a tool in assessing knowledge. *ICERI*.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford university press.

Baghaei, P., & Amrahi, N. (2011). The effects of the number of options on the psychometric characteristics of multiple-choice items. *Psychological Test and Assessment Modeling*, *53*(2), 192.

Behizadeh, N., & Engelhard Jr, G. (2014). Development and validation of a scale to measure perceived authenticity in writing. *Assessing Writing*, *21*, 18-36.

Bleske‐Rechek, A., Zeug, N., & Webb, R. M. (2007). Discrepant performance on multiple‐choice and short answer assessments and the relation of performance to general scholastic aptitude. *Assessment & Evaluation in Higher Education*, *32*(2), 89-105. https://doi.org/10.1080/02602930600800763

Budiyono, B. (2018). Multiple-Choice Questions (MCQs) ss Short-Answer Questions (SAQs) For Inferential Comprehension. *English Education: Journal of English Teaching and Research*, *3*(2), 71-83. https://doi.org/10.29407/jetar.v3i2.12506

Campbell, M.L., 2015. Multiple-Choice Exams and Guessing: Results from a One-Year Study of General Chemistry Tests Designed to Discourage Guessing. *Journal of Chemical Education* 92 (7), 1194-1200.

Canagarajah, S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly: An International Journal*, *3*(3), 229-242.

Chan, C. (2009). Assessment: Short Answer Questions, Assessment Resources HKU, University of Hong Kong. *Retrieved January*, *22*, 2015.

Cheng, L., & Lingyun, G. A. O. (2002). Passage dependence in standardized reading comprehension: Exploring the College English Test. *Asian Journal of English Language Teaching*, *12*, 163-178.

Cheng, L., Klinger, D. A., & Zheng, Y. (2007). The challenges of the Ontario Secondary School Literacy Test for second language students. *Language Testing*, *24*(2), 185-208. https://doi.org/10.1177/0265532207076363

Cohen, A. D. (2012). Test-taking strategies and task design. *The Routledge handbook of language testing*, 262-277.

Currie, M., & Chiramanee, T. (2010). The effect of the multiple-choice item format on the measurement of knowledge of language structure. *Language Testing*, *27*(4), 471-491. https://doi.org/10.1177/0265532209356790

Dávid, G. (2007). Investigating the performance of alternative types of grammar items. *Language testing*, *24*(1), 65-97. https://doi.org/10.1177/0265532207071512

Dehnad, A. et al, 2014. Comparison between Three-and Four-Option Multiple Choice Questions. Proceedings of the International Conference on Current Trends in ELTA. *Procedia - Social and Behavioral Sciences* 98, 398–403.

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessments: The limited scientific evidence of the impact of formative assessments in education. *Practical Assessment, Research, and Evaluation*, *14*(1), 7. https://doi.org/10.7275/jg4h-rb87

Famularo, L. (2007). The effect of response format and test taking strategies on item difficulty: a comparison of stem-equivalent multiple-choice and constructed-response test items. *Boston College*.

Faremi, Y. A. (2016). Reliability coefficient of multiple choice and short answer objective test items in basic technology: comparative approach. *Journal of Educational Policy and Entrepreneurial Research (JEPER)*, *3*(3), 59-69.

Funk, S. C., & Dickson, K. L. (2011). Multiple-choice and short-answer exam performance in a college classroom. *Teaching of Psychology*, *38*(4), 273-277. https://doi.org/10.1177/0098628311421329

Granville, S., Chamberlain, C., Delmont, E., Button, A., & Dison, L. (2004). The role of short answer questions in developing higher order thinking. A *Journal of Language Learning*, *20*(2), 28-45.

Grunert, M. L., Raker, J. R., Murphy, K. L., & Holme, T. A. (2013). Polytomous versus dichotomous scoring on multiple-choice examinations: development of a rubric for rating partial credit. *Journal of Chemical Education*, *90*(10), 1310-1315.

Hancock, G. R. (1994). Cognitive complexity and the comparability of multiple-choice and constructed-response test formats. *The Journal of experimental education*, *62*(2), 143-157. https://doi.org/10.1080/00220973.1994.9943836

Hastedt, D., & Sibberns, H. (2005). Differences between multiple choice items and constructed response items in the IEA TIMSS surveys. *Studies in Educational Evaluation*, *31*(2-3), 145-161.

Hickson, S., Reed, W. R., & Sander, N. (2012). Estimating the effect on grades of using multiple-choice versus constructive response questions: data from the classroom. Educational Assessment, 17(4), 200–213 https://doi.org/10.1080/10627197.2012.735915.

Huang, S. H. (2016). Communicative Language Teaching: Practical Difficulties in the Rural EFL Classrooms in Taiwan. *Journal of Education and Practice*, *7*(24), 186-202.

Hussey, E. K., Teubner-Rhodes, S. E., Dougherty, M. R., & Novick, J. M. (2010, March). Parsing under pressure: The role of performance pressure in cognitive control and syntactic ambiguity resolution. In *Poster presented at CUNY 2010: Conference on Human Sentence Processing, New York City, NY*.

Ibbett, N. L., & Wheldon, B. J. (2016). The Incidence of Clueing in Multiple Choice Test Bank Questions in Accounting: Some Evidence from Australia. *e-Journal of Business Education and Scholarship of Teaching*, *10*(1), 20-35.

In'nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing*, *26*(2), 219-244.

Kang, D. M. (2005). A study on short-answer-question format in testing EFL learners' L2 reading ability.

Kastner, M., & Stangla, B. (2011). Multiple choice and constructed response tests: Do test format and scoring matter? *Procedia-Social and Behavioral Sciences*, *12*, 263-273.

Kerkman, D. D., & Johnson, A. T. (2014). Challenging Multiple-Choice Questions to Engage Critical Thinking. *Insight: A Journal of Scholarly Teaching*, *9*, 92-97.

Kobayashi, M. (2002). Method effects on reading comprehension test performance: Text organization and response format. *Language testing*, *19*(2), 193-220. https://doi.org/10.1191/0265532202lt227oa

Liao, J. T. (2018). Multiple-choice and short-answer questions in language assessment: the interplay between item format and second language reading *(Doctoral dissertation, The University of Iowa)*.

Marvaniya, S., Saha, S., Dhamecha, T. I., Foltz, P., Sindhgatta, R., & Sengupta, B. (2018, October). Creating scoring rubric from representative student answers for improved short answer grading. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (pp. 993-1002).

Masoumi, G. A., & Sadeghi, K. (2020). Impact of test format on vocabulary test performance of EFL learners: the role of gender. *Language Testing in Asia*, *10*(1), 1-13.

McDaniel, R. (2021, February 18). Bloom's taxonomy. Retrieved March 01, 2021, from https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/

McKenna, P. (2019). Multiple choice questions: answering correctly and knowing the answer. *Interactive Technology and Smart Education*. https://doi.org/10.1108/ITSE-09-2018-0071

Moore, H., 2014. Comment on "Does the student a) know the answer, or are they b) guessing?" [The Conversation]. Available at: https://theconversation.com/does-the-student-a-know-the-answer-or-are-they-b-guessing-31893 [accessed April 19, 2018].

Mozaffari, F., Alavi, S. M., & Rezaee, A. (2022). Investigating the impact of response format on the performance of Grammar tests: Selected and constructed. *Journal of Teaching Language Skills*, *36*(2), 103-128. doi: 10.22099/jtls.2017.23918.2154

Mullen, K., & Schultz, M. (2012). Short answer versus multiple choice examination questions for first year chemistry. *International Journal of Innovation in Science and Mathematics Education*, *20*(3).

Palmer, E. J., & Devitt, P. G. (2007). Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple-choice questions? Research paper. *BMC medical education*, *7*(1), 1-7.

Parmenter, D. A. (2009). Essay versus multiple-choice: Student preferences and the underlying rationale with implications for test construction. *Academy of Educational Leadership Journal*, *13*(2), 57.

Pinckard, R. N., McMahan, C. A., Prihoda, T. J., Littlefield, J. H., & Jones, A. C. (2009). Short‐answer examinations improve student performance in an oral and maxillofacial pathology course. *Journal of dental education*, *73*(8), 950-961.

Rademakers, J., Ten Cate, T. J., & Bär, P. R. (2005). Progress testing with short answer questions. *Medical teacher*, *27*(7), 578-582. https://doi.org/10.1080/01421590500062749

Roediger III, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1155. https://doi.org/10.1037/0278-7393.31.5.1155

Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language testing*, *23*(4), 441-474. https://doi.org/10.1191/0265532206lt337oa

Sam, A. H., Westacott, R., Gurnell, M., Wilson, R., Meeran, K., & Brown, C. (2019). Comparing single-best-answer and very-short-answer questions for the assessment of applied medical knowledge in 20 UK medical schools: Cross-sectional study. *BMJ open*, *9*(9), e032550.

Sangwin, C. J., & Jones, I. (2017). Asymmetry in student achievement on multiple-choice and constructed-response items in reversible mathematics processes. Educational Studies in Mathematics, 94(2), 205–222.

Simkin, M. G., & Kuechler, W. L. (2005). Multiple‐choice tests and student understanding: What is the connection? *Decision Sciences Journal of Innovative Education*, *3*(1), 73-98. https://doi.org/10.1111/j.1540-4609.2005.00053.x

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher*, *32*(2), 3-13. https://doi.org/10.3102/0013189X032002003

Srivastava, A., Dhar, A., & Aggarwal, C. S. (2004). View Point-Why MCQ.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, *87*(4), 324-328. https://doi.org/10.1177/003172170508700414

Tankersley, K. (2007). *Tests that teach: Using standardized tests to improve instruction*. ASCD.

Tychonievich, L. (2012). Multiple choice or short answer? The intellectual impact of question format. Retrieved from http://www.cs.virginia.edu/~lat7h/blog/posts/206.html

Weimer, M. (2015). Advantages and Disadvantages of Different Types of Test Questions. Retrieved from https://www.facultyfocus.com/articles/educational-assessment/advantages-and-disadvantages-of-different-types-of-test-questions/

Wu, W. (2008). Misunderstandings of Communicative Language Teaching. *English language teaching*, *1*(1), 50-53.

Yaneva, V., Harik, P., Pandian, R., Morales, A., & Clauser, B. (2020, December). Automated Prediction of Examinee Proficiency from Short-Answer Questions. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 893-903).

Appendix A

Midterm Test

Section 1 Listening Comprehension (5 points)

Transcript

Unites 1 and 2: Listening test. Listen and choose the correct answer a, b, or c.

Example: I don't really know how she manages, but mom never seems to get annoyed. It doesn't matter what's going on, she just takes everything as it comes. To be honest, I can't remember the last time she lost her temper.

Speaker 1: We're all so different in our family. I'm quite quiet and relaxed about things, but my brother Dan is quite the opposite. If he wants something, he just goes on and on till he gets his way. I don't know how he does it but he often ends up doing what he wants, even if you don't want to do that.

Speaker 2: Tom is a new boy in our class. Perhaps it's too early to say but he seems rather pleased with himself. It might just be that he is very confident, but I think he has quite a high opinion of himself. He certainly looks a little conceited. He always knows all the answers and he's always the first to put his hand up.

Speaker 3: Well, I'd never trust my cousin Fiona. She always seems very friendly and open but actually she isn't quite what she seems. She's one of those people who says one thing but does another. You never really know where you are with her.

Speaker 4: My uncle Garry has had lots of ups and downs in his life. He's lost his job recently but you never hear him complain. He's just very positive and he's always got a smile on his face. I really like spending time with him and he makes me feel good. He tells great jokes and lots of really funny stories of when he was my age.

Speaker 5: I tell my best friend Julia everything. I've been having a few problems at home recently and she's always there for me. It doesn't matter what time I call her; she listens and know what to say. I don't know what I'd do if she wasn't around.

A Which adjective best describes the person each speaker is talking about? Listen and choose the correct answer: a, b or c. (2.5 points; 0.5 each)

0	a) cheeky	b) *calm*	c) careless
1	a) **intellectual**	b) **pretentious**	c) ***pushy***
2	a) ***smug***	b) **eccentric**	c) **bright**
3	a) **excitable**	b) ***hypocritical***	c) **inconsiderable**

Section2 Reading Comprehension (5 points)

Read the text. Are the sentences true (T) or false (False)?

Risking it all

There is a certain type of person who thrives on the addictive adrenaline rush of doing something dangerous. In the last few years there has been a noticeable rise in the number of so-called extreme sports like mountain boarding, sky surfing, whitewater kayaking and barefoot waterskiing, to name but a few. What they all have in common is an element of danger and whether they take place on land, water, or in the air, they all involve speed, height and a high level of physical exertion together with spectacular stunts.

The most spectacular (and dangerous) of these sports, however, must surely be base jumping. BASE stands for Buildings, Antennae, Spans (bridges) and Earth. In other words, anything that is high enough to be jumped off with a parachute.

Base jumpers belong to a secretive, 'underground' community that likes to play while the world sleeps. Two of Britain's most successful base jumpers are Dan Witchalls, a veteran with over ten years' experience, and his right-hand man Ian Richardson. To become a member of this elite group, you can apply for an official number once you've carried out all four genres of jumps successfully (namely: building, antenna, span and earth). As of 2011, approximately 1,400 base numbers have been allocated, but interestingly, not many to women.

So what is it that motivates someone to risk their life in this way? Dan and Ian explain that a successful jump gives an unbeatable high, but the lows are often fatal. One wrong turn or gust of wind can result in a very serious accident. Since base jumping was invented there have been 147 deaths, meaning one in six people who take up the sport are killed. Dan is a natural athlete, a roofer by profession, who has escaped serious injury in over 800 jumps. 'I can't drive past a new building or crane without thinking, maybe, maybe …' he says. But now the risks he's taking are becoming more and more extreme as he goes in search of new challenges and new thrills. Fine perhaps for him, but not so easy for friends and family who must surely wonder each time he goes out whether he'll come back safely.

Ian Richardson, on the other hand, has earned a reputation among his base jumper friends as being rather accident-prone. In 2009 he was almost killed when he smashed into the side of a building after his parachute opened incorrectly. This has not, however, stopped him from risking

it all, time and time again. To an outsider, taking such risks seems somehow irresponsible but clearly this sport has a psychological hold over its participants.

0	Everyone enjoys taking risks.	*F*
1	The number of extreme sports has not increased in recent years.	*F*
2	All extreme sports are dangerous.	*T*
3	Base jumping only involves jumping off buildings.	*F*
4	Base jumpers don't practise their sport openly.	*T*
5	You can become a member as soon as you've performed one jump.	*F*
6	Few women have applied for membership.	*T*
7	Approximately 15% of people taking up the sport have lost their lives.	*T*
8	Dan has been seriously injured.	*F*
9	His stunts are becoming more dangerous.	*T*
10	A mechanical failure almost killed Ian.	*T*

Section 3 Language Use (15 points)

1 Match the two parts of the sentences. (2.5 points)

	A		B
0	I forgot my wallet,	4	which means I'm often late for school.
1	I live a long way from school,	1	which makes it hard to meet up with friends during the week.
2	Cycling is an endurance sport,	5	which makes me angry.
3	I have to train hard every day,	2	which means you have to be really fit.
4	I don't like getting up early,	0	which meant I had to borrow some money.
5	My brother sometimes uses my mobile phone,	3	which means I often get tired.

2 Rewrite the sentences with a relative pronoun so that they have the same meaning. (2.5 points)

0 It is strange to me that Ben enjoys extreme sports like bungee jumping.
Ben enjoys extreme sports *like bungee jumping, which is* strange to me.

1 I'm running the New York marathon next year. This means I have to train every day.
I'm running **the New York marathon next year, which means I have to train** every day.

2 I missed my bus. This made me late for school.
I missed **my bus, which made me** late for school.

3 My school organises a lot of sports events. This makes it fun for us.
My school **organises a lot of sports events, which makes it** fun for us.

4 Swimming is a very competitive sport. That is why you have to train hard.
Swimming is **a very competitive sport, which is why you have to** train hard.

5 I'm afraid of water. This means I would never try surfing.
I'm afraid **of water, which means I would never** try surfing.

1 Put the words in the correct order. (2.5 points)

0 Jemma / generosity / What / about / like / is / I / her
What I like about Jemma is her generosity.

1 is / rudeness / amazes / John's / What / me
What amazes me is John's rudeness.

2 yesterday / I / what / remember / said / I / don't
I don't remember what I said yesterday.

3 wrong / doesn't / done / Matt / what / know / he's
Matt doesn't know what he's done wrong.

4 pushy / you / Sally / What / is / forget / really / that / is
What you forget is that Sally is really pushy.

5 I've / what / You / believe / won't / heard
You won't believe what I've heard.

2 Join the two sentences to make one. (2.5 points)

0 Luke's really considerate. I like that about him.
What *I like about Luke is that he's really considerate.*

1 I read something about Morocco. I don't remember it.
I don't remember **what I read about Morocco.**

2 Fred is quite pushy. I find it really annoying.
What **I find really annoying is that Fred is quite pushy**.

3 You're not always right. You should remember that!
What **you should remember is that you're not always right**.

4 English spelling is hard. I find it really challenging.
What **I find really challenging is that English spelling is hard**.

5 Katie is very kind. It makes her special.

What **makes Katie special is that she is very kind**.

1 Complete the sentences with the correct form of the verbs. (5 points)

talk	get	be	meet	help	meet	go	help	open	talk
		want	try	decide	suggest	stop			

0 I don't remember *seeing* that film before.

1 We tried **to get** tickets for the concert but it was sold out.

2 You always promise **to help** me, but then you never do.

3 I've never enjoyed **being** on my own.

4 I suggest **meeting** later and then we can decide what to do.

5 If you stop **talking** for just one minute, you'll know what we're meant to be doing.

6 Please stop **talking**. You're really getting on my nerves!

7 I tried **to open** my Facebook page, but there was no internet connection.

8 I don't remember **meeting** you. Are you sure we know each other?

9 You always promise **to help** me but then you never do.

10 Do you feel like **going** out this evening?

11 Please can you **try** to remember where you left your phone?

12 I'**ve / have decided** not to go to university this year as I want to have a gap year.

13 **Stop** talking and listen to me for once!

14 Matt **suggested** meeting up at 6, so we did.

15 Which film do you **want** (you) to see this evening?

Appendix B

Treatment 1

Section 1 Grammar (2.5 points)

1 Read the sentences and if possible, replace *used to* with *would*, otherwise put *X*.

1. My dad used to whistle while he was working.
   Answer: **My dad would whistle while he was working.**

2. The dog used to follow her everywhere.
   Answer: **The dog would follow her everywhere.**

3. They used to arrive at the end of June and stay until September.
   Answer: **They would arrive at the end of June and stay until September.**

4. She used to have green hair.
   Answer: **X**

5. I used to have a horse but I sold it.
   Answer: **X**

2 Complete the table with the information about what people used to do in the past and what we do nowadays. (2.5 points)

1. Sent telegrams	*Send emails*
2. Used candles	**Use electricity**
3. Made their own toys	**Buy toys from shops**
4. Travelled in the carriages	**Travel in cars and planes**
5. Wrote letter	**Write emails / use the phone**
6. Wore long skirts and dresses	**Wear jeans, T-shirts, short skirts, shorts**
7. Listened to records	**Listen to CDs / MP3 players**
8. Had lots of children	**Have only one or two children**

Section 2 Vocabulary (5 points)

3 Read the article and complete the missing words. (2.5 points)

Shoppers looking for a natural high were able to satisfy their needs by inhaling, or breathing in, flavored oxygen from the first oxygen bar to be (0) launched in Cardiff. Inhaling oxygen through a tube is a new (1) _ _ _ _ _ which is expected to (2) _ _ _ _ _ _ _ in nightclubs across the UK. The trend started in Tokyo, but has quickly (3) _ _ _ _ _ _ to Britain and the US. "The demand has been (4) _ _ _ _ _ _ _ _ _ _ _ _ _" says the manager of the bar, Neil Lucas. Some customers say they are already (5) _ _ _ _ _ _ on the experience. "You feel really good and full of energy afterwards", says Ray, a local businessman.

Answer:

1. **craze**
2. **catch on**
3. **spread**
4. **overwhelming**
5. **hooked**

4 Complete the sentences with adverbial phrases. (2.5 points)

You might be told it's wrong to look at people (0) in a rude way, but a new craze called Stare Master is winning fans across America. People are queuing up (1) _____ to take part in starting concerts. The Stare Master concerts are held (2) _____ and have really strict rules – you aren't allowed to laugh, close your eyes, nod or move. "It's great", said one fan, "because it tests your self-control (3) _____." It was invented by two friends because they were bored and wanted to spend their free time in (4) _____.

Answer: 1. With enthusiasm, 2. In public, 3. In a different way, 4. In a fun way

Appendix C

Treatment 2

Section 1 Grammar (5 points)

1 Make sentences comparing the actions. Use the *adverbs* and *as … as*.

0. *Mary (eats / drinks / quickly).*
   Mary eats as quickly as she drinks.

1. He (reads / talks / slowly)
   Answer: **He reads as slowly as he talks.**
2. Frank (doesn't work / plays / hard)
   Answer: **Frank doesn't work as hard as he plays.**
3. She (plays the guitar / sings / nicely)
   Answer: **She plays guitar as nicely as she sings.**
4. We (arrived / could / soon)
   Answer: **We arrived as soon as we could.**
5. I (don't speak French / speak Spanish / well)
   Answer: **I don't speak French as well as I speak Spanish.**

Section 2 Vocabulary (5 points)

2 Read and complete the sentences.

0. *The house was really dirty, so we have cleaned it from top to bottom.*
1. Your pullover looks kind of strange – I think it's inside _____.
2. Sorry, that's not right – the _____ way round.
3. You've got your T-shirt on _____ to front.
4. He was covered in mud from head to _____.
5. Turn the bottle upside _____ and shake it – then the sauce will come out.

Answer: 1. Out, 2. Wrong, 3. Back, 4. Toe, 5. Down

Section 3 Reading Comprehension (5 points)

3 Read the text and fill in the table. (2.5 points)

Fashion & music

We're used to seeing music stars like Madonna and Lady Gaga wear fashion which create an image for them. But music and fashion have always been closely linked. The tough, leather-wearing image of early rock stars like Gene Vincent Influenced a generation of young people in Europe and in the USA. A cultural war broke out in the mid-1960s in the UK over the reality between the "Mods" (who favored high-fashion, expensive styles) and the "Rockers" (who wore T-shirts and leather); followers of each style had their favorite musical acts, who fed into the rivalry by releasing records praising one style and criticizing the other. In the 1960s, The Beatles brought mop-top haircuts, collarless blazers and Beatle Boots into fashion.

Rock musicians were among the first people to wear hippie fashion, and introduced such styles as the Nehru jacket; bands like the Beatles wore custom-made clothes that had a strong influence on 1960s style. As rock music genres developed, what an artist wore became as important as the music itself in defining the artist's intent and relationship to the audience. The glam rock of the 1970s brought fashion to new level of importance in rock music with the "glitter" image of artists like T. Rex. Some artists who had been active in the late 1960s, such as David Bowie, also adopted a glam-influenced look.

In the late 1970s, disco acts helped make flashy urban styles fashionable, while new wave groups started wearing mock-conservative clothes including suit jackets and thin ties, in an attempt to be as different from mainstream rockers as possible.

In the early 1990s, the popularity of grunge brought in a fashion of its own. Grunge musicians and fans wore torn jeans, old shoes, flannel shirts and backwards baseball hats, and grew their hair against the clean-cut image that was popular at the time, together with a heavily commercialized pop music culture. Musicians continue to be fashion icons; pop-culture magazines such as Rolling Stone often have fashion sections featuring musicians as models.

Year	Style	Clothing items	Artists
1960s	1.	5.	1.
	2. "Rockers"	6.	2.
	3.	7.	3.
	4.	Nehru jacket	
1970s	1.		1. T. Rex
	2. Disco style		2.
	3.		
1990s	4.	Torn jeans, flannel shirts, backwards baseball hats	X

Answer:

Year	Style	Clothing items	Artists
1960s	1. "Mods"	High fashion and expensive cloths	1.Gene Vincent
	2. "Rockers"	T-shirts and leather	2. The Beatles
	3. Beatles style	Collarless blazers, Beatle Boots, custom-made cloths	3. David Bowie
	4. Hippie style	Nehru jacket	
1970s	1. Glam rock style	Glitter cloths	1. T. Rex
	2. Disco style	Flashy costumes	2. David Bowie
	3. Wave group style	Suit jackets, thin ties	
1990s	1. Grunge	Torn jeans, flannel shirts, backwards baseball hats	X

4 Read the text and write down the styles that matches the photos the best. (2.5 points)

a. 

Grunge

d. 

Beatles style

b. 

Modes

e. 

Glam rock

c. 

Rockers

f. 

Disco

Section 4 Listening Comprehension (5 points)

5 Listen to the radio phone-in. Four people talk about their favorites "songs about fashion" and complete the chart. (1.5 points)

Transcript

Host: So, good morning everyone. Welcome to the show and today we're inviting you to call us and tell us your favorite song, or songs, about fashion. Songs about fashion. OK? Well, I bet I know what some of answers will be, but let's see… and wow, that was quick, we have a caller already and it's Janine. Hi Janine, where are you calling from?

Speaker 1: Hi Mike. I'm in Manchester.

Host: OK, Janine and your song is …?

Speaker: Well, there's only one choice really, isn't there? It has to be "Vogue" by Madonna.

Host: Well. There's a surprise! I'd never have guessed that anyone would choose a Madonna song!

Speaker 1: Well right, but isn't it a great song?

Host: I'm not going to answer that, Janine! Thanks anyway for your call and next we have Andy from …?

Speaker 2: From Scotland, Dundee, Scotland.

Host: Morning Andy. So, tell us.

Speaker 2: Well, I'm a huge David Bowie fan …

Host: So, you're going to choose "Fashion" - right?

Speaker 2: That's the one. "Fashion" by David Bowie. Brilliant stuff. It reminds me of my days at university…

Host: Thanks, Andy. No time for all that kind of stuff! University days – pah! Next please – and it's Phyllis. Hi Phyllis.

Speaker 3: Hi Mike.

Host: Where are you, Phyllis?

Speaker 3: I'm in Brighton.

Host: And what's your song?

Speaker 3: Well, Mike, as you can probably tell I'm a bit older than your other callers so I'm going with a 60s classic. It's the Kinks …

Host: "Dedicated Follower of Fashion"?

Speaker 3: Yep! Still love it after all these years.

Host: Well, it's good to know someone still likes it, Phyllis. Thanks for the call and next we have …

Speaker 4: Mark. Hi Mike. I'm phoning from London and I'm going to choose "Freedom" by George Michael.

Host: What? That isn't about fashion! We want songs about fashion, Mark!

Speaker 4: I know Mike – but have you seen the video? It's full of fashion models.

Host: That's not the point, Mark. Oh dear. Let's forget that one. Next please! Actually, no, let's have some music and listen to one of those songs now – we're going to play "Fashion" by Bowie.

Listen to the radio phone-in. Four people talk about their favorite "songs about fashion". Complete the chart with the names in the box.

Speaker	Song name	Artist
Speaker 1: Janine		
Speaker 2: Andy		
Speaker 3: Phyllis		
Speaker 4: Mark		

Answer:

Speaker	Song name	Artist
Speaker 1: Janine	"Vogue"	Madonna
Speaker 2: Andy	"Fashion"	David Bowie
Speaker 3: Phyllis	"Dedicated Follower of Fashion"	The Kinks
Speaker 4: Mark	"Freedom"	George Michael

6 Listen to the radio presenter introduce Philippa Chandler. (1.5 points)

Transcript

Presenter: Welcome to the People Show. This week we'll be talking to Philippa Chandler. Philippa works as a forecaster for Next Big Thing, a company which predicts what the next big youth trends will be. Philippa acts as eyes and ears for some of the world's top companies. She lets them know what their clients like and dislike and what they listen to. Her job is to keep her clients up-to-date, so she regularly interviews people on the street, goes clubbing and surfs the net. The information she gives could help her clients create the next big thing. How did she find a job like that? Well, Philippa read an article on William Higham, the founder of Next Big Thing, Britain's first trend forecasting agency. She sent Higham an email saying she was perfect for the job as a trend spotter because she was very sociable and she had answered questions such as "Why we buy things" as part of her cultural Studies degree. He immediately offered her a part-time job as trainee and six months later she was part of the staff. So, Philippa tell us about your job …

Philippa: I love my job. It satisfies my curiosity but it also makes me even more curious.

Presenter: How do you start your research?

Philippa: I often start by surfing the net for new trends and fashions. I can spend an hour looking at a new website. It's work, but it's not really productive or useful if I spend too long on one thing.

Presenter: What's your advice to young people looking for a job?

Philippa: First of all, they should choose a course they are going to enjoy as well as something they think will be useful for work. Then they should try to get as much work experience as they can. The more experience you have, the easier it is for you to realise what you are good at and what you really like doing. If you haven't tried something, you'll never know.

7 Listen the audio and fill in the missing information.

The current episode of People Show has invited for an interview Philippa Chandler who works as a (1) **forecaster** for Next Big Thing. This company (2) **predicts trends** that are going to be popular among the young generation. Philippa informs the company about their clients' preferences. Her responsibility is to (3) **keep her clients up-to-date**, so she regularly interviews people on the street, (4) **goes clubbing** and (5) **surfs the net**. This information helps to create the next big thing. Philippa has found her job while (6) **reading an article** about the founder of Next Big Thing William Higham. She emailed the company saying that she is the best candidate for the job as she is very (7) **sociable** person and has a (8) **cultural Studies degree** on the topic of "Why we buy things". The company immediately offered her a part-time job and after six months she was part of the stuff.

8 Listen again and continue the sentences. (2 points)

1. The meaning of the "next big thing" is **looking for something new and exciting that a lot of people will want.**
2. Philippa's clients are interested in the information she collects because **it tells them what young people are interested in, and what the latest fashions are.**
3. Philippa's research is not productive when **she spends too long looking at one website.**
4. Philippa's advice to young people looking for a job is **to study something you enjoy, and get as much work experience as you can.**

Appendix D

Treatment 3

Section 1 Grammar (5 points)

1 Finish the sentences with a clause in a correct conditional. (5 points)

0. If you don't wear a coat in the winter, *you get sick*.
1. If it is sunny tomorrow **we'll go to the park**.
2. If you sit in the sun too long **you get burned**.
3. If I were you, **I would go out tonight**.
4. If I were the Prime Minister, **I would make the museums free**.
5. If she had studied harder, **she would have passed the exam**.
6. If I won the lottery, **I would buy a big house**.
7. If I hadn't gone to bed so late, **I wouldn't have been tired**.
8. If she hadn't stayed at home, **she wouldn't have gone shopping**.
9. If I go out tonight, **I'll go to the cinema**.
10. If I had listened to my mother, **I wouldn't have caught a cold**.
11. If I were British, **I would speak perfect English**.
12. If I have enough money, **I'll buy some new shoes**.
13. If I weren't studying English, **I would study French**.
14. If I were on holidays today, **I would go to the beach**.

Section 2 Vocabulary (5 points)

2 Read and sentences and fill in the missing words. (5 points)

1. Would you like to make a **donation** to the library fund?
2. The event is being held in **aid** of charity.
3. Our job will be a lot easier if we **collaborate** with each other.
4. Could you give us a **hand** with the party decorations?
5. I did a charity run and I **raised** out more than 150 € charity.
6. Nobody forced him to go into the army – he's a **volunteer**.
7. It's their problem, not mine. I don't want to get **involved**.
8. There's going to be a big **demonstration** against racism in London next weekend.
9. Would you like to **sign** the petition?

Section 3 Reading Comprehension (5 points)

3 Read the text and complete the sentences.

Giving to Charity at Christmas

The holiday season in America is traditionally a time of gift-giving among family and friends. It is also a time for many Americans to help others who are less fortunate than they are. One such charitable effort has been helping people in New York help other New Yorkers for the past century. Along with the smell of roasting chestnuts on Fifth Avenue, and the sound of hand-bells rung by the scores of Santa Claus look-alikes outside the lively Big Apple department stores, the re-appearance of New York Times Neediest Cases Fund is another signal for New Yorkers that the holiday season is here.

Every day from the beginning of November till the end of January, the New York Times newspaper highlights the plight of around ninety people who are down on the luck, who could be

helped through reader's generosity. "I think the unifying spirit here is one of public principles," says Jack Rosenthal, who administers the Neediest Cases Fund for the New York Times Foundation.

The fund was conceived on Christmas Day 1911, after the newspaper's publisher went for a walk following a big turkey dinner and was approached by a hungry man asking for money for food. The kindly publisher gave it to him and the following Christmas, he asked the stuff of his newspaper to go out and find the hundred neediest people in New York and to write stories about them. The publisher felt confident that New Yorkers, once presented with this information, would sympathize and respond.

That remains the spirit of this fund," says Mr. Rosenthal. The fund has received over two hundred million dollars since it began from donors large and small, and diverse as the city of New York itself. People from all walks of life have donated to the cause.

1. The nickname of New York is _____. (**the Big Apple**)
2. The Neediest Cases Fund is a _____. (**charity fund that helps the New Yorkers who are in need.)**
3. The people donating money to the fund are _____. (the readers of the New York Times newspaper).
4. The neediest Fund was founded on _____. (**25th of December 1911**)
5. The newspaper publishes articles about _____ hoping that the readers would sympathize and help them. (**a hundred poorest people in New York)**

Section 4 Listening Comprehension (5 points)

Transcript

So, do you want to be a hero today? There is a great need for blood donations around the world, and you can be the means of saving lives. According to the American Red Cross, over 40,000 blood donations are needed every single day in the United States alone, and without the help of volunteers like you, it is impossible to fill this need. About 9.2 million people donate every year in the US, and although approximately 38% of the population is eligible to donate, less than 10% of them actually do.

Sometimes people don't donate out of fear, but the process is relatively simple following a four-step process: registering, getting your medical history checked, donating, and then having some refreshments (food and drink that you are given right after you donate to replenish your fluids and energy). And once you donate, you have the ability to donate red blood cells every 56 days.

So, carefully consider becoming a hero today. Donate blood and you can save lives.
Listen to the "Blood Donations" announcement and complete the sentences.

4 Listen to the "Blood Donations" announcement and complete the sentences.

1. There are over **40000 people** in the US per day who need blood donations.
2. The large part of the population is capable to donate blood every year, however **only 10% of them actually do that**.
3. Sometimes people don't donate blood because **of fear / they are afraid**.
4. The four steps of blood donation process are: **registering, checking medical history, donating and having refreshments**.
5. After the blood donation you are allowed to repeat the process (**after 56 days**).

Appendix E

Treatment 4

Section 1 Grammar (5 points)

1 Read the text and put the verbs in the future perfect or future continuous tenses.

Technology is advancing so fast that by the year 2050 who knows what new technologies we (1) **will have invented** (invent). It is quite likely that by 2050 we (2) **will have used up** (use up) most of the earth's natural resources and so we (3) **will be relying** (rely) on wind and solar power. As there will be a shortage of energy, it is quite likely that at that time scientists (4) **will be trying** (try) to find a way for us to live outside the earth. Before the end of the next century, it's possible that people (5) **will be living** (live) in cities on the Moon or perhaps in cities on the seabed.

By 2050 it's possible that scientists (6) **will have already discovered** (already/discover) how to cure diseases such as cancer and, due to the advancement of genetic engineering, maybe hereditary diseases passed down from generation to generation (7) **will have disappeared** (disappear) forever. It is quite possible that by 2050 life expectancy (8) **will have increased** (increase) to 100 and that we (9) **will be enjoying** (enjoy) a healthier existence.

Another area likely to have affected by technology in the year 2050 is education. Many students (10) **will be studying** (study) online from their homes.

Section 2 Vocabulary (5 points)

2 Read the sentences and fill in the missing vocabulary items.

1. By 2030, many species will have **died out**.
2. If the world's **temperature** goes up any more, there won't be enough cold water in the sea.
3. I need to **get rid of** some old cloths.
4. We have the **resources** to end the hinger.
5. If we go on like this, we're going to **use up** all our planet's resources. There won't be any left.

Section 3 Reading Comprehension (5 points)

The Disappearing Honeybee

Do you run away when you hear a bee buzzing? Many people do. A bee sting hurts and some people are allergic to bee venom. But did you know that bees are very important to humans? Honeybees do more than just make honey. They fly around and pollinate flowers, plants, and trees. Our fruits, nuts, and vegetables rely on these pollinators. One third of America's food supply is pollinated by the honeybee.

Have you seen or heard a honeybee lately? Bees are mysteriously disappearing in many parts of the world. Most people don't know about this problem. It is called "colony collapse disorder" (CCD). Some North American beekeepers lost 80% of their hives from 2006-2008. Bees in Italy and Australia are disappearing too.

The disappearance of the honeybee is a serious problem. Can you imagine never eating another blueberry? What about almonds and cherries? Without honeybees, food prices will skyrocket. The poorest people always suffer the worst when there is a lack of food.

This problem affects other foods besides fresh produce. Imagine losing your favourite ice cream! Haagen Daaz is a famous ice cream company. Many of their flavours rely on the hardworking honeybee. In 2008, Haagen Daaz began raising money for CCD. They also funded a garden at the University of California called The Haven. This garden helps raise awareness about the disappearing honeybee and teaches visitors how to plant for pollinators.

Donating money to research is the most important thing humans can do to save the honeybee. Scientists need money to investigate the causes of Colony Collapse Disorder. Some scientists blame CCD on climate change. Others think pesticides are killing the bees. Commercial bee migration may also cause CCD. Beekeepers transport their hives from place to place in order to pollinate plants year-round.

Not everyone has money to donate regularly. There are other ways to help the honeybee. Spread the word by telling your friends and family about the problem. Tell your teacher about the disappearing bees too. Maybe your class can write a letter to the government.

3 Are the sentences True or False? Justify.

1. People are scared of honeybees because their sting causes allergy. (**False. People are scared of honeybees because their sting hurts and some people are allergic to bee venom.**)
2. Over 75 % of Americas' food crops are pollinated by honeybees. (**False. Only one third of America's food crops are pollinated by honeybees.)**

3. Colony collapse disorder (CCD) is the problem of extinction of honeybees. (**True. CCD is a big problem because it largely affects the food supply.**)
4. The disappearance of honeybees can cause an increase in amount of poor people. (**False. The food prices will increase which can affect poor people.**)
5. The donated money is invested in scientific researches to explore the problem of CCD. (**True. Scientists try to solve the problem of CCD and save honyebees from extinction.**)

Section 4 Listening Comprehension (5 points)

4 Watch the video about British broadcaster and natural historian David Attenborough and fill in the missing words.

"Tonight, we've got rather a different program for you. I am David Attenborough, and I am 93. I've had the most (1) **extraordinary** life. It's only now that I appreciate how extraordinary. The living world is a unique and spectacular marvel. Yet the way humans live on Earth is sending it into a decline. Human beings have (2) **overrun** the world. We're replacing (3) **the wild** with the tame. This film is my witness statement and my vision of future. The story of how we came to make this our greatest mistake. And how, if we act now, we can yet (4) **put it right**. Our planet is headed for (5) **disaster**. We need to learn how to work with nature, rather than against it. And I'm going to tell you how."

Appendix F

Final Test

Section 1 Listening Comprehension (5 points)

Transcript

Hannah: Martin. Hi. Sorry I'm late, but I didn't realise what the time was. I have to confess, I stopped off in that new shop on the High Street and just got engrossed.

Martin: No worries. I've already ordered a coffee and I've been reading the paper.

Hannah: Anything interesting?

Martin: Quite an interesting article on nostalgia and our love of the past. It kind of got me thinking about the old days. You know, now when I see all the kids with their mobiles and MP3 players, I can't help thinking about the things we had when we were their age.

Hannah: Yeah, I know what you mean. Everything was just so different, wasn't it? When I was at school, there was a mad craze for French skipping.

Martin: What on Earth was that?

Hannah: I think it must have been a girl thing, but we all did it and it was so simple. All you needed was some elastic and at least three people to take part. Though I used to do it at home using two chairs.

Martin: Really?

Hannah: Hours of fun. And it was good exercise too, as you had to jump over the elastic and the better you were, the higher it went.

Martin: Sounds weird to me, but each to his own. I was much more into the Rubik's Cube. Remember them?

Hannah: Of course. I suppose it's the equivalent of Sudoku today in a way.

Martin: Yeah, they're both totally addictive, that's for sure. When I look back, I can't believe how much time I used to spend fiddling around. It was so frustrating.

Hannah: I never really understood the attraction. I don't have the mind for puzzles like that. Give me something simple, like a crossword and I'm happy.

Martin: Well, I just don't have time now, though I do occasionally do Sudoku if I have to go anywhere by train. Anyway, look, the waitress is coming over. Have you decided what you want to drink?

1 Listen to two people talking about past crazes and answer the questions. (1.25 points: 0.25 points for each)

1. Hannah is late for an appointment with Martin because …
   a. she did not have a watch to look at the time.
   b. she visited the new shop on the High street.
   c. she was buying some products in the shop.

2. What does Hannah think about Rubik's Cube?
   a. It is very complicated.
   b. It is too easy.
   c. It is quite addictive.

3. Listen to the part of the dialogue and answer why does Martin use the phrase "What on Earth was that?"?
   a. He has never heard of that game.
   b. He thinks it was not popular in any country.
   c. He shows that he was never interested in it.

4. By saying "French skipping must have been a girl thing" Hannah emphasizes that …
   a. Only girls play that game.
   b. Boys were never interested in it.
   c. It was mainly popular among girls.

5. Hannah compares Rubik's Cube to
   a. French skipping
   b. Jigsaw puzzles
   c. Sudoku

2 Now, listen again and complete the chart. (1.25 points: 0.25 points for each)

	Craze	Opinion about the craze
Hannah	French skipping	It is a fun game and also a good exercise for the body.
Martin	Rubik's Cube	It is very addictive and takes a lot of time to solve.

Transcript

Mother: So, Tom. Have you had any more thoughts about what you want to do before you go to uni next year?

Tom: Well, I've been thinking of getting involved with some kind of environmental project somewhere.

Mother: Sounds interesting. I certainly think it would be a good idea to get some practical experience, something you can put on your CV.

Tom: Yeah, of course. But the main thing is to do something. It's so easy to sit at home and get worked up about all the stuff that's going on, but it's much harder to actually get out there.

Mother: Yes, you're absolutely right. So, have you thought where you'd like to go?

Tom: I'm thinking of South America.

Mother: Why there specifically?

Tom: Well, I'm interested in all those ancient civilizations. So, I could work for a while and then take some time out to travel.

Mother: Sounds brilliant. But how are you going to pay for all of this, might I ask? Even if you're volunteering, you still need money to cover your flights and living expenses. It won't be cheap, you know.

Tom: Well, mother, dear, that's where you come in.

Mother: And I thought as much. We were going to give you some money for your birthday anyway, so you can put that towards the trip, and perhaps you could ask some local companies to sponsor you.

Tom: That's a good idea. Or I could try and get some articles published. I know everyone does a blog these days, but I could do something a bit more serious, you know?

Mother: Absolutely. We'll have to make a list of all the people we know to see whether anyone could help out. In my experience, it's usually a case of who you know, and one thing inevitably leads to another.

Tom: So, you don't mind then?

Mother: Mind? Why should I? I think it's great that you want to get out there and make a difference.

3 Listen to Tom and his mother talking about his future plans. After read the sentences and choose the correct option. (1.25 points: 0.25 points for each)

1. Tom wants to apply to a university in another country.
   a. True
   b. False
   c. Not given
2. He prefers doing something useful instead of wasting his time.
   a. True
   b. False
   c. Not given
3. Tom wants to go to a country with a rich historical background.
   a. True
   b. False
   c. Not given
4. He does not expect that his parents are going to financially support him.
   a. True
   b. False
   c. Not given
5. Tom is planning to earn some money with blogging.
   a. True
   b. False
   c. Not given

4 Listen again and fill in the missing information in the text.

Tom is planning to apply for a university next year, but before that he would like to join an 1. **environmental project**. His mother finds it important to get some 2. **practical experience** for a better CV. Tom thinks about going to 3. **South Africa** to work and travel there. However, he needs money to accomplish his plans. His mother suggests contacting 4. **local companies** for sponsorship. But Tom also wants to publish some articles. They are going to make 5. **a list of people** who can sponsor him.

Section 2 Reading Comprehension (5 points)

Thank you

This is a huge thank you to all those people who have supported me, both financially and psychologically, on my journey. It wouldn't have been possible without all your generosity and help and I hope that you'll enjoy reading about the difference your donations have made to the lives of children in the poorest countries of the world.

As you know, we set off on our three-week fundraising trek a month ago but the journey started a year before that when I read an article about a local charity that was organising a trek to Everest Base Camp to raise money for a housing project in Rwanda. I'd always wanted to go there so this seemed the ideal opportunity – combining a personal ambition with charity. Those of you who know me will realise that this was indeed a challenge as my fitness levels were certainly not high enough to undertake such an adventure. So, after signing up, paying a deposit and getting my fundraising pack – which included lots of ideas on how to raise the money – I enrolled at the local gym so I could start my fitness programme. However, I soon dropped that idea and bought an exercise bike so I could practice whenever I had time! I'm not really the kind of person who enjoys a workout with lots of other people.

Probably the hardest part of this whole project was the actual fundraising. No one said it would be easy but it certainly was harder than I'd expected. It's amazing how many reasons people will give you as to why they are unable to help! 'Charity begins at home' was one of the phrases I heard most frequently, i.e. why should we help people in Africa when there are so many people here in need of help? I suppose I don't really have an answer to that other than to say that the world is just one big global village now and we're all connected.

To kickstart my fundraising effort, I wrote to everyone I know, both personally and professionally, and to all friends of the family. Many of you reading this will have sent in donations – thank you. I also wanted to involve people in my local community as many of them hadn't heard of the charity. I gave a talk at the local community center' and persuaded the local printer to produce leaflets and flyers for free which we handed out beforehand. There was a good turnout and the event, while not raising much money, certainly encouraged a lot of questions and I'm sure many of the people there will have gone away interested in finding out more.

Within six months I'd raised the money I needed, and of course I funded the costs of the expedition myself. And then the day arrived and the next stage of the adventure began. There is so much to tell you about the trek itself, so I'd like to invite anyone who's free to come along to a special trek evening on 15th August when you can find out more.

5 Read the text and answer the questions. (1.25 points: 0.25 points for each)

1. What is the author thankful for in his letter?
   a. For psychological support from his friends during his project.
   b. For the financial support for education of African children.
   c. For helping him to organize his journey to Africa.
   d. <u>For donating money and standing next to him during his project.</u>

2. How has the author decided to go on a journey to Africa?
   a. <u>He has read about a trek for Rwanda organized by a local charity.</u>
   b. He has come up with an article about poor people in Rwanda.
   c. He has published an article about Everest Best Camp trek.
   d. He has funded some money for a housing charity project.

3. What does the author imply about his physical conditions?
   a. He was not ready for the long flight.
   b. <u>He was not fit enough for journey in Rwanda.</u>
   c. He was not prepared for hiking in the mountains.
   d. He was not healthy enough for hot climate.

4. Who has donated money for the charity?
   a. The friends of the author
   b. The family members of the author
   c. <u>Most of the people the letter is addressed to</u>
   d. Local companies and organizations

5. Did the author manage to accomplish his mission?
   a. He was able to find money for the expedition but not the charity.
   b. He almost reached his goal but the remaining money was funded during the trek.
   c. <u>He succeeded to collect the money necessary to help poor people in Rwanda.</u>
   d. He has decided to get the money in several stages as it was hard to find it at once.

6 Read the sentences and tell whether they are True or False. Justify your answer. (1.25 points: 0.25 points for each)

1. The blog is a thank you only to people who have made donations. (**False. The author thanks everyone who have supported him not only financially but also psychologically.**)
2. A lot of people prefer to support local charities. (**False. Many people find various reasons to not fund money.**)
3. The writer believes we're one big family. (**True. The author thinks that people around the world are connected to each other.**)
4. The printer charged for producing the publicity material. (**False. The local printer did not take money for publication of the materials.**)
5. The writer managed to raise the money just before the start of the trek. (**True. The auther managed to collect the money within six months.**)

Section 3 Language Use (15 points)

7 Read the sentences and choose the correct option. (2.5 points: 0.5 points for each)

1. *'I can run really fast.'*
   a. Tom said me that he could run really fast.
   b. Tom told that he could run really fast.
   c. **Tom said that he could run really fast.**

2. *'There was an accident outside the supermarket'.*
   a. Karen said there has been an accident outside the supermarket.
   b. Karen said there were an accident outside the supermarket.
   c. **Karen said there had been an accident outside the supermarket.**

3. Which sentence is NOT correct?
   a. Maria said she didn't have a key.
   b. George said he was going to be late.
   c. **Mike said he will help me.**

4. *Megan apologised for not coming. She said she had been busy* ____.
   a. previous day
   b. the following day
   c. **that day**

5. *Tom said 'I'm going to a football match tomorrow.'*
   Five days later you say...
   a. Tom said he was going to a football match tomorrow.
   b. Tom said he is going to a football match the following day.
   c. **Tom said he was going to a football match the following day.**

8 Rewrite the sentences in reported speech (2.5 points: 0.5 points for each)

0. *Tim: 'She works in an office.'*
   Tim said she worked in an office.
1. Teacher: 'Rachel has never been in Philadelphia.'
   The teacher said Rachel had never been to Philadelphia.
2. Kathy: 'Jack can speak French.'
   Kathy said Jack could speak French.
3. Sarah: 'I'm trying to find time to go on holiday.'
   Sarah said she was trying to find time to go on holiday.
4. Craig: 'Time goes so slowly when I'm at school.'
   Craig said time went so slowly when he was at school.
5. Mum: 'I didn't have time to go shopping after work.'
   Mum said she hadn't had time to go shopping after work.

9 Choose the correct option for each sentence. (2.5 points: 0.5 points for each)

1. When I was a child I like vegetables.
   a. **didn't use to**
   b. wouldn't
   c. both are correct

2. In the morning she run in the park.
   a. Used to
   b. Would
   c. **Both are possible**

3. Which sentence is **NOT** correct?
   a. I would go shopping with my best friend every weekend.
   b. I would have very short haircut.
   c. She would call me after class to chat.

4. I _____ wear glasses when I was at university.
   a. **Didn't use to**
   b. Wouldn't
   c. Both are correct

5. When the weather was bad, I _____ stay home and watch my favorite TV shows.
   a. Used to
   b. Would
   c. **Both are correct**

10 Complete the sentences using the verb in brackets with the correct form of *used to* or *would*. If both *used to* and *would* are possible, use WOULD. (2.5 points: 0.5 points for each)

When I was a child, we *0. used to live (live)* in a cabin in the mountains. Back then, we **1. didn't use to have** (have) much money; that's why we **2. used to have** (have) vegetables and chickens in our back garden. Every day we **3. would get up** (get up) very early and my father **4. would take** (take) some eggs for breakfast.
The weather in summer **5. used to be** (be) really nice. There was a lake near the house; we **6. would go** (go) there to have a swim every morning. Later, in the afternoon my brothers and I **7. would go** (go) fishing, and in the evening my mother **8. would cook** (cook) whatever we had caught. It was fantastic!

11 Complete the sentences with phrases. (2.5 points: 0.5 points for each)

0. *I cleaned my bedroom from top to bottom the other day.*
1. You've got the numbers **the wrong way round**. It's 348, not 438.
2. If you turn your pullover **inside out**, no one will see the coffee stain.
3. If you hang **upside down** for long enough, all your blood will rush to your head.
4. After playing rugby I was covered in mud **from head to toe**.
5. You must have your T-shirt on **back to front** – it looks very strange.

12 Choose the correct answer: (2.5 points: 0.5 points for each)

0. I've always done _____ work when I can.
   a) **voluntary** b) volunteer c) volunteering
1. I'll be able to walk _____ a stick once I feel a bit better.
   a) **with the aid of** b) with aid of c) with the aid from
2. Please ____ our plan for the redevelopment of this site.
   a) do b) **back** c) give
3. We can ____ you a hand if you'd like us to.
   a) Make b) hold c) **give**
4. _____ workers are brave people.
   a) **Aid** b) Aids c) Aided
5. If we all _____, we'll achieve great results.
   a) correlate b) **collaborate** c) commiserate

13 Read the text and choose the correct word.

Endangered animals are those *0. species* that are in danger of going 1**. extinct**. Their reproductive rates are lower than their mortality rates over long periods of time, so their numbers are decreasing. The reasons for this are varied. Very often, it involves a loss of 2. **habitat**, as people invade their living areas. When a species is listed as 3. **endangered** or threatened, it is not a menace sentence. Many animals, like the bald eagle and the American alligator, were on verge of extinction and are now 4. **recovering**. Many species, however, will not recover, and could be lost forever. Animal species have been going extinct throughout time. Paleontologists estimate that well over 90 percent of all plant and animal species that ever existed have gone extinct. 5. **Pollution** also threatens the survival of many species. Oil spills, acid rain and water pollution have been devastating for many species of fish and birds.

14 Complete the sentences.

0. *I'd like to make a small donation to a charity of your choice.*
1. Please come and sign our petition. We desperately need your support.
2. Would you be able to hand out some leaflets in the town centre tomorrow?
3. Let me give you a hand with your fundraising. I think you need some more help.
4. We're collecting money in aid of the Red Cross so please give generously.
5. You don't have to help out at the weekend – it's completely voluntary.

15 Read the sentences and choose the correct form.

0. *Paul cried. He had been hit | had hit | has been hit by his friend.*
1. They escaped. They had warned | warned | **had been warned** by somebody.
2. The monkey felt safe after it **had hopped** | had been hopped | has hopped on the tree.
3. Jill didn't see the man. He was hidden | had hidden | **had been hidden** behind the wall.
4. I didn't like the hotel which I preferred | had been preferred | **had preferred** before.
5. We didn't want the car that we had shown | **had been shown** | have shown.

16 Read the text and put the verbs in the brackets in past perfect simple or past perfect continuous.

I'm sorry I left without you last night, but I told you to meet me early because the show started at 8:00. I *0. had been trying* (try) to get tickets for that play for months, and I didn't want to miss it. By the time I finally left the coffee shop where we were supposed to meet, I 1. **had been waiting** (wait) for more than half an hour. I had to leave because I 2. **had arranged** (arrange) to meet Kathy in front of the tether.

When I arrived at the theater, Kathy 3. **had already picked up** (pick up, already) the tickets and she was waiting for us near the entrance. She was really angry because she 4. **had been waiting** (wait) for more than half an hour. She said she 5. **had almost given up** (give up, almost) and 6. **had gone** (go) into the theater without us.

Kathy told me you 7. **had been** (be) late several times in the past and that she would not make plans with you again in the future. She mentioned that she 8. **had missed** (miss) several movies because of your late arrivals. I think you owe her an apology. And in the future, I suggest you be on time!

17 Complete the conditional sentences.

0. *I would have told you, if I _____ him.*
   - *a. See*
   - *b. Saw*
   - *c. Would have seen*
   - ***d. Had seen***
1. If they had gone for a walk, they _____ the lights off.
   - **a. Would have turned**
   - b. turned
   - c. have turned
   - d. Had turned

3. My friend _____ me at the station if he gets the afternoon off.
    a. Meets
    b. Will meet
    c. Would meet
    d. Met
4. If I _____ it, nobody would do it.
    a. Don't do
    b. Won't do
    c. Hadn't do
    d. Didn't do
5. If my father _____ me up, I'll take the bus home.
    a. Doesn't pick
    b. Didn't pick
    c. Won't pick
    d. Wouldn't pick

18 Read the text and write the verbs in the brackets in the correct conditionals.

Did you hear about that guy who won 180 million dollars in the lottery? If I *0. won* (win) that much money, I 1**. would quit** (quit) my job the next day. I 2. **would travel** (travel) around the world and 3. **stay** (stay) in the most luxurious hotels. If I 4. **wanted** (want) anything, I 5. **would buy** (buy) it. If I 6. **saw** (see) a beautiful Mercedes that I wanted, I 7. **would buy** (buy) it. If I wanted to stay in a beautiful hotel and the hotel 9. **was** (be) full, I 8. **would buy** (buy) the hotel and make them give me a room. I 10. **could** (can) do anything in the world if I had 180 million dollars.

Appendix G

Survey Questions

1. Group Name

   Խումբը

2. Age

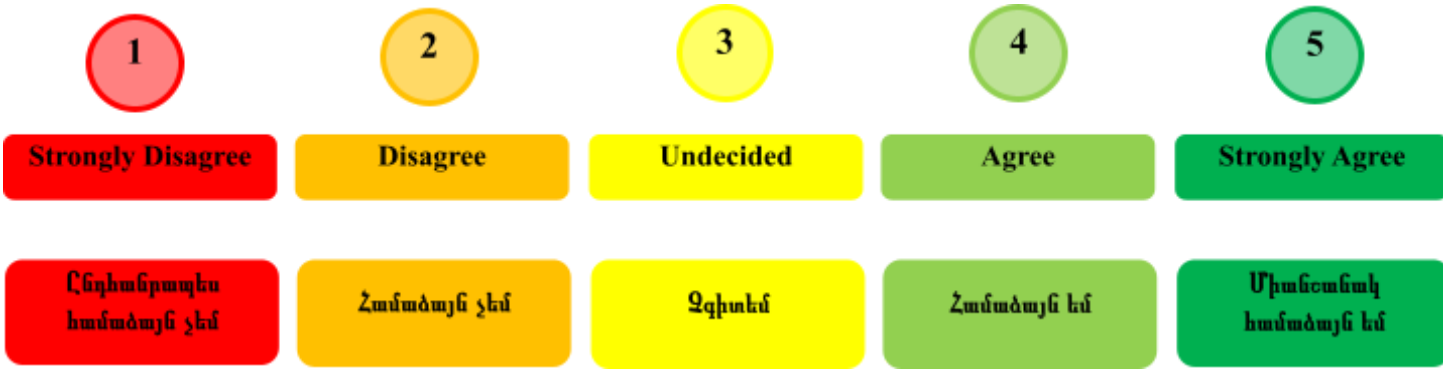
   Տարիքը

3. Gender

   Սեռը

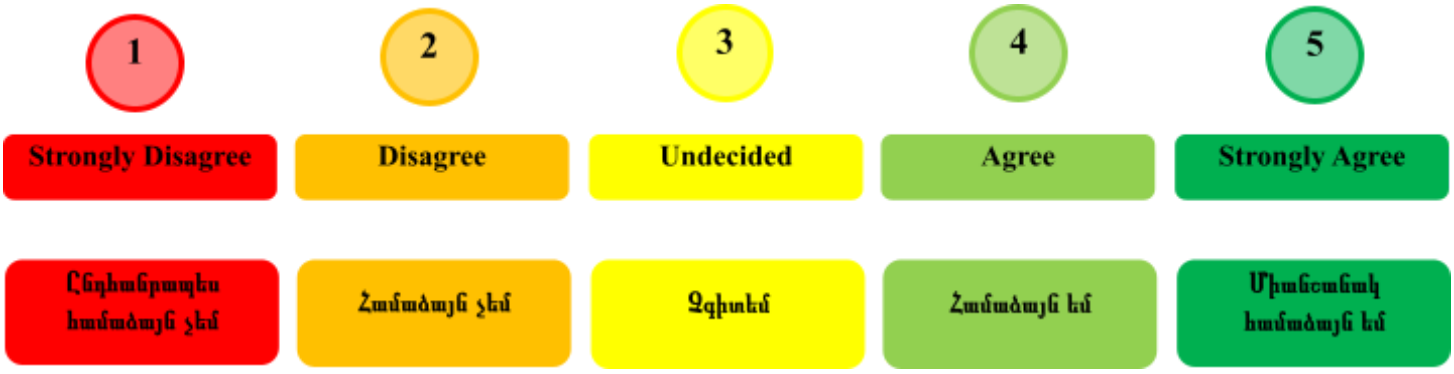
4. I prefer answering multiple-choice (MC) questions in my English language test.

   Ես գերադասում եմ անգլերեն լեզվի թեստի ընթացքում պատասխանել ընտրովի տարբերակներով հարցերին:



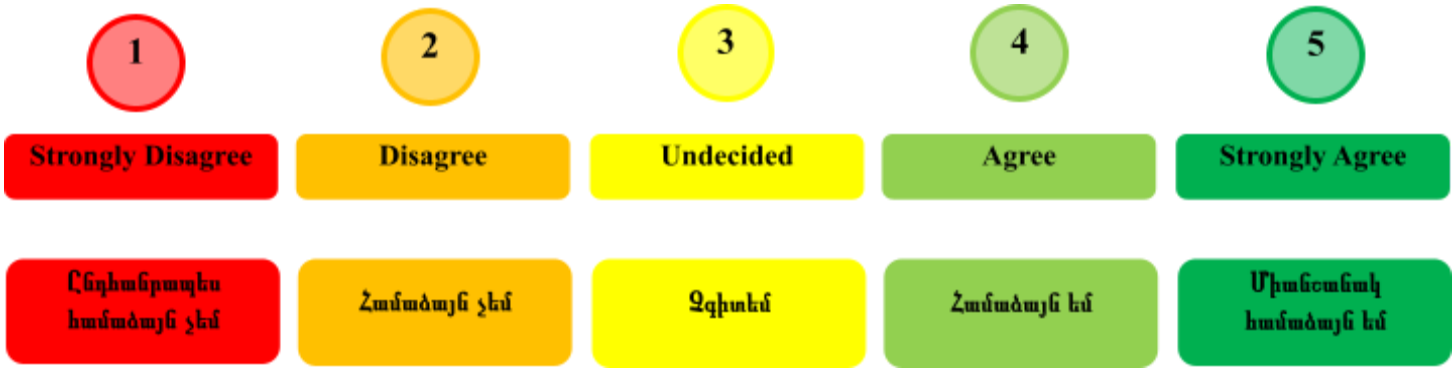
5. Multiple-choice (MC) questions take much time to answer.

   Ընտրովի տարբերակներով հարցերին պատասխանելը բավականին ժամանակատար է:



6. Multiple-choice (MC) questions are easy to answer.

Ընտրովի տարբերակներով հարցերին պատասխանելը հեշտ է:

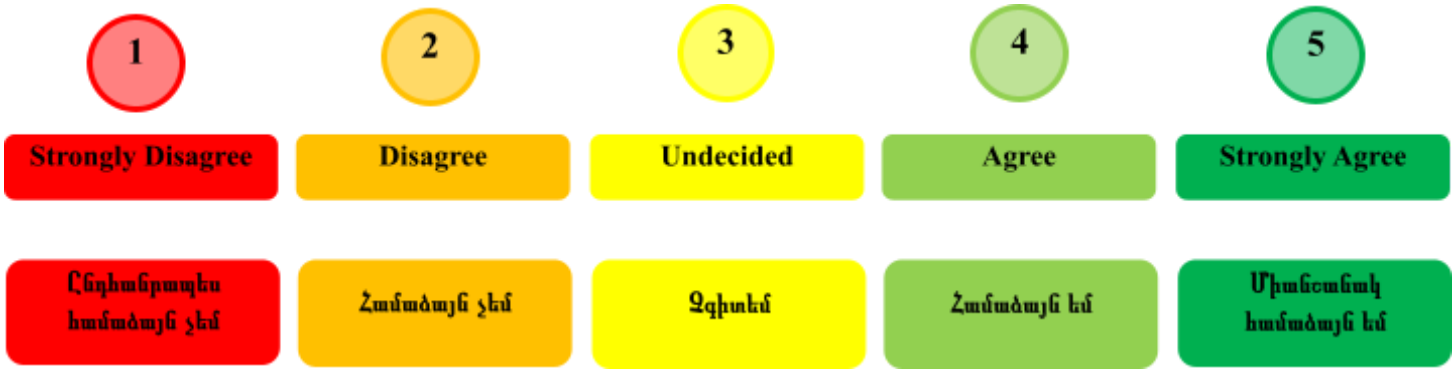


7. Multiple-choice (MC) questions check my understanding of the course material.

   Ընտրովի տարբերակներով հարցերը ստուգում են իմ անգլերենի դասընթացների ընթացքում ձեռք բերած գիտելիքների ըմբռնումը:

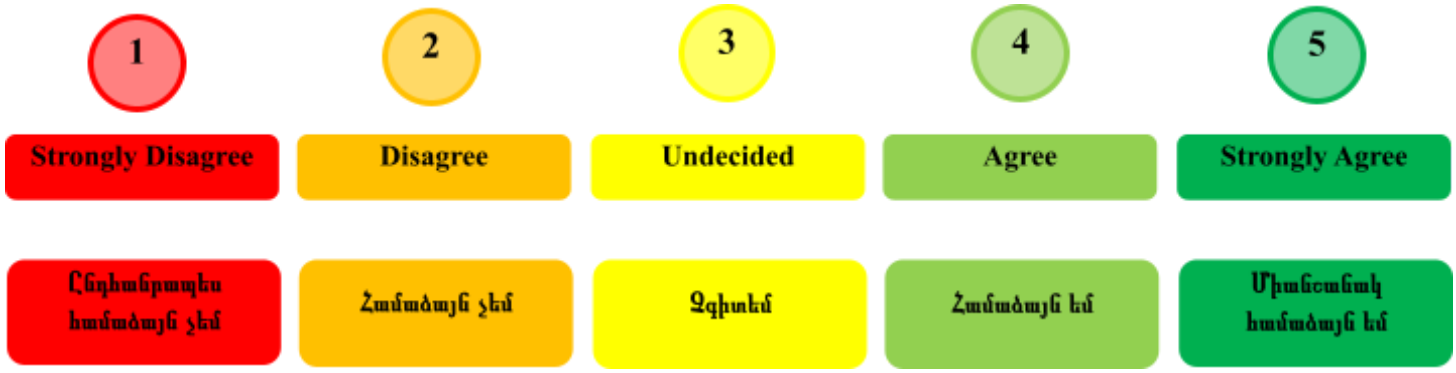
8. Multiple-choice (MC) provide me information about my language gaps.

   Ընտրովի տարբերակներով հարցերը թույլ են տալիս հասկանալ իմ անգլերեն լեզվի գիտելիքների բացթողումները:



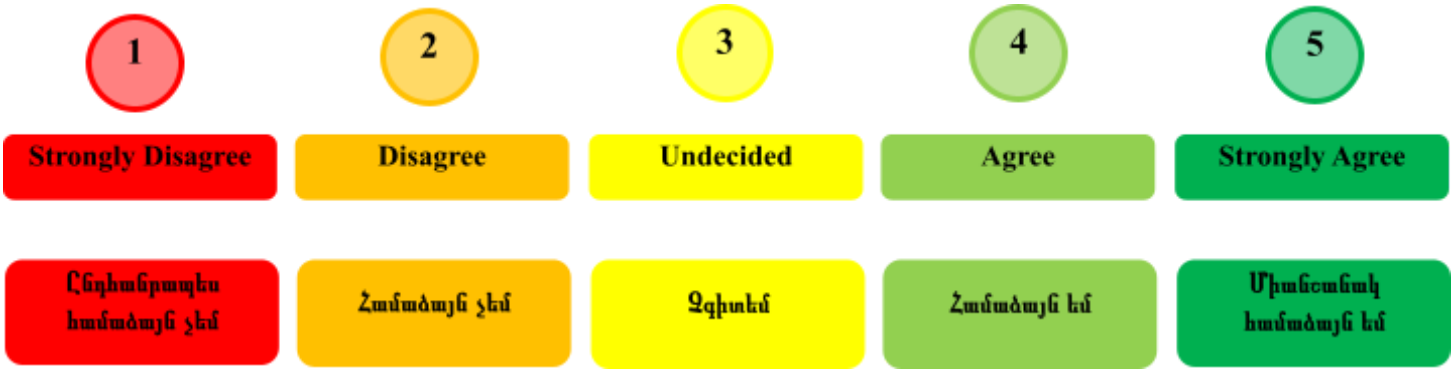
9. Multiple-choice (MC) questions show my language performance in real-life.

   Ընտրովի տարբերակներով հարցերը տրամադրում են իրական կյանքում իմ անգլերենի գիտելիքների օգտագործման ունակությունը:



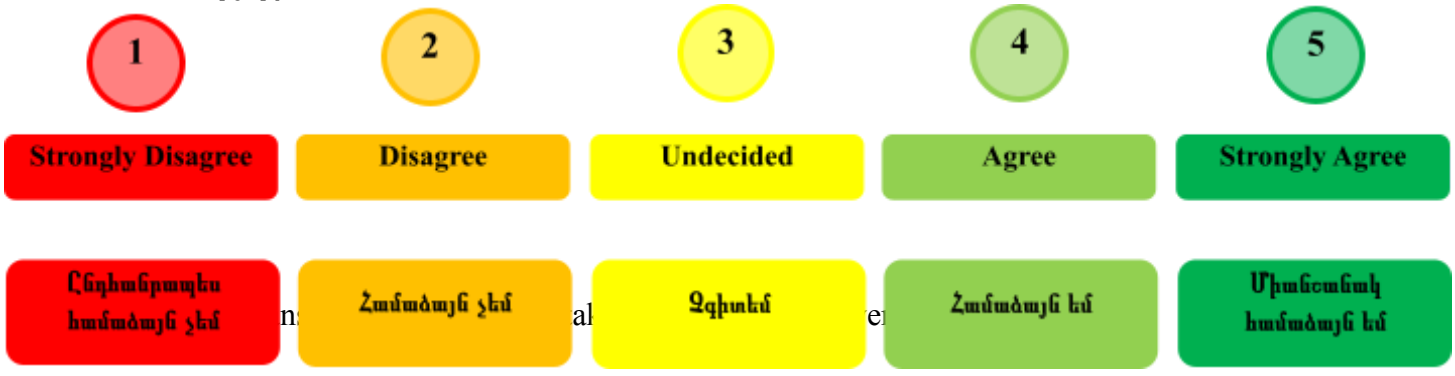
10. Multiple-choice (MC) questions help me to perform better during language tests.

    Ընտրովի տարբերակներով հարցերն ինձ օգնում են ավելի լավ հանդես գալ անգլերեն լեզվի թեստի ժամանակ:



11. I prefer answering short-answer (SA) questions in my English language test.

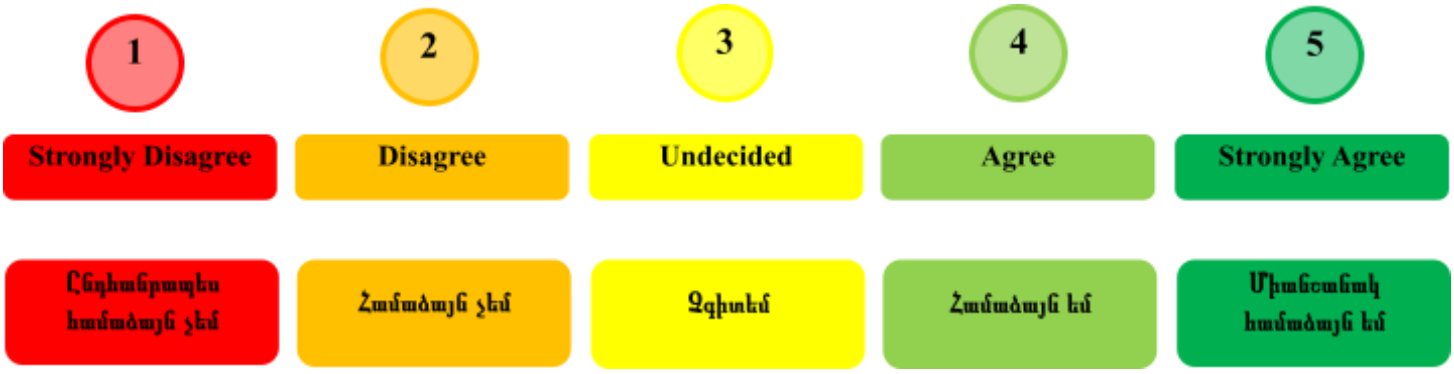
    Ես գերադասում եմ անգլերեն լեզվի թեստի ընթացքում պատասխանել կարճ պատասխանով հարցերին:



Կարճ պատասխանով հարցերին պատասխանելը բավականին ժամանակատար է:

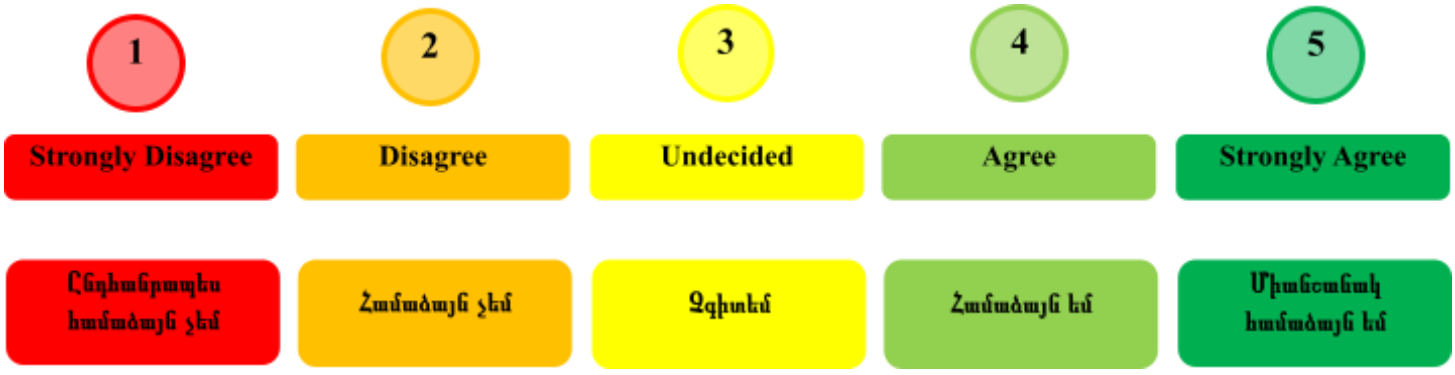
13. Short-answer (SA) questions are easy to answer.

Կարճ պատասխանով հարցերին պատասխանելը հեշտ է:



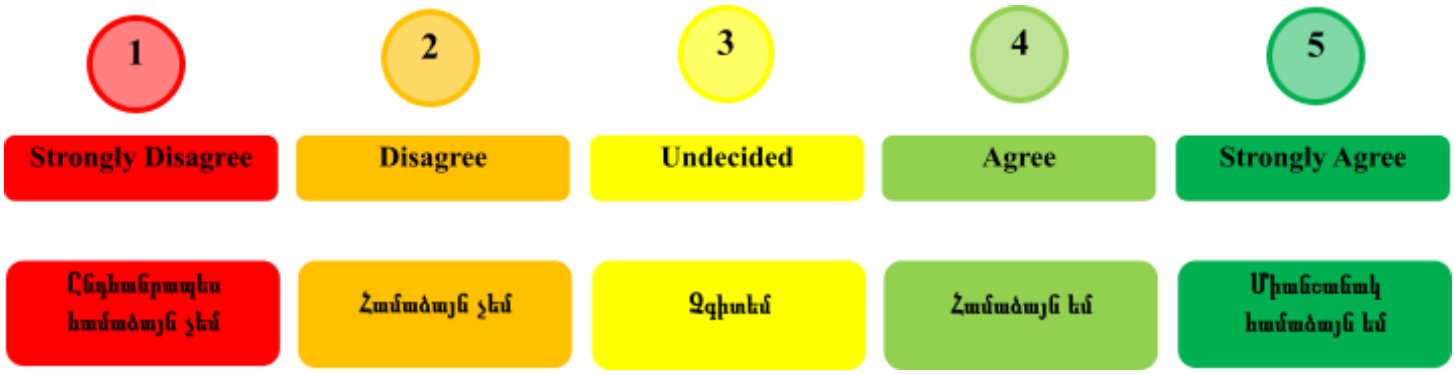
14. Short-answer (SA) questions check my understanding of the course material.

Կարճ պատասխանով հարցերը ստուգում են իմ անգլերենի դասընթացների ընթացքում ձեռք բերած գիտելիքների յուրացումը:



15. Short-answer (SA) provide me information about my language gaps.

Կարճ պատասխանով հարցերը թույլ են տալիս հասկանալ իմ անգլերեն լեզվի գիտելիքների բացթողումները:

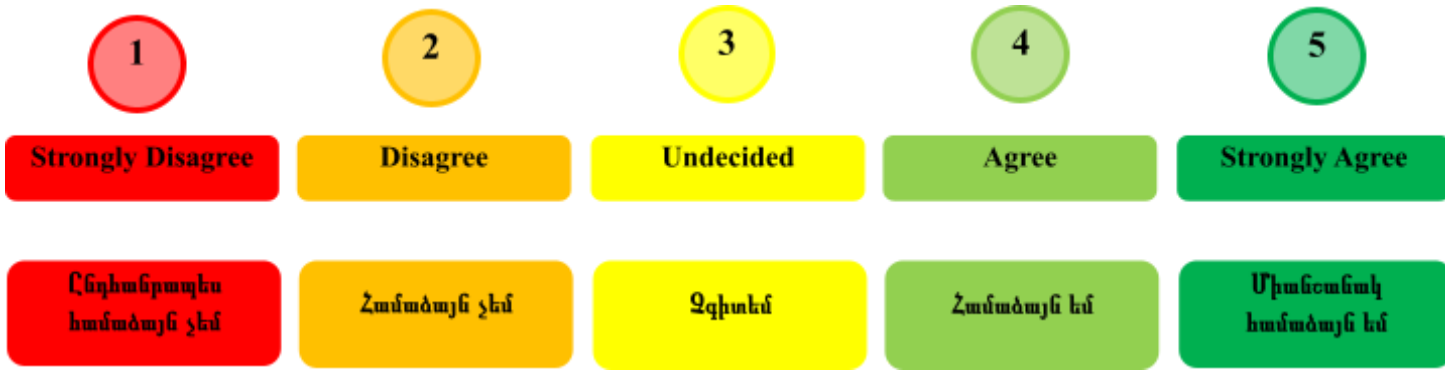


16. Short-answer (SA) questions show my language performance in real-life.

Կարճ պատասխանով հարցերը տրամադրում են իրական կյանքում իմ անգլերենի գիտելիքների օգտագործման ունակությունը:

17. Short-answer (SA) questions help me to perform better during language tests.

Կարճ պատասխանով հարցերն ինձ օգնում են ավելի լավ հանդես գալ անգլերեն լեզվի թեստի ժամանակ:



Appendix H

Interview Questions

1. Which question format allows you to answer faster to the questions?

   Ո՞ր հարցի ֆորմատն է թույլ տալիս ավելի արագ պատասխանել հարցերին:

2. In which question format are you able to answer the question easier?

   Ո՞ր հարցի ֆորմատում ես կարողանում ավելի հեշտ պատասխանել հարցին:

3. In which question format do you feel more confident when answering the questions?

   Ո՞ր հարցի ֆորմատի ժամանակ եք առավել վստահ պատասխանում հարցերին:

4. Which question format expresses your thoughts and ideas the best?

   Ո՞ր հարցի ֆորմատն է ավելի լավ արտահայտում Ձեր մտքերը և գաղափարները:

5. Do you think that MC questions show your comprehension of the material? Why?

   Արդյո՞ք ընտրովի տարբերակներով հարցերը ցույց են տալիս թե ինչքանով եք յուրացրել նյութը: Ինչու՞:

6. Do you think that SA questions show your comprehension of the material? Why?

   Արդյո՞ք կարճ պատասխանով հարցերը ցույց են տալիս թե ինքնաին եք յուրացրել նյութը: Ինչու՞:

7. What are the advantages / disadvantages of MC question format?

   Որո՞նք են ընտրովի տարբերակներով հարցի ֆորմատի առավելությունները / թերությունները:

8. What are the advantages / disadvantages of SA question format?

   Որո՞նք են կարճ պատասխանով հարցի ֆորմատի առավելությունները / թերությունները:

9. Which question format do you prefer more? Why?

   Ո՞ր հարցի ֆորմատն եք նախընտրում: Ինչու՞:

SS	R	Student A	Student B	Student C	Student D	Student E	Student F
6	R1	1.25	1.25	1.25	1.25	1.25	0.625
	R2	1.25	1.25	1.25	1.25	1.25	0.625
12	R1	1.25	0.75	1	0.75	1.25	0
	R2	1.25	0.5	0.75	1	1.25	0
18	R1	0.5	0.25	0.25	0	0.25	0.5
	R2	0.5	0.25	0.25	0	0.25	0.5
19	R1	0.5	0	0	0	0	0
	R2	0.5	0	0	0	0	0
20	R1	0.5	0.25	0.5	0.5	0.5	0
	R2	0.5	0.25	0.5	0.5	0.5	0
21	R1	0.5	0	0.25	0.25	0.25	0
	R2	0.5	0	0.25	0.25	0.25	0
22	R1	0.5	0	0.5	0	0	0.25
	R2	0.5	0	0.25	0	0	0.25
28	R1	0	0.25	0.25	0.25	0.25	0.25
	R2	0	0.25	0.25	0.25	0.25	0.25
29	R1	0	0.25	0.25	0.25	0.25	0.25
	R2	0	0.25	0.25	0.25	0.25	0.25
30	R1	0.25	0.25	0.25	0.25	0	0
	R2	0.25	0.25	0.25	0.25	0	0
31	R1	0	0	0	0	0	0
	R2	0	0	0	0	0	0
32	R1	0	0	0	0	0	0
	R2	0	0	0	0	0	0
38	R1	1.1	1.1	0.85	0	0.31	0.625
	R2	1	1.125	0.75	0	0.31	0.625
44	R1	0.25	0.25	0	0	0.75	0
	R2	0.25	0.25	0	0	0.75	0
46	R1	0.75	0.5	0	0	0.5	0
	R2	0.75	0.5	0	0	0.5	0
52	R1	0.85	0.9	0.85	0	0.2	0
	R2	0.9	1	0.9	0	0.2	0.125
58	R1	1.25	0	1.125	0.9	0.375	1.125
	R2	1.25	0	1.125	0.85	0.375	1.125

Appendix I

Inter-rater Reliability Comparison Group

Green – task is not attempted

SS	R	Student A	Student B	Student C	Student D	Student E	Student F	Student G
6	R1	1.25	0.9	1.25	1.25	1.25	1.25	1.25
	R2	1	0.5	1.25	1.25	1.25	1	1.25
12	R1	1.25	1	1.25	1.25	1.25	1.25	1.25
	R2	1.25	0.75	1.25	1.25	1.25	1.25	1.25
18	R1	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	R2	0.5	0.5	0.25	0.5	0.5	0.5	0.5
19	R1	0.375	0	0	0	0	0.375	0.5
	R2	0.25	0	0	0	0	0.30	0.5
20	R1	0.5	0.25	0.375	0.5	0.5	0.25	0.5
	R2	0.5	0.25	0.25	0.5	0.5	0.25	0.5
21	R1	0.5	0	0.5	0.5	0.5	0.5	0.5
	R2	0.5	0	0.5	0.5	0.5	0.5	0.5
22	R1	0.5	0.5	0.25	0.5	0.5	0.25	0.5
	R2	0.5	0.25	0.25	0.5	0.5	0.25	0.5
28	R1	0.25	0.25	0.25	0.25	0.25	0	0.25
	R2	0.25	0.25	0.25	0.25	0.25	0	0.25
29	R1	0.25	0.25	0.25	0	0.25	0	0.25
	R2	0.25	0.25	0.25	0.125	0.25	0	0.25
30	R1	0.25	0.25	0.25	0.25	0.25	0	0.25
	R2	0.25	0.25	0.25	0.25	0.25	0	0.25
31	R1	0.25	0.25	0.25	0	0.25	0	0.25
	R2	0.25	0.20	0.25	0.125	0.25	0	0.25
32	R1	0.25	0.25	0.25	0.125	0.25	0	0.25
	R2	0.25	0.25	0.25	0.125	0.125	0	0.25
38	R1	0.625	0.47	1.25	1.25	0.9	1	0.9
	R2	0.47	0.47	1.25	1.25	1	1	1
44	R1	1.25	1.25	0.75	1	1	1.25	0.5
	R2	1.25	1.25	0.75	1	1	1.25	0.5
46	R1	1.25	1	1	1.25	1.25	1	1.1
	R2	1.25	1	1	1.25	1.25	1	1
52	R1	1.17	0.7	1.15	1	1.1	0.31	1.25
	R2	1.12	0.8	1.15	1.12	1	0.31	1.25
58	R1	0.5	1	0.625	1.125	1	0.125	1.25
	R2	0.5	1	0.5	1.125	1	0.125	1.125

Inter-rater Reliability Treatment Group

References for Appendices

Englishpage.com. (n.d.). Conditional exercise: Present unreal conditional. Retrieved March 4, 2021, from https://www.englishpage.com/conditional/conditional2.htm

Englishpage.com. (n.d.). Past perfect and past perfect continuous. Retrieved March 4, 2021, from https://www.englishpage.com/verbpage/verbs13.htm

Environmental Activist David Attenborough. (n.d.). Retrieved February 29, 2021, from https://www.englishclub.com/reading/environment/activist-david-attenborough.php

ESL Cyber Listening Club. (2020, March 22). Blood donations. Retrieved February 22, 2021, from https://www.esl-lab.com/academic-english/blood-donations-script/

ESOL Cources. (n.d.). Giving to Charity at Christmas. Retrieved February 22, 2021, from https://www.esolcourses.com/content/topics/christmas/upper-intermediate/christmas-charity-reading.html

Jiménez, A. H. (2010, February 25). Endangered species (PART 1/5): Reading - ESL worksheet by avidrox. Retrieved April 18, 2021, from https://www.eslprintables.com/vocabulary_worksheets/environment_and_nature/endangered_species/Endangered_species_Part_1_5__374599/

Lingo4you. (n.d.). Exercises on conditional sentences (mix). Retrieved March 3, 2021, from https://www.ego4u.com/en/cram-up/grammar/conditional-sentences/exercises?05

Low Intermediate 3 Final test [PDF]. (n.d.). Yerevan: American University of Armenia.

Puchta, H., Stranks, J., & Lewis-Jones, P. (2011). English in mind 4: Workbook. In *English in mind 4: Workbook* (2nd edition ed., p. 41). Cambridge: Cambridge University Press.

Puchta, H., Stranks, J., & Lewis-Jones, P. (2011). English in mind 4: Workbook. In *English in mind 4: Workbook* (2nd edition ed., p. 28). Cambridge: Cambridge University Press.

Puchta, H., Stranks, J., & Lewis-Jones, P. (2011). English in mind 4: Workbook. In *English in mind 4: Workbook* (2nd edition ed., p. 27). Cambridge: Cambridge University Press.

Puchta, H., Stranks, J., & Lewis-Jones, P. (2011). English in mind 4: Workbook. In *English in mind 4: Workbook* (2nd edition ed., p. 26). Cambridge: Cambridge University Press.

Puchta, H., Stranks, J., & Lewis-Jones, P. (2011). English in mind 4: Workbook. In *English in mind 4: Workbook* (2nd edition ed., p. 24). Cambridge: Cambridge University Press.

Retrieved from Exam English Ltd, A. (n.d.). Exam English. Retrieved March 5, 2021, from https://www.examenglish.com/grammar/B1_reported_speech.htm

Retrieved from *Past perfect passive* [PDF]. (n.d.). ESL worksheets on e-grammar.org.

Test-English. (2020, March 22). Future continuous and future perfect – TEST-ENGLISH. Retrieved February 29, 2021, from https://test-english.com/grammar-points/b1-b2/future-continuous-and-future-perfect/3/

Would and used to - past habits and repeated actions – TEST-ENGLISH. (2020, March 22). Retrieved March 6, 2021, from https://test-english.com/grammar-points/b1-b2/would-and-used-to/2/

Would and used to - past habits and repeated actions – test-English. (2020, March 22). Retrieved March 5, 2021, from https://test-english.com/grammar-points/b1-b2/would-and-used-to/