Macrothink Institute™

International Journal of Linguistics
ISSN 1948-5425
2012, Vol. 4, No. 3

Direct and Indirect Assessments of Listening: An Investigation of the Nature of Short-Answer Questions and Dictation as Different Types of Testing Items

Abdullah Al Fraidan

Assistant professor, English Language Department

King Faisal University, Saudi Arabia

E-mail: afridan@kfu.edu.sa

Nurah I. Al Mulhim

Teaching Assistant, English Language Department

King Faisal University, Saudi Arabia

E-mail: green_ran@hotmail.com

Received: May 25, 2012 Accepted: June 6, 2012 Published: September 1, 2012

doi:10.5296/ijl.v4i3.2200 URL: http://dx.doi.org/10.5296/ijl.v4i3.2200

Abstract

This paper investigates two different methods of testing listening comprehension: the direct test, which consists of short-answer questions (SAQs), and the indirect test, which consists of dictation. The results of the study are localised, as the research is conducted at King Faisal University, a local university in Saudi Arabia, with four students from the Department of English Language. After the data were collected using questionnaires and think-aloud techniques, the information was analysed. The performance of each participant on the direct test was compared with his or her performance on the indirect test. The study was conducted to determine whether the performance of the participants on the indirect test is superior to their performance on the direct test. However, the results showed that the assumption underlying the question must be rejected. Moreover, the results led us to conclude that the indirect testing technique (dictation) is more complex and requires greater cognitive ability than we anticipated.

Keywords: Direct & indirect testing, Listening comprehension, Assessment, Think aloud protocols

1. Direct and Indirect Assessments of Listening

We consider listening and reading to be passive or receptive skills. However, it does not follow that these skills do not involve active participation on the part of listeners or readers; rather, it is simply that there is no visible product of these activities. Because comprehension occurs in the mind, we cannot observe this activity. Writing and speaking also involve mental activity; however, unlike listening and reading, these activities yield a written or spoken product. The occurrence of comprehension in the mind thus presents a special problem for language testers. To demonstrate their comprehension, test takers must complete a task.

The items used to test listening have changed during the last 40 years (Douglas 2003, p. 92). Language test developers have studied and introduced a number of test items that differ in purpose and nature, and both direct and indirect listening test items have developed.

In this study, we will consider only one type of each listening test item: short-answer questions (SAQs) as one type of direct item and dictation as one type of indirect item. The nature of these test items will be investigated, as the different types of items are assumed to yield different test scores. This classification of items as direct and indirect should suggest differences both in the types of processes involved, which have been reported in the literature on listening tests, and in the levels of performance and indicators of the listening abilities of the individuals in question, which are the focus of this study. Our investigation will occur at King Faisal University (KFU) in Saudi Arabia, where we will test the listening comprehension of students.

Before presenting the study, we will review previous studies that have discussed related issues and explain how the current study is related to the literature on listening comprehension.

The purpose of listening tests has changed during the past 40 years (Douglas, 2003, p. 92). The focus has changed from sound discrimination to listening comprehension, both at the sentence level and at the levels of discourse and pragmatics. The issue of communicative competence and the mastery of communication units have yielded what is known as communicative language testing. Regarding communicative language testing, Bachman and Palmer (1996, as cited in Bachman, 2003, p. 64) indicate the need for a correspondence between language test performance and language use. For a particular language test to be useful for its intended purposes, test performance must correspond to language use in non-test situations in demonstrable ways. This communicative approach has led to the criticism of integrative tests such as the cloze test and dictation Weir (2001, p. 9) as failing to communicate any information that is directly related to student performance.

Another major issue that has been raised as a result of communicative testing development is the concept of "authenticity" in both listening texts and tasks. According to McDonough and Shaw (1993, p. 43), authenticity can be described as "a term which loosely implies as close an approximation as possible to the world outside the classroom, in the selection both of language material and of the activities and methods used". The issue of authenticity has been raised in relation to the need for communicative performance. Some scholars argue that

communicative tests do not exist and that tests address only certain aspects of communicative ability (Buck, 2001, p. 23). Communicative testing presents challenges to test designers, and Buck's comment appears to indicate that it is nearly impossible to fully capture communicative performance. However, the authenticity of a text is considered an essential component of the communicative characteristics of the entire task. For instance, Thrasher (2001, p. 14) and Weir (2003, p. 137) consider authenticity to be a critical characteristic of listening texts. Thus, in the current study, we selected the text carefully to ensure its authenticity. In the "Method" section, we have provided a description of the text that was used in this study and the specific characteristics that ensure its authenticity.

Moreover, the bottom-up and top-down processes are believed to be important approaches to the interpretation of texts, whether spoken or written, and successful listeners use both top-down knowledge and outside-the-head knowledge (Nunan, 1995, p. 17). Therefore, listeners are expected to use both top-down and bottom-up listening processes to appropriately and successfully interpret texts. However, we cannot guarantee the application of these approaches to every type of text or test or eliminate the possibility that the use of a single approach will be a key factor in the successful processing or interpretation of a given text. In the bottom-up approach, a listener segments a stream of speech into its constituent sounds. In the top-down approach, a listener uses his inside-the-head knowledge (Nunan, 1995 p.17 & 18). The concept of top-down processing is related to schema-theoretic models, another topic that this approach emphasises. These models are based on *"schema theory*", which suggests that knowledge acquisition and text interpretation occur through the activation of schemata: networks of information that are stored in the brain and that act as filters for incoming information (Alderson, 2000, p. 45). Different types of schemata are used to interpret a given text. One type is "content schemata", which is an individual's general background knowledge regarding a topic. The text that we have provided for this study contains information that is likely to be familiar to the participants and that may be assumed to influence their processing of the text. In the "Method" section, we will elaborate more on the nature of the text that we used and discuss the extent of the participants' knowledge of the subject matter contained in the text.

Listening test items have been classified both in terms of how authentic or communicative they are and in terms of whether they are direct or indirect. The product of listening is not spoken or written responses. Douglas (2003, p. 118) indicates that the product of receptive skills lies within the brain and is thus impossible to observe. Furthermore, he adds that we can observe the result of meaningful input only in the form of spoken or written output.

Although all types of tests can be considered indirect indicators of an individual's ability, test designers tend to classify tests as either direct or indirect. Direct measures include student products or performances that demonstrate that specific learning has occurred. In contrast, indirect measures (e.g., student perceptions of learning) may imply that learning has occurred but do not specifically demonstrate this learning or skill.

Dictation is one type of indirect listening. In this study, we are interested in comparing dictation to short-answer questions (SAQs), which comprise a direct test. According to

Thrasher (2000, p. 7), Oller (1987) claimed that this testing type was a good example of what he termed a "pragmatic test". He defined a pragmatic test as "any procedure or task that causes the learner to process sequences of elements in language that conform to the normal textual constraints of that language". Thus, according to Oller, dictation is beneficial because it employs text that is longer than a single sentence and allows test takers to use their knowledge of the world and other context clues to reconstruct a message. Oller is likely referring to one method of conducting dictation. However, to understand the nature of dictation and the types of processes that are involved in this type of testing, we must recognise that there are two ways of administering a dictation test. Buck (2001, p. 77) considers Oller's view of dictation to be "unconvincing" because he does not consider dictation to be "a pragmatic test" and does not believe that it reflects real life. He further explains that dictation works in a number of ways: both the length of the segments and the degree to which they challenge test takers can vary. He claims that short segments (fewer than seven chunks of information) do not challenge listeners and that writing ability is less important than word recognition in these instances. In contrast, when chunks or segments become longer, listeners must rely on their short-term memory to store segments and then write them down (Buck 2001, p. 77 & 78). Thus, the difficulty of this technique can be manipulated by the length of segments (or "bursts", as they are technically known). Therefore, the first level on which dictation works will be the easiest to examine. Buck's argument will consequently be used as one of the theoretical frameworks of this study.

Dictation, then, requires the integration of both listening ability and writing ability. At the second level of listening, this task becomes integrative or "quasi-extensive" in that it requires a certain degree of linguistic sophistication to produce correct responses (Douglas, 2003, p. 131). The cognitive processes that are required for dictation are clearly important at all levels, whether such processes involve mere word recognition and memorisation or whether they are real interpretations that require the use of linguistic ability. However, according to (Buck, 2001), "the question in focus is how well these processes operationalises the listening construct" (p. 79).

The other testing technique that is of interest in this study is short-answer questions (SAQs). The SAQ test is a type of direct listening test that, unlike dictation, involves three ability clusters: listening comprehension, writing, and reading. Consequently, the absence of one of these abilities or a deficiency in one of these areas may influence individual performance on such a test (Thrasher, 2000, p. 12). Moreover, the complexity of this testing item results from the abilities that are required to complete the task, the order and types of questions, and the reliance on note-taking. Thrasher (2000) offers the following observation:

"SAQ tasks involve more than merely listening and writing. There is the necessity to summarise or identify the main points of what was said. This is a separate skill that even some native speakers never develop" (p. 11).

The order of questions is another factor that is assumed to influence listener performance and that may affect the reliability of such tests. Thompson (1995) explains as follows:

"It seems widely acceptable that questions will ask for information in the same order it occurs in the passage … Test developers who break these conventions may confuse test takers, which could lead to unreliable performance" (as cited in Buck, 2001, p. 139).

The final factor that contributes to the complexity of SAQs is the types of questions that are asked. SAQs are considered a realistic means of testing listening comprehension (Weir, 2003, p. 140), especially in instances in which it is possible to simulate real-life activities by creating a written record of the explicit main ideas and important details of a spoken message (e.g., we a lecture), as is the case in this study. Although writing ability may interfere with listening comprehension in SAQs, test designers are responsible for specifying and limiting written responses through the types of questions that are used. Questions that require a greater amount of propositional inferencing (passage dependence) require designers to determine in advance what is considered an acceptable interpretation of the text and what constitutes an adequate written response (Weir, 2003, p. 140).

The practice of previewing questions before listening to a text is controversial. There are two perspectives on the practice of previewing questions with students, and it has been argued that the practice may have a negative effect on construct validity, as suggested by Jane Sherman (1997, p. 121). However, when Sherman conducted her study, she found that previewing questions actually proved more useful than expected. Because of these results, we decided to present the questions to the subjects before they began listening to the test text.

A *construct* is an underlying ability that has been hypothesised in language proficiency theory (Hughes, 2003, p. 5). The main construct on which this study focuses is listening comprehension. However, this construct is general and must be further explained to ensure a reliable and valid assessment.

Listening comprehension is a skill that is difficult to separate into its constituent parts. Listening comprehension involves extracting meaning from spoken language, and all of the sub-skills that constitute listening are subservient to this quest for meaning. However, there is no consensus in the literature on listening skill regarding the precise definition of listening comprehension ability.

Bachman (2003, p. 51) discusses the difficulty of choosing a theoretical framework for classifying listening comprehension sub-skills. Bachman argues that although this variability in the classification of listening comprehension ability has not received significant empirical support, considering this variation does help to provide a general perspective regarding the constituents of this skill.

One of the frameworks that are intended to classify these sub-skills is that provided by Richards (1983), who argues that the purpose of the instance of listening affects the sub-skills that are used. For instance, the sub-skills that are used for social actions differ from those that are used for "academic listening" and "listening for leisure time" (as cited in Hughes, 2003, p. 133).

For this study, we used Richards' taxonomy for classifying sub-skills. The type of listening is clearly academic listening; therefore, we chose the following specific constructs from the sub-skills that Richards provided for this purpose:

1-The ability to identify the topic of a lecture

2-The ability to identify relationships among units within discourse and follow topic development

Given that the segments that are used in dictation tests are typically short and following Buck's view of this type of dictation in addition to the complex nature of SAQs as previously discussed, we propose the following questions:

Is the performance of test subjects on an indirect test superior to their performance on a direct test?

2. Method

2.1 Participants

The participants in this study are 4 female students (age: 21 years) at King Faisal University in their first year in the Department of English Language. All of the participants are our students and are enrolled in a core course in listening comprehension as part of their specialisation in English. I chose these four participants because two of these students are expected to perform better in our listening class and are considered excellent students, whereas the other two students are expected to demonstrate lower academic performance.

2.2 Materials and Procedure

2.2.1 The Text

We used a text that is not considered formal in style and that captures the characteristics of everyday speech that the subjects may encounter outside of the classroom -- for instance, when listening to radio broadcasts or interacting with native speakers. The text is a lecture on "entrepreneurs" that was obtained from a listening comprehension course book that students utilise in their first year of English. In sum, we provided an "authentic text".

The nature of the text is important because it determines its content validity and whether "beneficial backwash" is achieved. Therefore, we have provided a simple description of the text that is used in this study, including the following:

1-Text type: monologue, planned, structured and transactional in function

2-Text form: descriptive and expository. The text explains the meaning of "entrepreneurs" by providing a definition and well-known examples. In addition, the text describes the achievements of these entrepreneurs and the similarities and differences among them.

3-Length: the duration of the text is 4 minutes and 16 seconds.

4-Speed of speech: 26 words per minute

5-Dialect: American dialect

Adapted from Hughes (2003)

As previously noted, successful listeners are those who use both the top-down and the bottom-up approach. One part of the top-down approach is the use of knowledge of the subject matter, which is "entrepreneurs" in this study. The recorded text that is provided to the participants includes entrepreneurial names such as"Yahoo.com" and "The Body Shop"; when the participants were asked about these two examples to activate their schemata before they listened to the passage, they indicated their familiarity with them.

2.2.2 Thinking-Aloud Technique

This type of technique is used in the direct (SAQ) test when the subjects listen to the text for the second time and write down their responses. "Thinking-aloud techniques" (or, as they are sometimes known, "think-aloud protocols") ask participants to think aloud while performing a set of specified tasks. Users are asked to describe what they are observing, thinking, doing, and feeling as they complete their tasks. This approach enables observers to directly observe the process of task completion rather than only being privy to its final product (a definition cited from Wikipedia.com). The use of this technique is limited to research; therefore, we explained the basic concept to the subjects, although I did not train them in the use of the technique.

2.2.3 The Tests

2.2.3.1 The direct test (SAQs)

The participants were given the direct test, which contains SAQs that the participants were asked to answer after they had listened to the recorded text twice. The first time that they listened to the text, the subjects were asked to take notes that would not be marked as in their regular listening course (because note taking is considered a different skill and sometimes a different testing technique) (Douglas, 2003, p. 129). During the second listening period, the participants were asked to listen and write down their responses. The participants were tested individually because they would be asked to verbalise their thoughts for each question; that is, they would be asked to use the "thinking-aloud technique" in writing their responses, which were to be recorded for data analysis. For this reason, during the second listening period, the tape was stopped 3 times to allow me to record the subjects' thoughts. The participants were given 5 minutes to take notes and were then given 10 minutes to write their responses to the questions after the tape was stopped.

2.2.3.2 The Indirect Test (Dictation)

The indirect test, which involves the use of the dictation technique, was conducted after 2 days to reduce the familiarity of the participants with the text content and to limit possible memorisation effects. (More than 2 days would have been preferable in reducing the memorisation effect; however, the available time was limited, especially because we needed to examine each subject individually due to class schedule conflicts.) During the test, the participants were asked to listen to the text three times. The first listening period was

intended to induce a *general idea* (or the "*gist*") of the text. During the second period, the subjects were asked to begin writing down words and phrases to make as many meaningful constructions as possible. Eventually, they were asked to write down the entire text. The tape was stopped for every generic structure (internal discourse structure): the introduction, the definition of the term 'entrepreneur,' the characteristics that entrepreneurs share, the possible differences among them and the conclusion. The pauses generated segments of different lengths or bursts that the participants were asked to write down. The shortest burst contains seven content words, whereas the longest burst contains 14 such words. In the third listening session, the participants reviewed all of the text that they had written down. After the first listening period, they were given 3 minutes, whereas after the second and third listening periods, they were given 10 minutes to create their own versions of the text.

A brief review of both types of tests as they pertain to studying and teaching listening skills at King Faisal University may be helpful.

Dictation. We administered a questionnaire to determine whether our students and the instructors in the English Language Department are familiar with this testing item. The questionnaire yielded the following results:

1-The students in the Section 1 listening course have never been introduced to this technique. Therefore, we trained our students on how to take such a test to ensure that their limited experience would not generate a test-item effect.

2- This technique is not commonly used amongst listening comprehension teachers at KFU. The results show that only one instructor uses this technique for sound discrimination purposes.

As mentioned, although this type of technique is infrequently used in the context of teaching and learning listening at KFU, we have chosen it because the literature review revealed that many linguists have suggested the need to investigate the usefulness of this testing technique.

SAQs. Using the same questionnaire that is described in Appendices 1 and 2,we reached the following conclusions regarding SAQs:

1- SAQs are the type of testing technique with which students are most familiar in listening tasks.

2- This type of test is the most commonly used technique among English language teachers who teach listening courses at KFU.

Because of the highly frequent use of and familiarity with this technique at KFU, we decided to use SAQs as the only type of direct testing in this study.

2.2.4 The Rating Scale

2.2.4.1 Direct Test

For both the direct and indirect tests, we used numerical rating scales. In the direct test, the total mark is out of 10. Q1 and Q2, which are marked on a scale from 1 to 5, assess the ability

of the students to identify the topic of the lecture. The remaining questions assess the ability of the subjects to identify the relationships among the units and to follow topic development, and these questions are again marked on a scale from 1 to 5.

2.2.4.2 Indirect Test

The indirect dictation test is marked on a scale from 1 to 24. This test is marked semantically rather than lexically. In other words, the participants are marked on the basis of how many meaningful word clusters or clauses they form. For this test, we followed Weir's description (2003, p. 135) of this type of marking. Weir (2003) states that "[m]arking may be problematic if one wishes to adopt a more communicatively oriented marking scheme" (p. 135). Although this type of marking is considered problematic because it is not word-for-word marking and may thus increase subjectivity, this issue can be overcome if those doing the scoring provide instructions regarding how to take such tests (Weir, 2003, p. 135). We actually did provide such training during our discussion of the choice of dictation. As with most listening tests, no marks were deducted for the use of recognisable standard or personal abbreviations; the omission of communicatively redundant items, such as articles; or mechanical errors related to grammar, punctuation or spelling.

This choice of marking style is a function of the specific sub-skills that we assessed, which require marking based on the number of sentences that the students actually write down to create a version that is similar to the original text. The same criterion and construct division on which the direct test was based was also used for the indirect dictation test. Therefore, we based our marking on how the participants manifested the two sub-skills or constructs: (1) their ability to identify the topic by writing down the word "entrepreneurs" with a definition and examples (e.g., "Jeff Bezos and Jerry Yang") and their contributions was marked on a scale of 1 to 12 ; and (2) their ability to identify the relationships among the units was marked using the same scale, and the participants were instructed to write down the similarities and differences among the entrepreneurs.

We used the same scale in the direct test to obtain a verbal description of the performance of the participants and to ensure a common basis for comparing the findings of the direct and indirect tests. Therefore, the ability to identify the topic was marked using a scale of 0 to 12. A mark of 0 was assigned to the category of "not found", marks 1-2 were assigned to the "very poor" category, marks 3-5 were assigned to the "average" category, marks 6-8 were assigned to the "good" category, marks 9-10 were assigned to the "very good" category, and marks 11-12 were assigned to the "excellent" category. The same marking system will be used to assess the ability of the subjects to identify the relationships among the units. To avoiding confusing readers, we will refer to the first scale (0-5) as scale 1 and the second scale (0-12) as scale 2.

2.2.5 Procedure

The procedures used with the two types of tests differed slightly. For the SAQ test, the participants were asked to answer the questions that were provided within the recorded text, and as they attempted to answer these questions, their responses were recorded using the

"thinking-aloud technique". For the dictation test, in contrast, the students were asked to write down as many meaningful words and phrases as possible and eventually to write down the entire text; again, as they attempted to write, they were asked to employ the "thinking-aloud technique". For both tests, the tape was played three times, and the first time was generally intended to enable the participants to grasp the gist of the text. However, unlike with the SAQ test, in the dictation task, the tape was stopped for every generic structure during the second listening period to test the ability of the participants to write down the different series of words, sentences and phrases. Finally, a numerical scale was used for both tests as a means of quantifying the performance of the participants.

3. Results

3.1 SAQs

For the direct test, we were able to record the participants' responses using the "thinking aloud technique". A transcription of each participant's response is presented below, followed by an analysis. Each participant was examined individually, and we did not provide any assistance because the purpose was to attempt to investigate how the students actually think. The tape was stopped when the participants had finished answering the questions.

3.1.1 The First Participant

The first participant did not answer any of the questions correctly except for Q3, for which her response was partially correct in that she supplied the word 'vision'. Therefore, she was given only half a mark (out of a total of one mark) for answering part of the question. The first question was difficult for her, likely because she experienced difficulty reading the keyword "entrepreneurs". Her difficulty reading the questions or some of the words in the questions likely resulted in some difficulty processing the questions and providing answers to Q1 and Q4.

For some questions, such as Q2 and Q4, the participant expressed her lack of knowledge or understanding by providing responses such as "I don’t know".

She expressed her understanding of Q5 by providing an Arabic equivalent; however, she failed to provide the appropriate answer and thus demonstrated her inability to follow the development of the topic.

3.1.2 The Second Participant

The second participant answered Q1 correctly, indicating her ability to identify the topic. However, she failed to follow the development of the topic and to identify the relationships among the units. She provided only part of the answer to Q3; thus, she received only half of a mark for this question (out of a total of one mark).

3.1.3 The Third Participant

The third participant answered most of the questions correctly. She provided the correct answers to Q1 and Q2, which indicated her ability to identify the topic, although her answer to Q2 was incomplete. The remainder of the questions were also answered correctly, except

for Q4, for which the participant provided a general answer rather than a specific example that illustrates the difference. Therefore, this subject was not given a full mark for answering this question.

3.1.4 The Fourth Participant

This participant's answers to Q3 and Q4 clearly indicate that she understood the idea at some point but lacked the equivalent English expression to appropriately express the idea.

In addition, the participant was sometimes unable to read a word, which could have influenced her performance because a word that cannot be read clearly cannot be understood. This situation occurred in Q4, as the participant did not know how to read the words "factors", "entrepreneurs" and "differ" and thus stated that she did not know the answer. Although this participant provided the appropriate Arabic equivalent response, as in Q3 and Q4, we did not mark her answer as a correct response because it was not sufficient for her to demonstrate her understanding of the questions in her L1; rather, she was required to also use written English in this context.

The test is marked out of 10, with 5 marks for the first construct and 5 marks for the second construct. In addition, each mark indicates a verbal description of the performance of the participants. The tables below present the numerical performance indicators for each participant on this test.

The first participant

The construct	not found	very poor	average	good	very good excellent
1-The ability to identify the topic of a lecture	**0**	1	2	3	4 5
2-Identify relationships among units	0	**1**	2	3	4 5
		The total mark			1

The second participant

The construct	not found	very poor	average	good	very good	excellent
1-The ability to identify the topic of a lecture	0	1	**2**	3	4 5	
2-Identify relationships among units	0	**1**	2	3	4	5
		The total mark				3

The third participant

The construct	not found	very poor	average	good	very good	excellent
1-The ability to identify the topic of a lecture	0	1	2	3	**4**	5
2-Identify relationships among units	0	1	2	3	4	**5**
The total mark						9

The fourth participant

The construct	not found	very poor	average	good	very good	excellent
1-The ability to identify the topic of a lecture	0	1	2	**3**	4	5
2-Identify relationships among units	0	1	**2**	3	4	5
The total mark						5

The following descriptive account is provided based on this table:

1- The first participant did not demonstrate the ability to identify the topic of the lecture, and her ability to identify the relationships among the units was very poor.

2- The ability of the second participant to identify the topic of the lecture was average, whereas her ability to identify the relationships among units was very poor.

3- The ability of the third participant to identify the topic of the lecture was very good, and her ability to identify the relationships among units was excellent.

4- The ability of the fourth participant to identify the topic of the lecture was good, whereas her ability to identify the relationships among units was average.

3.2 Dictation

Most of the participants were able to write down the short segments correctly, except for the third and fourth participants, who demonstrate some difficulty writing down some of the long segments. The first and second participants also demonstrated some difficulty writing down the short bursts; however, most of the difficulty that they experienced was related to the long bursts. Some of the long sentences were also missing words that were also absent from the

responses of the first participant. Because "semantic scoring" was used, only half of a mark was given for such constructions.

The following presents illustrates the participants' marks on the indirect test for both sub-skills based on scale 2.

The participant	Identifying the topic of the lecture	Identifying the relationships among the units	The total mark
First participant	4	3	7
Second participant	6	10	16
Third participant	12	8	20
Fourth participant	4	5	9

The tables below show the results for scale 1.

The first participant

The construct	not found	very poor	average	good	very good	excellent
1-The ability to identify the topic of a lecture	0	1	**2**	3	4	5
2-Identify relationships among units	0	1	**2**	3	4	5

The second participant

The construct	not found	very poor	average	good	very good	excellent
1-The ability to identify the topic of a lecture	0	1	**2**	3	4	5
2-Identify relationships among units	0	1	2	3	**4**	5

The third participant

The construct	not found	very poor	average	good	very good	excellent
1-The ability to identify the topic of a lecture	0	1	2	3	4	**5**
2-Identify relationships among units	0	1	2	**3**	4	5

The fourth participant

The construct	not found	very poor	average	good	very good	excellent
1-The ability to identify the topic of a lecture	0		1	**2**	3	4 5
2-Identify relationships among units	0	1	**2**	3	4	5

The following descriptive account of the performance of the participants on the indirect test is based on scale 1:

1-The ability of the first participant to identify the topic of the lecture and to identify the relationships among units was average.

2- The ability of the second participant to identify the topic of the lecture was average, whereas the participant's ability to identify the relationships among units was very good.

3-The ability of the third participant to identify the topic of the lecture was excellent, and her ability to identify the relationships among units was good.

4-The ability of the fourth participant to identify the topic of the lecture and the relationships among units was average.

4. Discussion

The results generate significant insights. The first participant did not demonstrate an ability to identify the topic of the lecture by answering Q1 and Q2 because her difficulty reading a word in Q1 likely influenced her subsequent performance on Q2. The participant's ability to identify the relationships among units and follow the development of the topic was very poor. Another major problem that generated this low performance was difficulty reading and processing the question and consequently providing an appropriate answer, as in Q4. The same factor was at play in the fourth participant's answer to Q4; again, this problem influenced her ability to follow the development of the topic.

Although writing ability can be strongly reflected in questions that require inferences regarding meaning, this ability is also reflected in answers to simple SAQs. The writing ability that is required to answer SAQs can also be observed in the performance of the fourth participant. Although she did not answer the third and fourth questions correctly, she provided some Arabic equivalents. Nevertheless, this actually demonstrates the participant's weak ability to identify the relationships among units and to follow the development of the topic using the necessary grammatical elements. Of course, a deficiency in either reading or writing ability alone does not indicate weak performance in such a task. Consequently, the second subject experienced difficulty reading Q1 and Q6 but still answered these questions correctly.

In analysing the results, we can observe that the overall performance of the participants and their total scores differ for the SAQs and the dictation test. The first and second participants performed better on the indirect test, whereas the third and fourth participants performed better on the direct test than on the indirect test. Although the third and fourth participants still demonstrated stronger performance on the indirect test than did the other participants, our purpose is only to compare the performance of each subject in both tests.

In analysing the sub-skills under investigation for the first and second participants, we observed that the first and second abilities (1) ability to identify topic and 2) ability to identify relationship) are generally stronger and indicate higher-level skills on the indirect test than on the direct test. We can discuss these final results with reference to two factors: the nature or types of questions and the order of the questions on the SAQ test. Q1 and Q2, which reflect the ability of the participants to identify the topic, are straightforward and are clearly stated by the speaker. However, these questions were not answered by the first subject and were not answered completely by the second subject. The other questions were not straightforward; rather, they required the ability to connect and follow the different generic structures of the text. The order of the questions should also be considered because it can confuse students, hinder their performance and generate unreliable responses (Thompson 1995, as cited in Buck, 2001 p. 139). In this study, some of the questions were not arranged according to the order in which the information was presented in the passage. This factor may account for the low performance and listening abilities of the first and second participants on the direct test. However, dictation operationalises listening in the narrower of the two levels of listening (Buck, 2001, p. 77). The participants were asked to listen to segments of speech that were relatively short compared with the long flow of speech that they heard during the direct test. In the dictation test, as we noted previously, the tape was stopped at different intervals to create sentences of varying length. However, the number of short segments or sentences was greater than the number of longer segments or sentences. All of the participants were somewhat successful in processing and writing down the short bursts or segments, but their performance in processing the longer segments varied. Like Buck, who suggests that the dictation of short, segmented speech requires only word recognition, Ellis (2001, p. 106) believes that in this type of dictation, listeners "rely on their short-term memory"; thus, listeners merely memorise sentences. Moreover, the use of linguistic ability to link words and units together, which is an ability that more advanced learners utilise in much longer segmented speech as a result of the limits on the capacity of working memory, is not employed when one listens to short, segmented speech. In fact, such arguments explain the stronger performance on the indirect test as compared with the direct test, as observed for the first and second listening abilities among the first and second participants on the indirect test. Furthermore, the third participant demonstrated stronger capabilities with regard to the first skill on the indirect test, whereas her second ability was found to be stronger when evaluated using the direct test. The abilities of the third participant are another matter: her first ability was found to be better on the direct test, but her second ability did not change; rather, it was average on both tests. The total scores of the third and fourth subjects contrasted with those of the first and second subjects. The total scores of the third and fourth subjects were higher on the direct test. However, poor performance on the indirect test was sometimes

observed among these individuals for the short segments that require only word recognition. (In contrast, most of the long segments required a greater reliance on short-term memory) We have stated that dictation in this context is believed to require the mere memorisation of words or mere word recognition for short bursts (phrases containing seven chunks or fewer). However, the number of words or chunks of information in a single burst is not significant unless it is related to the length of the words themselves (Ingram, 2007, p. 121). If a single burst consisted of six long content words, this burst would be more difficult to process than other bursts consisting of the same number of shorter words.

These findings provide an answer to the question of whether the overall performance of the subjects is stronger on the indirect test than on the direct test, the answer is no. The differences in the nature of the two types of tests in this study were assumed to influence the performance of the listeners, as the latter were expected to perform better on the indirect test. This is similar to Al Fraidan (2011) justification with similar subjects when they did different direct and indirect writing tests. However, as noted previously, the results lead us to reject this assumption and thus provide us with the answer to the question underlying this research.

The familiarity of the students with some elements of the topic, such as "FedEx" and "Yahoo.com", did not have a positive effect on their performance, as the first and second subjects received very low scores on the direct test. This result is consistent with the observation regarding top-down processes and prior knowledge that was offered by Saville-Stoike (2006): "the unconscious and automatic access that listeners have to prior knowledge of content may be inappropriate and could account for some misunderstanding" (p. 160).

5. Conclusion

This study has several limitations and has yielded several significant findings. The complex nature of listening comprehension contributes to the complexity of its sub-skills and, consequently, to the types of testing items that are used to assess them. This investigation highlights the nature of direct and indirect testing items and the different processes that are involved in this testing, and our interpretation of the results reveals differences in student performance on the different testing items. This performance difference may be predictable, but it is assumed in this study that the indirect type of testing used is likely the actual cause of its remarkable significance. However, the results of this investigation show that students vary in their performance and that indirect dictation tests are likely more complex and impose a greater cognitive burden (as a result of the length of the words and bursts) than we have assumed.

References

Al Fraidan, A. (2001). Direct and indirect assessment of writing. Lab Lampert. Germany

Bachman and Palmer. (1996). in Bachman, L (2003). Fundamental considerations in Language testing, Sixth edition. NY: Oxford University Press.

Brown, H. Douglas (2003). Language Assessment: Principles and classroom Practice. White plains, NY*: Longman*.

Buck, G. (2001). Assessing listening. Cambridge, U.K: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511732959

Charles, J. Alderson (2000). Assessing reading. Cambridge, U.K: Cambridge University Press.

Davis, A. (1996). "The competing claims of accuracy and fluency in the Construction of performance tests in language proficiency". *Melbourne Papers in Language Testing*, *12*(16), 1-22.

Ellis, R.C. (2001). Memory for language. Cambridge, U.K: Cambridge University Press.

Hughes, A. (2003). Language testing for teachers. Cambridge, U.K: Cambridge University Press.

Ingram, J. (2007). Neurolinguistics: An introduction to spoken language processing and its disorders . Cambridge , U.K :Cambridge University Press.

McDonough, J., & Shaw, C. (1993). Materials and methods in ELT BlackwellRichards, J. C. and Lockhart, C. (1994). Reflective teaching in second language classrooms. CUP.

Nunan , D. Language teaching methodology. Hertfordshire , U.K :Phoenix LTD.

Richards. (1983) in Hughe (2003). Language testing for teachers. Cambridge, U.K: Cambridge university press.

Saville-Troik, M. (2006). Introducing second language acquisition. Cambridge ,U.K: Cambridge University Press.

Sherman, J. (1997)."The Effect of question preview in listening comprehension tests". *Testing Journals*, *4*(16), 12-37.

Thomson. (1995). Reflective teaching in second language classrooms in Buck, G. (2001). Assessing listening .Cambridge, U.K : Cambridge University Press .

Thrasher, R (2000). The testing of listening comprehension. NY: International Christian University Press

Weir , C. (2005).Language testing and validation .London: Palgrave McMillan

Wikipedia (available at: http//www.wikipedia.com)(‘viewed on 1 /1/2012).

Appendices

The following appendices are listed according to the numbers provided within the chapters of this report. The appendices related to the performance of the subjects are arranged according to the order in which they participated in the study

Appendix 1

Questionnaire

Have you taken this type of test before?

Yes ( )

In which learning stage did you previously take this type of test?.....

No ( )

Appendix 2

Place a checkmark beside the test types with which you are familiar with from studying listening throughout your stages of learning.

Dictation …… Matching items……..

Multiple-choice items ……. True or False……..

Complete the sentences ……..Short-answer questions ……

Others……
……..

Appendix 3

Questionnaire

Place a checkmark next to the inventions with which you are familiar:

Yahoo.com ( )

FedEx ( )

Body Shop ( )

Amazon.com ( )

Write about what you know in your L1 (Arabic).

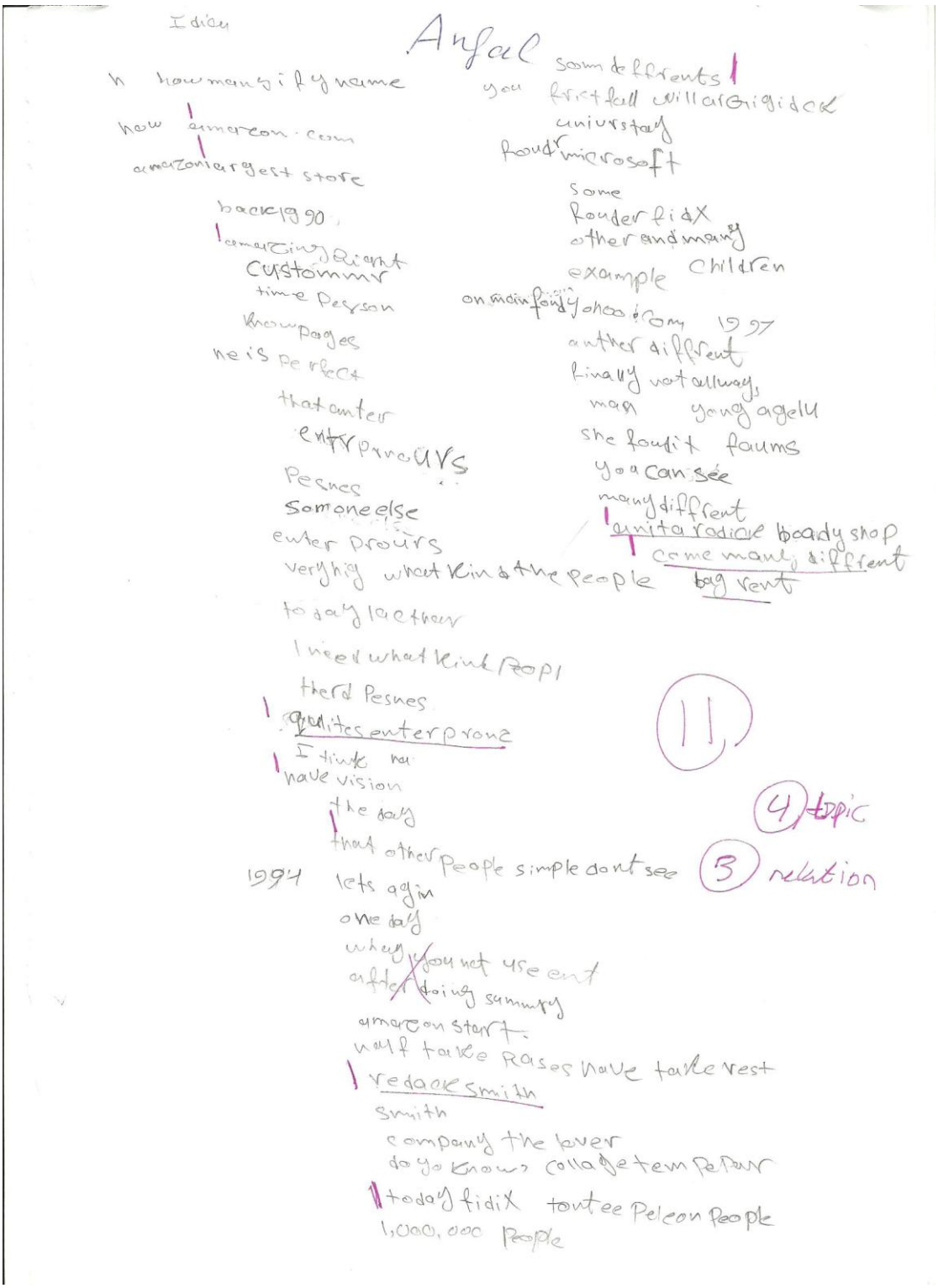
……
……
……
……
……..

Appendix 4

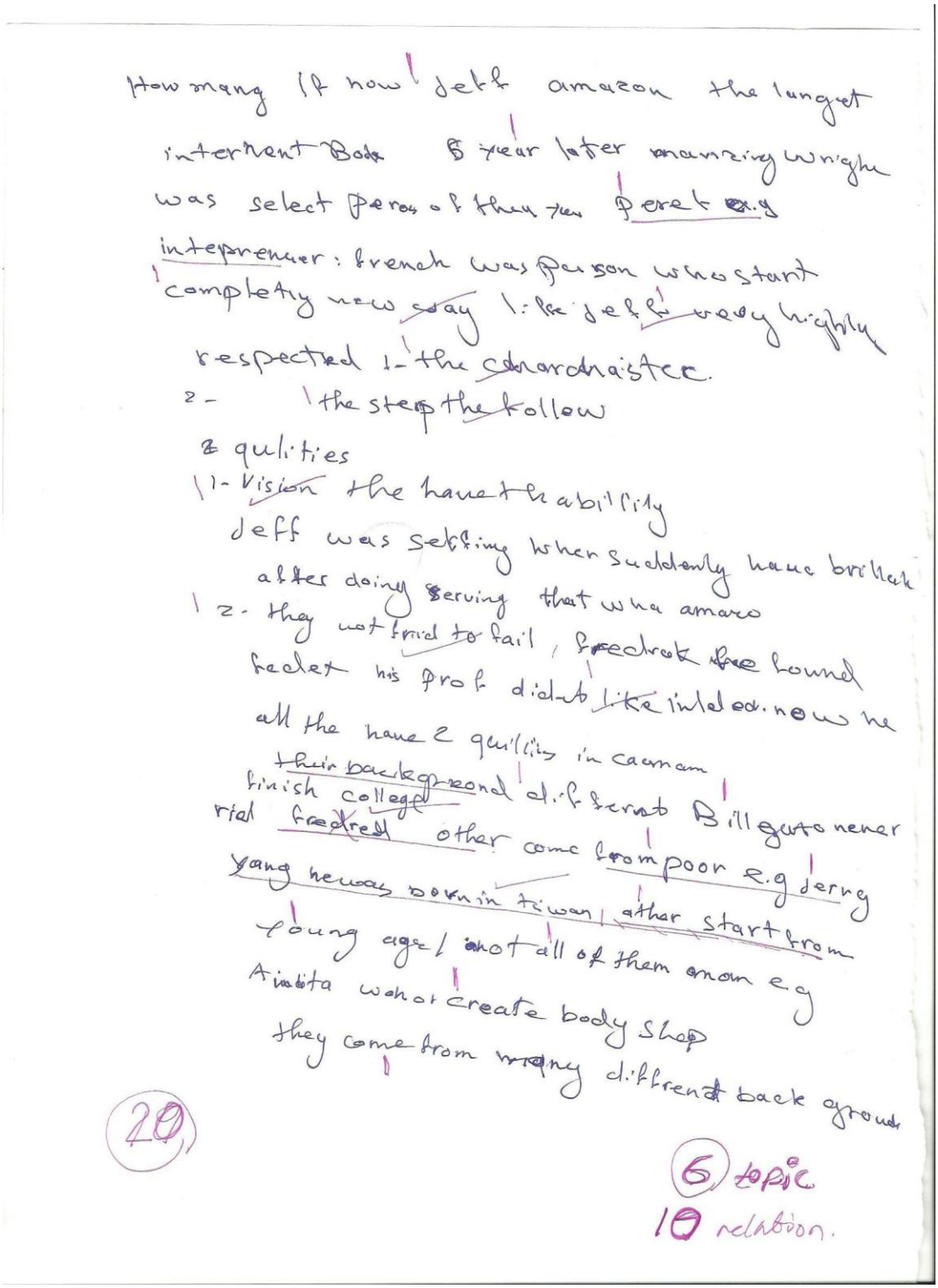
This section includes the following:

-The performance of the students on the indirect test (the names displayed on the papers displayed represent the first subject, the second subject, the third subject and the fourth subject in that order.)

-The transcriptions of the students' use of the thinking-aloud technique on the direct test, which are followed by their performance on the papers



International Journal of Linguistics
ISSN 1948-5425
2012, Vol. 4, No. 3



How many if now Jeff amazon the langest
internent Bode 5 year later manzing wright
was select peros of them yes peret e.g
intepreneur: french was person who start
completly new way like Jeff very highly
respected 1- the charchaister.
2- the step the follow
2 qulities
1- Vision the have the abillity
Jeff was selfing when suddenly have brillah
after doing serving that wha amazo
2- they not fried to fail, fredrok the found
fedex his prof didnt like inlded. now he
all the have 2 quillity in caman
their background different Bill gato never
finish college
rial fredred other come from poor e.g Jerry
yang he was born in tiwan, ather start from
young age / not all of them anom e.g
Anita wah or create body shop
they come from many diffrent back ground

20

6 topic
10 relation.

Enterpeneurs.

Jeff bazo is the Founder of amazon site.
Amazon is the largest books store.

- A person who start a completely new business or trading industry and some who does something no one has done before

characteristics the enterpeneurs have:

1) They have vision. what that means they see opportunity which people can't see
2) They're not afraid to take risks. like Fredrick smith the Founder of Fedex. the company that delivers packages

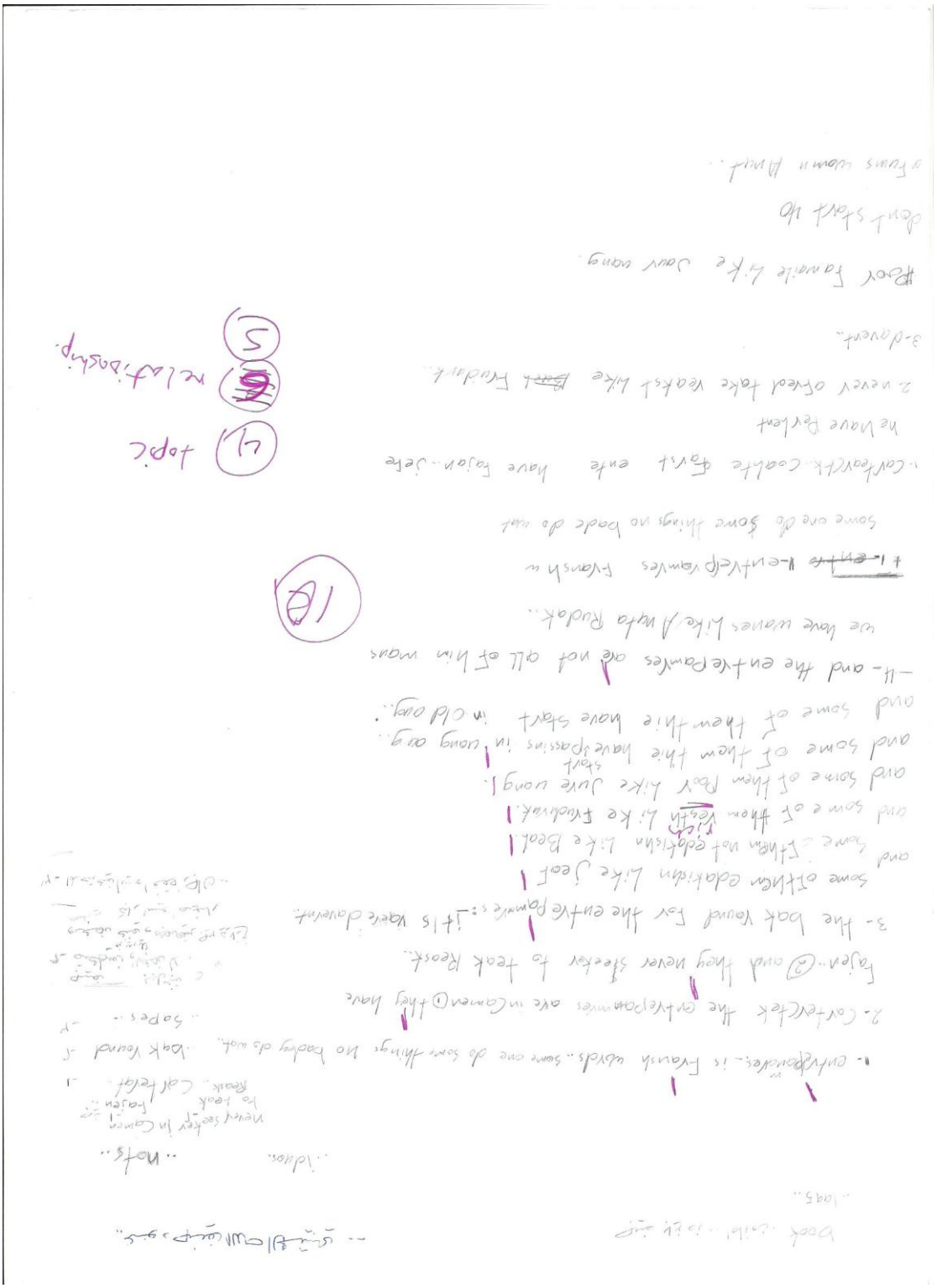
Their backgrounds: some of them are well educated like jeff bazo, but some not. bill gates did not finish college. jerry yang is immigrant and he is the formder of the yahoo website.

2) and some of them rich and some not.
3) Not all of them are men. There are women like anita roddick. the Founder of the bodday shop.
and you can see the entrepeneurs have many different backgrounds their process:

22

12 identifying topic
باقي ← 8

Amal - ramadan



Appendix 5

The first participant:

Q1: What does the word……(she does not read the word) mean?

..urrmm…It is frunch (mispronouncing it), and you can see ...or think of people in a different way.

Q2: Examples in the passage……

Hmmm…..I don't know…..What example?.....(she did not write anything)

Q3:What qual…qualities ..do entom..entro…have in common?

They have…vesion….only vesion (she wrote "vision")

Q4: What factors make entro…sometimes differ from each other? (she slowly reads the questions word by word)

Example…maybe Jeff Bezos…no…no…hmmm…what…OK.wedon't know.

Q5: A magazine……

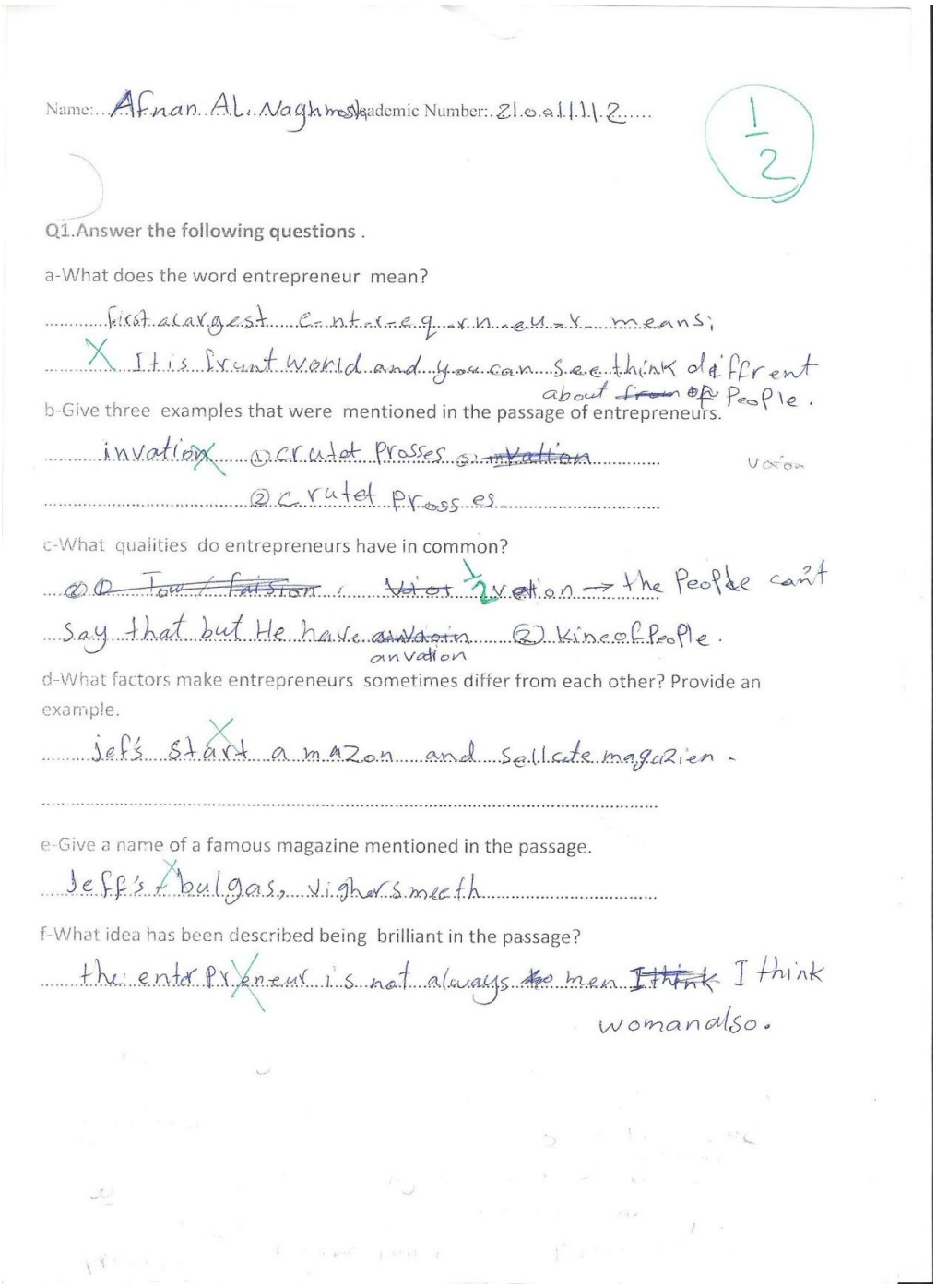
(Answering in Arabic (مافي مجله ولا جريده), which means no magazine or a newspaper was mentioned)

Q6: What idea has been des…described being bril…(difficulty with pronunciation) brillant …What is this, brillant?

Hmmm, maybe woman also work…..

International Journal of Linguistics
ISSN 1948-5425
2012, Vol. 4, No. 3

The first participant



Name: Afnan AL Naghmosh Academic Number: 21001112

1/2

Q1.Answer the following questions .

a-What does the word entrepreneur mean?

first a largest entrepreneur means;
It is frunt world and you can see think diffrent about ~~from~~ of People.

b-Give three examples that were mentioned in the passage of entrepreneurs.

invation ① crutet Prosses ② ~~invation~~ Voion
② crutet prosses

c-What qualities do entrepreneurs have in common?

~~① Tow / faision~~ ~~Voior~~ ½ vation → the People can't say that but He have ~~anvaoin~~ anvation ② kine of People.

d-What factors make entrepreneurs sometimes differ from each other? Provide an example.

jef's start a mazon and sellcate magazien.

e-Give a name of a famous magazine mentioned in the passage.

Jeff's, bulgas, vighersmeeth

f-What idea has been described being brilliant in the passage?

the entrpreneur is not always ~~to~~ men ~~I think~~ I think womanalso.

The second participant

Q1: What does the word entapenors (mispronouncing the word)?

OK….we have someone do no one does it before…and whatwewrite this better

(she writes down this answer). It means a person who start business completely in a different way.

Q2: Give examples for them….OK…means (she writes down) characteristics what kind of people. Background, they come from different background, and they create business.

Q3: What qualities ...OK…means features…do they have…OK... (she writes down) vision, the ability to see other people can't see..like Jef Bezos he uses internet for books…else...else…hmmm ...next...

Q4: What factors make entrapernars (mispronouncing the word)…having in common?

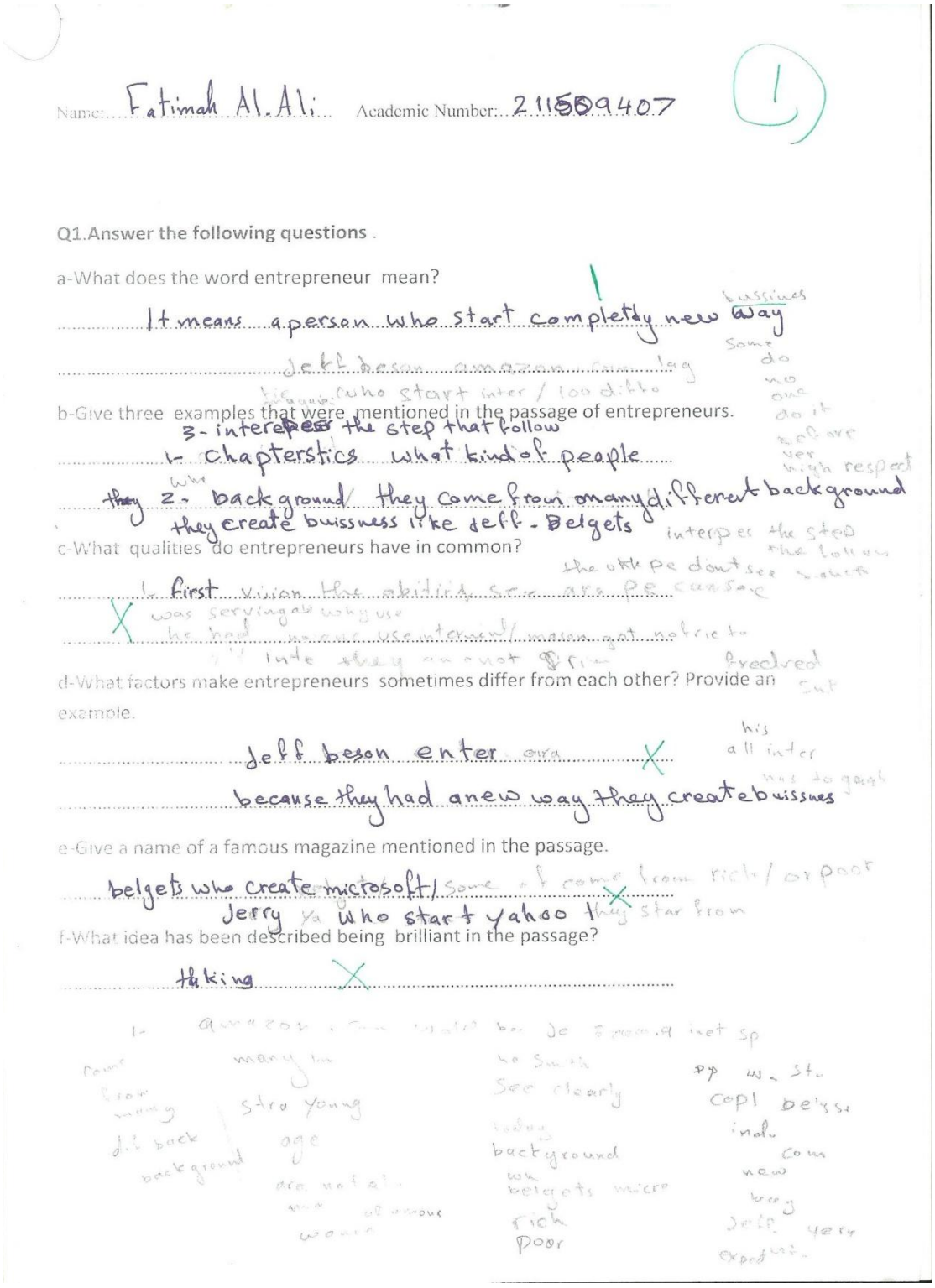
Maybe Jeff Bezos is example…(she writes)...Jeff Bezos and all have new ways of creating things….Next question.

Q5: A famous magazine (she pauses)...in the text?

Where is it (she says مافيه in Arabic, which means it is not in the text. She writes nothing)?

Q6: What idea has been described bri…brillant (mispronouncing the word) in the passage?

…..brillant means…what ...maybe smart. OK…(she writes) smart, thinking of ideas.



Name: Fatimah Al.Ali Academic Number: 211509407

Q1.Answer the following questions .

a-What does the word entrepreneur mean?

It means a person who start completly new way

Jeff beson amazon

b-Give three examples that were mentioned in the passage of entrepreneurs.

3- interepes the step that follow

1- chapterstics what kind of people

2- background they come from many different background

they create buissness like Jeff - Belgets

c-What qualities do entrepreneurs have in common?

1- first vision the abiliti

d-What factors make entrepreneurs sometimes differ from each other? Provide an example.

Jeff beson enter era

because they had a new way they create buissnes

e-Give a name of a famous magazine mentioned in the passage.

belgets who create microsoft/

Jerry ya who start yahoo

f-What idea has been described being brilliant in the passage?

taking

<u>The third participant:</u>

Q1: What does the word entrepreneurs (she pronounces it correctly) mean?

A person who starts a new business or industry (she writes)….OK.

Q2: Give three …three, hmm….OK, examples.

Jeff Bezos and the invention Amazon….who else…Fredrick, right... Number three…ermm… (she writes "Jeff" and "Bezos")

Q3: What qualities do entrepreneurs have in common….OK

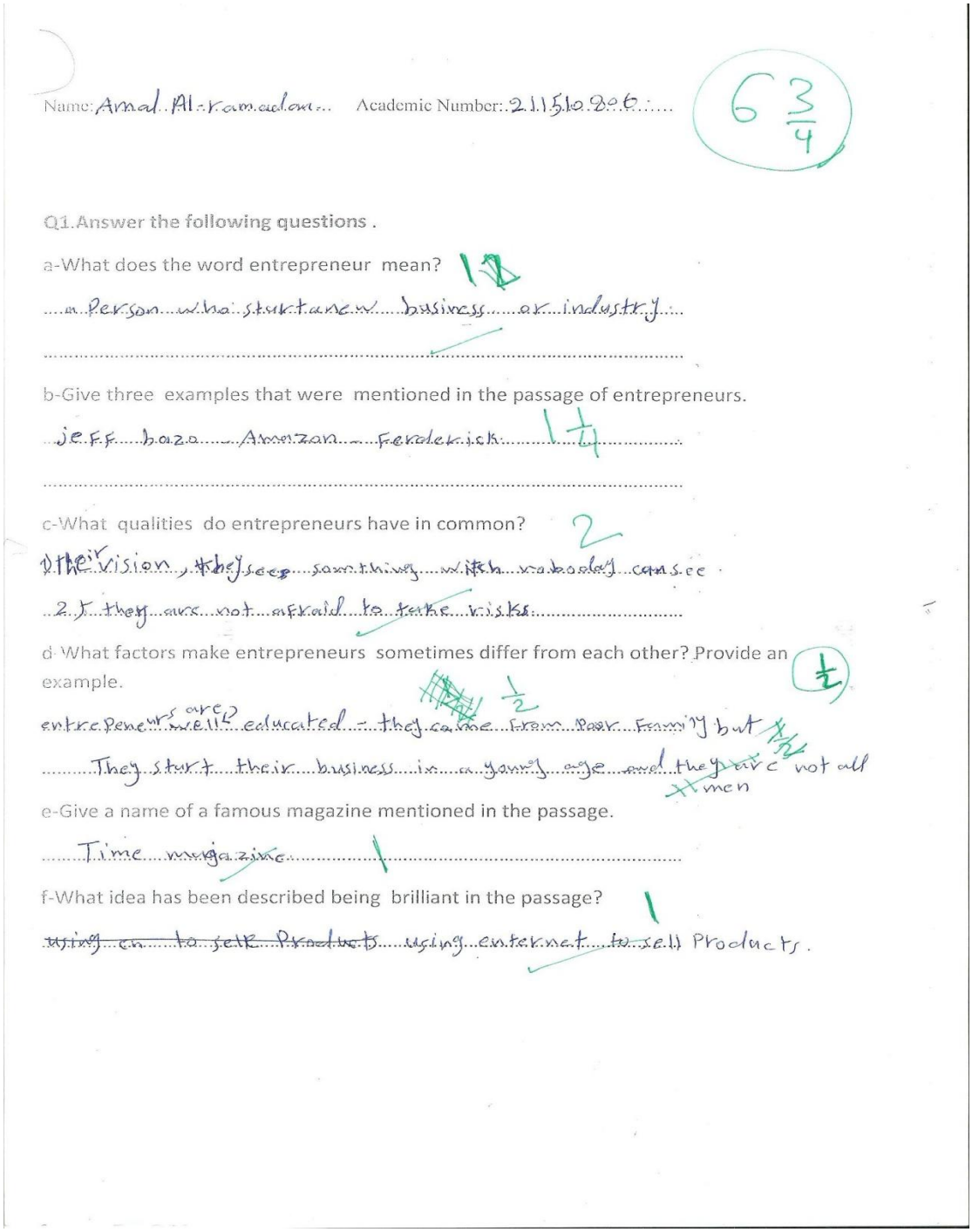
We have two. (she writes) One, their vision. They see something which no one can see, and we have number two …it is not afraid to take risks, I mean, dangers.

Q4: Hmmm…how they different?..(she writes) Some entrepreneurs are well educated, and they come from poor families …OK, they are poor… (she writes)…but start their business in a young age …What else ...OK, this is also example; (she writes) some are women, not men.

Q5: Name of a magazine...famous magazine (she writes) ...Time magazine.

Q6: A brilliant idea is…..the Amazon, of course…(she writes) Using the internet to sell products.

International Journal of Linguistics
ISSN 1948-5425
2012, Vol. 4, No. 3



Name: Amal Al-ramadan Academic Number: 211510806

6 3/4

Q1. Answer the following questions .

a-What does the word entrepreneur mean? 1

a Person who start a new business or industry

b-Give three examples that were mentioned in the passage of entrepreneurs.

jeff bazo Amazon Ferderick 1 1/4

c-What qualities do entrepreneurs have in common? 2

1) their vision, they see something with nobody can see

2) they are not afraid to take risks

d-What factors make entrepreneurs sometimes differ from each other? Provide an example. 1/2

entrepeneurs are well educated - they came from poor family but 1/2

They start their business in a young age and they are not all men

e-Give a name of a famous magazine mentioned in the passage.

Time magazine 1

f-What idea has been described being brilliant in the passage? 1

~~using en to sell Products~~ using enternet to sell Products.

<u>The fourth participant</u>

Q1: What does the word en…en…prnor mean? (she had some difficulty in reading the word)

It is Fransh word; it mean (she provides an Arabic equivalent definition: شئ.... عمل شئ من لاشئ)…. Somethings…..somethings from nothing. (she writes "something from nothing")

Q2: Examples….mmm…Jeef Bezos …else ...Amazon.com…no Amazon (crossed out later)…Fredrick Smith…female…female…..what is the name…….OK…..(she writes "Jeff Bezos" and "Fredrick")

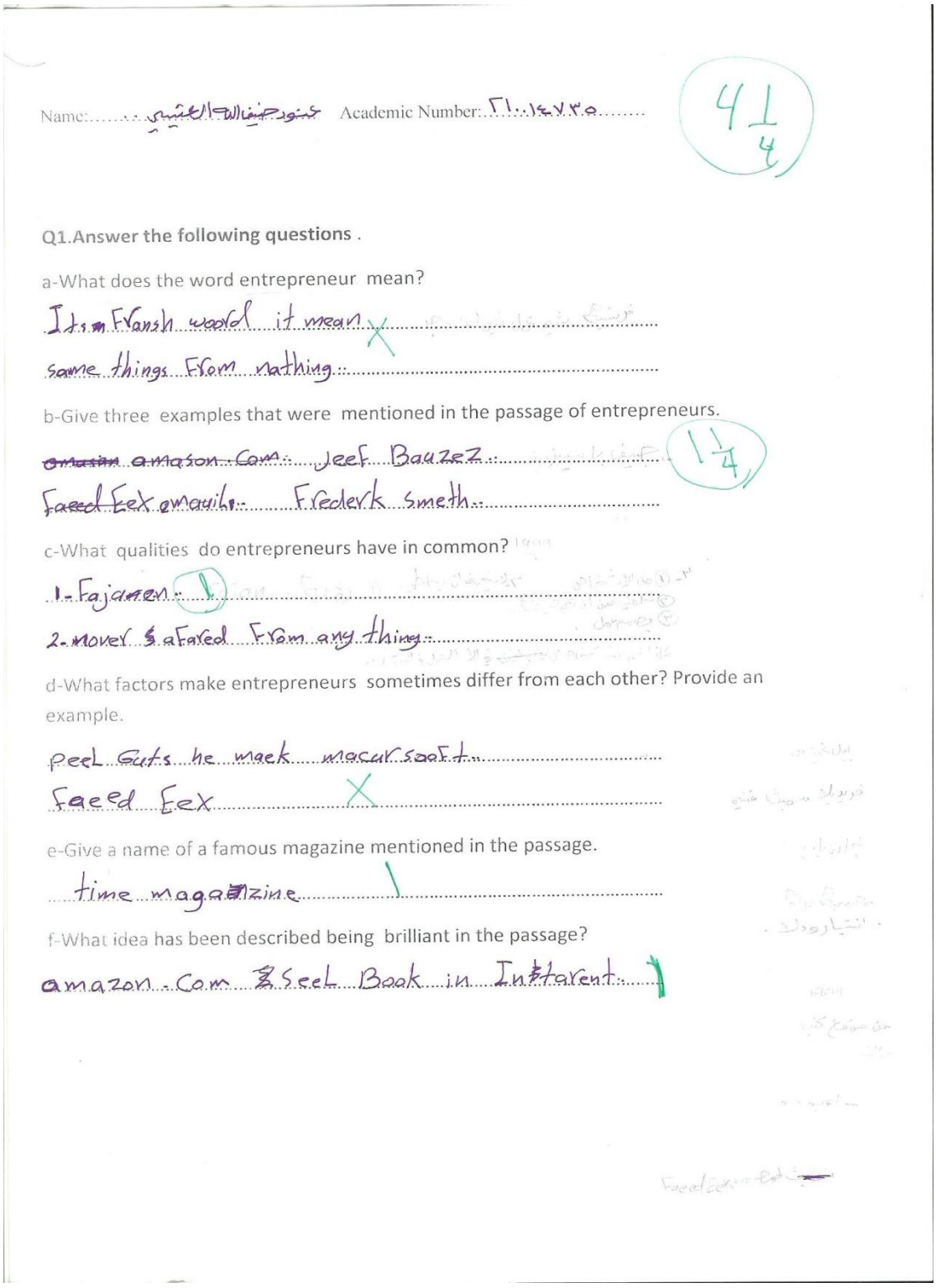
Q3: Fajan (mispronounced the word, again providing the Arabic equivalent .. رؤيه (vision)) ... لايخاف من المخاطر…(she provided the correct Arabic equivalent but failed to provide the exact answer in English; instead, she provided the following)….not afraid …Smith and Bezos…hmmm…Not afraid …no...never afraid from anything…OK…yes… (she writes "vision" and "not afraid from anything")

Q4: What… fa...factors make…entpenrs duffir maybe different …maybe (mispronouncing the words and then speaking in Arabic: "what is this? …I don't know")…example …what example…different…Bezos and Smith…hmmm…(then she answered in Arabic and wrote it on her paper: فردريك سميث غني...ووانيتا امراه....جيمي يانغ..ايش ..؟..نسيت, which means "Fredrick Smith is rich, and Anita Rodick is female….and Jemmy Yang is what?...I forgot").

Q5: Magazine, easy….Time magazine (she writes)…

Q6: …What idea has been described being brilliant…brilliant ….beginning of the passage…..very good idea is Amazon… (she writes) Amazon, sell books in internet ..this is brilliant…

International Journal of Linguistics
ISSN 1948-5425
2012, Vol. 4, No. 3



Name: عنود ضيف الله العتيبي Academic Number: ٢١٠٠١٤٧٣٥

4 1/4

Q1.Answer the following questions .

a-What does the word entrepreneur mean?

Its Fransh woord it mean X

same things From nathing.

b-Give three examples that were mentioned in the passage of entrepreneurs.

~~amasin~~ amason.Com. Jeef Bauzez.

Faeed Eex emauil. Frederk smeth.

1 1/4

c-What qualities do entrepreneurs have in common?

1- Fajamen. 1

2- mover & aFared From any thing.

d-What factors make entrepreneurs sometimes differ from each other? Provide an example.

peeL Guts he maek macursooft.

Faeed Eex X

e-Give a name of a famous magazine mentioned in the passage.

time magazine 1

f-What idea has been described being brilliant in the passage?

amazon.Com Seel Book in Intarent. 1