Investigating the Consistency between Rule Recognition Tasks and Language Production Tasks in Testing Grammatical Ability

Dr.Alshareef Edrees Ali Almllah [1]

1. Lecturer in the English Language Department, Faculty of Arts and Sciences, Salouq - University of Benghazi, Alshareef.almallah@uob.edu.ly

Abstract:

Different testing tasks are used to measure the learners' awareness of English language grammar. These tasks can be classified into mainly two types: tasks that test the recognition of the correct language structures and tasks that test the ability to produce acceptable language forms. This research applied an empirical approach to measure the consistency between rule recognition tasks and language production tasks in testing English language grammar. It also aims to shed light on the validity and reliability of these testing tasks as means for testing grammatical ability. The author used a split-half exam to test a group of 30 students. Half of the exam consisted of rule recognition items while the other half consisted of language production items. The correlation between the scores of the two halves was calculated and the results were very consistent. The results suggest that both of the two methods can give a clear evaluation of the learners' grammatical ability.

Key words: grammar testing, rule recognition tasks, language production tasks

التحقق من الاتساق بين مهام التعرف على القواعد ومهام إنتاج اللغة في اختبار القدرة النحوية

أ.الشريف إدريس علي الملاح [1]
1.محاضر بقسم اللغة الانجليزية كلية الآداب والعلوم سلوق – جامعة بنغازي

ملخص:

يتم استخدام أسئلة مختلفة لقياس وعي المتعلمين بقواعد اللغة الإنجليزية ويمكن تصنيف هذه الأسئلة إلى نوعين رئيسيين: أسئلة تختبر التعرف على التركيبات اللغوية الصحيحة وأسئلة تختبر القدرة على إنتاج تركيبات لغوية مقبولة. يطبق هذا البحث منهجًا تجريبيًا لقياس التوافق بين أسئلة التعرف على القواعد وأسئلة تكوين الجمل كأداة لاختبار قواعد اللغة الإنجليزية. كما يهدف البحث لتسليط الضوء على صلاحية ومدى الثقة في نتائج هذه الأسئلة كأداة لاختبار قدرة الطالب على استخدام قواعد اللغة. استخدم الباحث اختبارًا مكونا من جزئين لاختبار مجموعة من 30 طالبًا. احتوى نصف الاختبار الأول على أسئلة التعرف على القواعد بينما احتوى النصف الآخر على أسئلة لتكوين الجمل. تم حساب العلاقة بين نصفي الاختبار وكانت النتائج متسقة مما يشير إلى امكانية الحصول على تقييم جيد لمدى قدرة الطالب على استخدام قواعد اللغة باستخدام إحدى الطريقتين.

Scan QR & Read Article Online.

Alshareef Edrees Ali Almllah

1. Introduction:

Cambridge defines grammar as the rules about how words change their forms and combine with other words to make sentences. Traditional teaching methods focused on teaching grammar and considered mastering grammar as the core of mastering a language. It was unthinkable not to test grammar at that time (Rutherford, 1988). Even now with the new teaching methods that focus on teaching language using communicative approaches; testing grammar is still a very useful way to evaluate students' proficiency.

Grammar tests are designed to assess student proficiency in various aspects of language, including inflections and syntax. These tests utilize a range of formats, such as multiple-choice questions, error recognition exercises, rearrangement tasks, completion tests, transformation tasks, pairing and matching exercises, as well as combination and addition tests (Jabu, 2008). These testing items and many others can be classified into mainly two types: items that test the recognition of the correct language structures and items that test the ability to produce acceptable language forms. Despite the fact that rule recognition items measure the learners' understanding of the rules of the language (grammatical competence) while language production items measure their ability to use these rules (grammatical performance); different teaching institutions including universities use both of these types of items to test learners' grammatical ability.

1.1 Aims of the Research

This research aims to examine the consistency between two types of grammar testing tasks: rule recognition tasks and language production tasks as tools to measure grammatical ability. It also seeks to shed light on the validity and reliability of these testing tasks.

1.2 Statement of the Problem

The choice of the best method to test grammar depends on many factors such as the test purpose, the material to be covered, the number of learners, etc. Rule recognition questions such as multiple choice and true or false questions can be applied to cover many grammatical rules in a relatively shorter time. They are also practical for large numbers of test-takers because they are easily scored. But however such questions do not test directly the learners' ability to produce correct forms of language. On the other hand, language production questions such as describing a picture or translating a text can directly test the learners' ability to produce correct forms of language, but they are not practical for large numbers of learners or when there are a lot of rules to cover in the test. Understanding the grammatical rules a language and the ability to use these rules to produce grammatical sentences are two different things. The general assumption is that if the learners can recognize

Alshareef Edrees Ali Almllah

the correct grammatical structure then they are capable of producing correct sentences. This research tests this assumption.in other words, it tests the validity of rule recognition items as a tool to evaluate the learners' ability to produce grammatical sentences.

1.3 Methodology

This research used a quantitative approach to investigate the consistency between rule recognition items and language production items as tools for evaluating the learners' proficiency in English language grammar. The data was collected using a split-half grammar exam. The first part of the exam consisted of rule recognition questions while the second part consisted of language production questions. The test focused on testing two basic tenses: present simple and present continuous. The participants were thirty English language students at the Faculty of Arts and Science, Suloug - University of Benghazi. The correlation coefficient between the two halves of the exam was calculated to measure the consistency between them. Also the average of the difference between the participant's scores in the two halves was calculated to give a simpler indication of the degree of consistency.

2. Literature Review

The teaching of grammar has had a long and important role in the history of second language and foreign language teaching. It was thought that to learn another language meant to know the grammatical structures of that language (Rutherford, 1988). Even with the development of the communicative approach where the main purpose of learning a language changed to improve communicative competence in real-life settings; teaching grammatical rules is still considered an important ingredient to learning a language. But what to test under the title of grammar and how the test should be carried out depend on many factors some of them related to the purpose of the test others related to practicality.

2.1 Teaching Grammar:

Grammar teaching has been a source of much debate between language educators for a long time. Some educators believe that the most effective way to learn a language is by explicitly teaching grammar rules, which students then memorize and apply; others support an inductive approach, where students learn by analyzing language examples and deriving grammatical principles (Purpura, 2004). Teachers who adopted inductive language teaching believed that languages are best acquired in the same way that children acquire their native language, through extensive exposure to the language and through interaction. This belief gave rise to the 'direct method', and, ultimately, to the 'natural approach' to language acquisition (Krashen and Terrell, 1983). But even with the development of the communicative approach; teaching grammatical rules is still considered an important ingredient to learning a language. And

most of language teaching books still maintain a grammar section in every lesson where the rules are explained explicitly. So to summarize, grammar teaching have shifted from a predominant focus on structures to a greater concern for communicative effectiveness, and then to a balanced approach where meaning and communication are still emphasized, with a clear role of form and meaning-focused instruction (Doughty, 2002).

2.2 The Purpose of Testing

The main purpose of tests is to provide information about the test taker's abilities. Ur (1996:33) defines tests as: "an activity whose main purpose is to convey (usually to the tester) how well the testee knows or can do something". There are many reasons for testing language, including to meet diagnostic, proficiency, and promotional needs. Bachman and Palmer (1996) stated that: "There are two main purposes for language tests, the primary purpose is to make inferences about language ability, and the second purpose is to make decisions based on those inferences (Cited in Weigle, 2002:40). But if the tests are not capable of providing reliable and valid information about the testee's level, then the results of these tests won't be useful to make further decisions. Therefore, reliability and validity are very important aspects in language testing. According to Bachman (1990) reliability is a quality of test scores showing us the extent to which the scores from a test are free from errors of measurement, and validity is a quality of test interpretation and use providing us with information about the meaningfulness, appropriateness and usefulness of the results from the tests.

2.3 Testing Grammar

Tomlinson (2005) argues that a useful grammar test should provide an opportunity for students to reflect on the knowledge they have learnt and to recognize whether or not they are able to use the knowledge for communicative purposes. It is important to note that performance on grammar tests can be influenced by a variety of factors other than grammatical ability. In fact, test scores can vary depending on the personal attributes of test-takers such as age (Farhady, 1983; Zeidner, 1987), gender (Kunnan, 1990; Sunderland, 1995), and language background (Zeidner, 1986, 1987). Additionally, scores may fluctuate due to the test-takers' strategies (Cohen, 1994; Purpura, 1999), motivation (Gardner, 1985), and levels of anxiety (Gardner, Lalonde, Moorcroft and Evans, 1987). However, one of the most significant factors that can affect grammar-test scores is characteristics of the test itself. It is a well-known fact that the type of questions on a test can significantly affect the performance of the test-taker. For instance, some individuals may perform better on multiple-choice tasks compared to oral interviews, while others may excel in writing essays rather than filling in blanks. Addition-

Alshareef Edrees Ali Almllah

Alshareef Edrees Ali Almllah

ally, some people may score higher if they are asked to write a letter instead of interpreting a text. These variations in performance are due to the unique characteristics of each test task, which are known as test-task characteristics. Therefore, it is crucial for teachers to have a comprehensive understanding of the test tasks and their influence on the performance of test takers. And it is imperative that teachers meticulously select test items and thoughtfully evaluate their impact on the performance of the students. This ensures that the test accurately measures students' knowledge, skills, and abilities, and provides a fair evaluation of their performance.

2.4 Grammatical Knowledge and Grammatical Ability

Knowledge refers to a set of informational structures available for use in long-term memory. Ability, however, encompasses more than just a domain of information in memory; it also involves the capacity to use these informational structures in some way. Therefore, language ability refers to an individual's capacity to utilize mental representations of language knowledge built up through practice or experience in order to convey meaning. Given this definition, language ability, by its very nature, involves more than just language knowledge (Purpura, 2004).

Bachman and Palmer (1996) characterize language ability as a combination of language knowledge and strategic compe-

tence, defined as a set of metacognitive strategies (e.g., planning, evaluating). Language educators must acknowledge that a test taker's language ability can be affected by factors other than their language knowledge. These "non-linguistic factors in performance," as Carroll (1968) calls them, can reduce the reliability of the test taker's language ability interpretations. In other words, a test taker's performance is a result of their language knowledge interacting with the test task's characteristics and other non-linguistic features (such as their strategic competence, knowledge of the topic and personal attributes - as explained by Bachman (1990).

2.5 Validity

According to Hughes (2003), a test is considered valid if it accurately measures what it is intended to measure. Language tests are designed to assess theoretical constructs such as reading ability, speaking fluency, and control of grammar. The term "construct validity" refers to the extent to which we can interpret a given test score as an indicator of the abilities or constructs we aim to measure. Construct validity is also concerned with the domain of generalization to which our score interpretations apply (Bachman and Palmer, 1996).

Cronbach and Meehl (1955) identified four types of validity: predictive validity, concurrent validity, content validity, and construct validity. Predictive validity pertains to

how well a test can forecast candidates' future performance. Concurrent validity is evaluated when both test scores and criterion scores are determined at essentially the same time, and it can be studied when one test is proposed as a substitute for another. Content validity reflects the extent to which the items on a test are appropriate for the content domain it assesses, established by demonstrating that the test items are good representatives of the universe in which the researcher is interested.

Construct validity is critical whenever we want to interpret a test as a measure of an attribute or quality that is not clearly defined operationally. "Construct" is the core aspect of validity, studied when the tester seeks to demonstrate that a measure is valid by relating it to another measure that is thought to be valid.

2.6 Reliability

Bachman and Palmer (1996) defined reliability as the consistency of measurement, meaning that a test is considered reliable when it yields the same results upon repeated administration under the same conditions. Nunally (1982) described reliability as the degree to which a test consistently and accurately measures what it is intended to measure. Maduekwe (2007) emphasized that test reliability implies that a quality language test should produce consistent results. In her view, a reliable English test is one that consistently measures the intended constructs under all conditions.

Hughes (2003) identified two primary reasons for test unreliability: first, the interaction between the individual taking the test and the characteristics of the test itself; and second, the scoring process, which can also introduce sources of unreliability.

2.7 Grammar Testing Tasks

Tasks in grammar testing are activities designed to elicit linguistic (short answers) or non-linguistic (circle the answer) responses, that reflects test takers' abilities in mastering a particular language structure. The goal of grammar assessment is to measure students' grammatical ability. To achieve this, we need to create test tasks that accurately reflect the differences in students' grammatical abilities, while avoiding any irrelevant variability that may be caused by the types or quality of the tasks used. Language teachers know that the types of tasks used in a test and their quality can greatly affect students' performance. Bachman and Palmer (1996) proposed a model that analyzes tasks based on five key elements, each defined by specific characteristics. These five elements describe the following **characteristics:**

1. The characteristics of the setting include the physical characteristics, the participants, and the time of the task.
2. The test rubrics include the instructions, the overall structure of the test, the time allotment and the method used to score the response.

Alshareef Edrees Ali Almllah

Alshareef Edrees Ali Almllah

These characteristics can obviously influence test scores in unexpected ways (Madden, 1982; Cohen, 1984, 1993).

3. The characteristics of the input (sometimes called the stimulus) are critical features of performance in all tasks. The input is the part of the task that test-takers must process in order to answer the question. It is characterized in terms of the format and language.

4. The characteristics of the expected response are also considered in terms of the format and language. Similar to the input, the expected response of grammar tasks can vary according to channel (aural or visual), form (verbal, non-verbal), language (native or target) and vehicle (live or reproduced).

5. The relationship between the input and response can be direct or indirect. If the response is based primarily on information in the input, the relationship between the input and response is direct. If, however, the response cannot be based on the input, but rather needs other kinds of topical or pragmatic information, the relationship between the input and response is characterized as indirect.

2.8 Types of grammar testing tasks

Traditionally, various methods have been used to categorize the types of tasks found on tests. One common approach classifies tasks based on their scoring procedures. For instance, objective test tasks, such as true-false questions, are those that do not require expert judgment to evaluate performance based on correctness criteria. In contrast, subjective test tasks, like essays, necessitate expert judgment to interpret and assess performance according to the standards of correctness (Purpura, 2004). Others have classified tasks according to what they test as tasks that measure grammatical knowledge (rule recognition tasks) and tasks that measure grammatical ability (language production tasks). This research adopted the second classification as it's related directly to the purpose of testing. These two types of tasks are discussed with regard to Bachman and Palmer's framework of task characteristics described above.

2.8.1 Rule Recognition tasks

These tasks include Multiple-choice (MC) tasks, Error recognition tasks, matching tasks, discrimination tasks, noticing tasks etc... They present input in the form of an item, and test-takers are required to select the response. These tasks aim to assess the recognition or recall of grammatical forms and meanings. They are well-suited for evaluating various discrete aspects of grammatical knowledge. Scoring is typically done as right or wrong, based on a single criterion for correctness. Scoring of such tasks is relatively easy and very reliable. Hence, they are very practical for testing a large number of test-takers. However, developing items for rule recognition tasks can be challenging and time-consuming. Additionally, the format may encourage guessing, which could inflate scores due to the test-taker's fa-

miliarity with test-taking strategies. This raises important concerns about the validity of the conclusions drawn from these scores. (Cohen, 1998). Some educators argue that these tasks are inauthentic language-use tasks. Regardless of these limitations rule recognition tasks are still widely used to assess grammatical knowledge and to make inferences about grammatical abilities of test-takers.

2.8.2 Language production tasks

These tasks can be further classified into two types:

1) Limited-production tasks such as gap-filling tasks and short-answer tasks.

2) Extended-production tasks such as describing a picture, reporting tasks, story-telling, simulation tasks and role-play.

Production tasks provide input in the form of items that include language and/or non-language information, which can vary in length or topic. Unlike rule recognition tasks, production tasks require a response that reflects a limited amount of language production. Limited-production tasks aim to assess the examinee's ability in a specific grammatical structure, while extended-production tasks are more suited for evaluating the examinee's ability to use various grammatical forms to convey meanings in speaking and writing. Scoring language production tasks is generally more complex than scoring rule recognition tasks. Production tasks can be scored in several ways. For items with a single criterion for correctness, responses can simply be marked as right or wrong. When multiple criteria for correctness exist, separate scores (for example, one for grammatical form and another for meaning) can be compiled to create composite scores for each criterion. This approach allows an item to receive full, partial, or no credit based on the quality of the response. The main advantage of language production tasks is that they provide direct evaluation of the test-takers' abilities to use grammatical structures. But they have limitations regarding practicality of use especially with large numbers of test-takers. Also they cannot be scored in an objective method which questions their scoring reliability.

3. Data collection

In order to find out if rule recognition tasks can provide good evaluation of the participant's grammatical ability compared to language production tasks a split-half exam was prepared. The spilt-half testing method was used because it tests the participants under the same conditions except only the variable in question which is the task type. The first half of the exam used rule recognition tasks and consisted of two questions: the first question was 12 multiple choice items. And the second question contained 8 pairs of sentences from which the participants have to choose the correct one. The second half of the exam used language production tasks and it also contained two questions: in the first ques-

Alshareef Edrees Ali Almllah

Alshareef Edrees Ali Almllah

tion the participants had to write sentences to describe four pictures using the present continuous tense. And in the second question they had to write six sentences about John's daily routine using present simple tense. The details of John's daily routine were presented in a time table. The test was designed to test the participant's grammatical ability in two tenses only: present simple and present continuous. A copy of the exam is found in the appendix.

The research participants consisted of 30 English language students from the University of Benghazi, Faculty of Arts and Science Suloug. They were randomly selected from various semesters to ensure a range of proficiency levels.

Each half of the exam was scored out of ten. For multiple choice items and choose the correct sentence the participants get 0.5 mark for every correct response. And for the second half which is writing sentences the participants get 1 mark for every correct sentence. The chart below shows the participants scores arranged according to their score in the rule recognition tasks.

4. Data Analysis

To find out the consistency between the scores of the two types of tasks the correlation coefficient was calculated. The correlation coefficient is a value that ranges from -1 to 1, indicating the strength and direction of the relationship between two sets of data. It reflects how closely the measurements from two or

The sign (+/-) indicates the type of correlation between the variables as shown in the table.

Correlation value	Correlation type	Meaning
1	Positive correlation	When a variable changes, the other variable changes in the same direction.
0	No correlation	The variables have no relationship.
-1	Negative correlation	When a variable changes, the other variable changes in the other direction.

The absolute value of the correlation coefficient indicates the strength of the relationship between the variables. The greater the absolute value, the stronger the correlation. Labeling systems exist to roughly categorize r values as:

Correlation Coefficient	Correlation Strength
r < 0.35	Weak Correlation
0.36 to 0.67	Moderate Correlation
0.68 to 1.0	Strong Correlations
r > 0.90	Very Strong Correlations

The following table shows how the correlation coefficient was calculated using the correlation coefficient equation

The Correlation Coefficient Equation:

r = Correlation Coefficient

n = number of observations or samples

P.	Rule recognition tasks (x)	Language production tasks (y)	x.y	x^2	y^2	Δxy
1	9	10	90	81	100	1
2	9	9	81	81	81	0.5
3	9	9	81	81	81	0.5
4	8.5	9	76.5	72.25	81	1
5	8.5	9	76.5	72.25	81	1
6	8	9	72	64	81	1
7	8	7	56	64	49	1
8	7.5	9	67.5	56.25	81	2
9	7.5	8	60	56.25	64	0.5
10	7.5	7	52.5	56.25	49	0.5
11	7	9	63	49	81	2
12	7	8	56	49	64	1
13	7	7	49	49	49	0
14	6.5	9	58.5	42.25	81	2.5
15	6	9	54	36	81	3
16	6	5	30	36	25	1
17	6	4	24	36	16	2
18	6	4	24	36	16	2
19	6	4	24	36	16	2
20	5.5	9	49.5	30.25	81	3.5
21	5.5	8	44	30.25	64	2.5
22	5.5	7	38.5	30.25	49	1.5
23	5.5	4	22	30.25	16	1.5
24	5.5	3	16.5	30.25	9	2.5
25	5.5	2	11	30.25	4	3.5
26	5	6	30	25	36	1
27	5	4	20	25	16	1
28	5	2	10	25	4	3
29	4	4	16	16	16	0
30	3	2	6	9	4	1
	$\sum x = 195$	$\sum y = 196$	$\sum xy = 1359$	$\sum x^2 = 1335$	$\sum x^2 = 1476$	Avg. Δxy=1.51
		The Correlation Coefficient = 0.739				

Alshareef Edrees Ali Almllah

Alshareef Edrees Ali Almllah

The correlation coefficient between the two methods used to test the participants' grammatical proficiency is 0.739, also the average difference between the marks in only 1.51. This indicates a relatively strong correlation. The results suggest that rule recognition tasks can be used as an indicator of the participant's grammatical ability, similar to language production tasks. As seen in the chart, participants who scored high in rule recognition tasks tended to score high in language production tasks as well. However, there were a few cases where participants scored inconsistently, which could be attributed to their individual attitude and how they approach different types of questions.

5. Conclusion

The appropriate method for assessing grammatical proficiency is determined by various factors, such as the test's purpose, the material being evaluated, and the number of learners. There are mainly two types of grammar testing tasks; both of them have advantages and disadvantages. Rule recognition tasks, such as true or false and multiple-choice questions, have many benefits; they can cover many grammatical rules in a short time and are practical for large numbers of test-takers, as they are scored easily and objectively. However, they have been criticized for their indirect measure of grammatical ability. On the other hand, language production tasks, which test directly the ability of an individual to produce correct grammatical sentences, have also some disadvantages; such as difficulty and subjectivity of scoring.

The findings of this research revealed a strong correlation between the two types of grammar testing tasks; this suggests that both of the tasks can potentially serve as a valid indicator of an individual's grammatical ability. The findings also support the assumption that if an individual has adequate knowledge of grammatical rules of a language, them he will have the ability to use this knowledge to construct acceptable sentences. However, it is still recommended to use tasks that test directly what need to be tested if possible, or at least use different types of tasks to get a valid and reliable evaluation.

Reference

1.Bachman, F.L. (1990). Fundamental Considerations in Language Testing. Oxford: Oxford University Press.

2.Bachman, F.L & Palmer, A.S. (1996) Language Testing in Practice. Oxford: Oxford University Press.

3.Carroll, J. B. (1968). The psychology of language testing. In A. Davies (ed.), Language Testing Symposium: A Psycholinguistic Approach (pp. 46–69). London: Oxford University Press

4.Cohen, A. (1984). On taking language tests: What the students report. Language Testing, 1, 70–81.

5.Cohen, A. (1993). The role of instructions in

testing summarizing ability. In D Douglas and C. Chapelle (eds.), A New Decade of Language Testing Research (pp. 132–60). Alexandria, VA: TESOL.

6.Cohen, A. (1994). Assessing Language Ability in the Classroom. Boston: Heinle & Heinle Publishers.

7.Cohen, A. (1998). Strategies in Learning and Using a Second Language. New York: Longman.

8.Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological bulletin, 52(4), 281.

9.Doughty, C. (2002). Focus on form: the latest word for teachers. Third Annual Apple Lecture, Teachers College, Columbia University, New York City

10.Douglas, D. (2000). Assessing Language for Specific Purposes. Cambridge: Cambridge University Press.

11.Farhady, H. (1983). New directions for ESL proficiency testing. In J. Oller (ed.), Issues in Language Testing (pp. 253–69). Rowley, MA: Newbury House.

12.Gardner, R. C. (1985). Social Psychology and Language Learning: The Role of Attitudes and Motivation. London, Ontario: Edward Arnold.

13.Gardner, R. C., Lalonde, R. N., Moorcroft, R. and Evans, F. T. (1987). Second languageattrition: the role of motivation and use. Journal of Language and Social Psychology, 6, 29–47.

14.Hughes, A. (2003). Testing for Language Teachers 2nd Ed. Cambridge: Cambridge University Press.

15.Jabu, B.(2008). English Language Testing. Makassar: the UNM Publisher.

16.Krashen, S. D. and Terrell, T. D. (1983). The Natural Approach: Language Acquisition in the Classroom. San Francisco: Alemany Press.

17.Kunnan, A. (1990). DIF in native language and gender groups in an ESL placement test. TESOL Quarterly, 20 (4), 740–6.

18.Madden, H. (1982). Determining the debilitative impact of test anxiety. Language Learning, 32, 133–43

19.Maduekwe, N. (2007). The Deregulation Policy in the Nigerian Downstream Petroleum Sector and Its Effects. Available at SSRN 2812770.

20.Nunan, D. (1989). Designing Tasks for the Communicative Classroom. Cambridge: Cambridge University Press.

21.Nunnally, W. C. (1982). Magnetic switches and circuits (No. LA-8862-MS-Rev.). Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

22.Purpura, J. (2004). Assessing grammar. Cambridge: Cambridge University Press.

23.Purpura, J. and Pinkley, D. (1999). On Target 1 (2nd edition). White Plains, NY: Addison Wesley Longman.

24.Rutherford, W. (1988). Grammatical consciousness raising in brief historical perspective. In W. Rutherford and M. Sharwood Smith (eds.), Grammar and Second Language

Alshareef Edrees Ali Almllah

Alshareef Edrees Ali Almllah

Teaching (pp. 15–18). Boston, MA: Heinle and Heinle Publishers.

25.Sunderland, J. (1995). Gender and language testing. Language Testing Update, 17, 24–35.

Tomlinson, B. (2005). Testing to learn: a personal view of language testing. ELT Journal, 59(1), pp. 39-46.

26.Ur, P. (1996). A Course in Language Teaching. Cambridge: Cambridge University Press.

Weigle, S.C. (2002). Assessing Writing. Cambridge: Cambridge University Press.

27.Zeidner, M. (1986). Are English language aptitude tests biased toward culturally different minorities? Some Israeli findings. Language Testing, 3, 80–9.

28.Zeidner, M. (1987). A comparison of ethnic, sex, and age bias in the predictive validity of English aptitude tests: some Israeli data. Language Testing, 4, 55–71.

Appendix

Rule recognition Items:

Q1: Choose the correct answer.

1.I .... orange juice.

a) am liking   b) likes   c) like
d) is like

2.Andy and Lucy .... with us these days.

a) are staying   b) stay   c) is staying
d) stays

3.We .... to school every day.

a) aren't going   b) don't go   c) not going   d) doesn't go

4.I .... my friends in the café after school on Fridays.
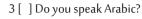
a) meet   b) am meeting   c) meating
d) meets

5.Look! John .... his school uniform!

a) wears   b) is wearing
c) wear   d) do wear

6.Do you want to play football? ~ Not now. I .... dinner.

a) am having   b) have   c) has
d) haves

7.Do you have any ice cream? ....

a) Yes, I am having   b) Yes, I am
c) Yes, I do   d) Yes, I has

8................. Jamie Chandler?

a) is John knowing   b)   does   John
know   c) does John knows d) do John know

9.I .... for a present but I can't find anything good.

a) am look   b) look
c) I am looking   d) looks

10.My brother and I .... early at the weekend.

a) don't get up   b) aren't get up
c) don't getting up   d) am not getting up

11.Are your parents working at the moment? No, they ....

a) don't   b) aren't
c) aren't being   d) doesn't

12.My baby sister .... all the time.

a) crying   b) is crying   c) crys
d)cries

Q2 Tick (✔) the correct sentence.

1[ ] Does he has a car?

5[ ] Are you playing football?

[ ] Does he have a car?

[ ] Do you playing football?

2[ ] Where do you come from?

6[ ] My friends are coming.

[ ] Where does you come from?

3 [ ] Do you speak Arabic?

7[ ] he is knowing him.

[ ] Are you speak Arabic?

[ ] he knows him?

4 [ ] I don't speak Chinese.

(ride – a bicycle)   Ex. He is riding a bicycle . . . .

(play – basketball) 1. ....

(take – a photo)   2. ....

(paint – a picture) 3...

(carry – a parcel)   4. ....

Q2. Write 6 sentences about John's daily routine. Use the information in the table below:

Activity	Time
Get up	7:00 AM
Have a shower	7:15 AM
Have breakfast	7:30 AM
Go to school	8:00 AM
Get home	1:00 PM
Have lunch	1:30 PM
Have dinner	8:30 PM
Do homework	9:00 PM
Go to bed	11:00 PM

1. ....

2. ....

3. ....

4. ....

5. ....

6. ....