The Validity of Grammaticality Judgment Task on Saudi EFL Learners

Sami Alanazi
School of Language, Literature and International Studies
University of Central Lancashire
14 The Room Apartments, Lawson Street, PR12QF, Preston, UK
E-mail: sami.alenezi@gmail.com

Received: 01-04- 2015           Accepted: 23-06- 2015                        Advance Access Published: August 2015
Published: 01-11- 2015          doi:10.7575/aiac.ijalel.v.4n.6p.78       URL: http://dx.doi.org/10.7575/aiac.ijalel.v.4n.6p.78

Abstract
The purpose of this paper is to shed light on the validity of the controversial data-gathering tool Grammaticality Judgment Task on measuring the grammatical competency of Saudi EFL learners. Despite the widespread use of GJT in SLA research, it is surrounded by a great deal of criticism. The present paper is part of a larger study investigating the acquisition of past verb forms by Saudi EFL learners. Thirty-six Saudi EFL learners took part in the study and were divided into three groups as follows: guided-planning group, semi-guided planning group, and control group. The task used in the study consisted of twenty test items: 10 control test items, and 10 experimental test items. The results did not reveal significant statistical differences between the three groups. Also, the results did not reflect the actual grammatical competency of the participants.

Keywords: Grammaticality Judgment Task, GJT; Second Language Acquisition, SLA; Linguistics; Saudi Learners

1. Introduction
Grammaticality Judgment Task (GJT) is a widely used tool by researchers in the area of Second Language Acquisition SLA. In this task, L2 learners are presented with a set of sentences and they are required to identify the grammatically deviant sentences. GJ Tasks are conducted to: (a) assess the speakers’ reactions to sentence types that only occur rarely in spontaneous speech; (b) obtain negative evidence on strings of words that are not part of the language; (c) distinguish production problems (e.g., slips, unfinished utterances, etc.) from grammatical production; and (d) isolate the structural properties of the language that are of interest by minimizing the influence of the communicative and representational functions of the language (Schütze, 1996). However, some concerns have arisen regarding the validity of GJT. GJT can be influenced by extra grammatical factors, the link between metalinguistic judgments and grammatical knowledge is not clear, and there is a lack of control techniques (Tremblay, 2005). Another concern about the validity of GJT is that learners may base their judgment on extraneous factors, such as sentence complexity or semantic irregularity (Ellis, 1991).

The present study investigates the validity of Grammaticality Judgment Task on Saudi EFL Learners. The paper is part of a larger study investigating the acquisition of past verb forms by Saudi EFL learners. A GJT was utilised in the larger study to measure the participants’ competency in identifying grammatically deviant sentences. The larger study was carried out to explore the factors that hinders the acquisition of past verb forms by Saudi EFL learners. Also, the larger project was set out to highlight the extent of L1 interference when Saudi EFL learners report in English using the past tense. The larger study tested the receptive and productive knowledge of the participants by using GJT and Picture-Cued Storytelling task. It concluded that Saudi EFL learners have difficulty in acquiring the past verb form in English. However, the GJT did not provide sufficient understanding of the participants L2 competency.

2. Background
The use of Grammaticality Judgment Task (GJT) goes back to the late 70’s and early 80’s (Bialystok, 1979; Gass, 1983). Since then, GJT have been widely used by researchers as a data-collection tool to test theoretical claims (Tremblay, 2005). Grammaticality judgments (GJT) comprise one (although definitely not the sole) kind of metalinguistic function, or objectification of language. In other words, the one means of objectifying language is to declare if a provided sentence is suitable or not. Grammaticality judgments comprise intricate behavioural performance: the participants in GJT may guess when not sure, lose patience if the test is too long, or try to keep balance between the number of sentences they judge to be grammatical and deviant (Ellis, 1991).

A great amount of SLA studies have employed grammaticality judgment tasks (for an early review of these see R. Ellis, 1991a) as a means of quantifying L2 learners’ knowledge. It is apparent that the preferred technique of exploring L2 explicit awareness as a conscious knowledge comprises the grammaticality judgment feat. Currently there is considerable literature on GJT. Following the reviews of Chaudron (1983) and R. Ellis (1991a) of GJT studies, several supplementary researches have been done (e.g., Bard, Robertson, & Sorace, 1996; Cowan & Hatasa, 1994; Davies &
Thirty-six Saudi students from different universities in the UK participated in this study. They were all adult males and learners in assessing individual sentences is considered.

### 3. Methodology

#### 3.1 Participants

Thirty-six Saudi students from different universities in the UK participated in this study. They were all adult males and...
females who had completed a one-year general English language course in the UK as a requirement to achieve the sufficient English language proficiency level. Their average score in IELTS were between 5.0 and 6.0. The participants were studying different disciplines, and had lived in the UK for more than two years. The homogeneity in IELTS results and the number of years spent in the UK for the participants was to maintain a balanced level of L2 proficiency and to ensure they had the adequate competency to identify grammatically deviant and well-formed sentences in English. The nature of the study was explained to the participants and they signed a form giving their consent to take part in this study. The participants were divided into three groups:

**Guided-Planning Group** consisted of twelve participants with different levels of English language proficiency (judged by their IELTS scores). The guided-planning group was provided with a detailed explanation about the task. Participants in this group were asked explicitly to focus on the past verb forms and they were provided with example. The guided-planning group was allowed time to plan their answers prior to the task and was provided with guidance. The implementation of pre-task planning was to get positive reflection on the participants’ performance (Willis & Willis, 1988).

**Semi-Guided Planning Group** consisted of twelve participants with different levels of English language proficiency (judged by their IELTS scores). The semi-guided planning group was also provided with detailed instructions and the participants were asked to pay attention to the past verb forms but no examples were provided. The participants in this group were also given time to plan their answers before engaging with the task.

**No-Planning Group** This was a control group consisting of twelve participants with different levels of English language proficiency (judged by their IELTS scores). The group received general instructions about the task with no further details or planning time.

### 3.2 Test Items

The task included 20 test items: 10 control items, and 10 experimental items. The items were presented in a counterbalanced order. All the test items were formulated in the past tense including three aspects: past simple, past progressive, and past perfect because the task was part of a larger study investigating the acquisition of past verb forms by Saudi EFL learners. Generally, the control test items were simple, because they, as suggested by Bullock et al. (2005), serve as a comparison point for the experimental items. In contrast, the verbs in the experimental items were manipulated because they were the focus of the study.

### 3.3 Procedure

The task was conducted in many cities in the UK and at different times because the participants were assembled from different cities in the UK. The participants sat for the tasks individually, and occasionally in groups for the GJT when the researcher travelled to meet a group of participants. The participants were given an introduction regarding the aims and purpose of the research and the tasks they were about to undertake. The first task given was the Grammaticality Judgment Task. The participants were given a list of sentences and they were asked to determine whether they were grammatically correct or incorrect. The same task was repeated after eight weeks to find out if the instruction on using the past verb forms had taken place (Schmitt, 2010).

As mentioned earlier, the GJT used in this paper was part of a larger study carried out to investigate the acquisition of past verb forms by Saudi EFL learners. The participant sat for the tasks individually, and occasionally in groups. Each participant was given a brief introduction about the purpose of the study. The participants were given a maximum of 10 minutes to plan their answers and provide their judgment. The same task was repeated in eight weeks’ time.

The Participants were instructed to base their judgment on their intuition and whether they would, or would not, use the sentence in normal situations. To avoid rejecting sentences “on the basis of prescriptive rules of English”, the participants were instructed not to reject a sentence because they might know a better way to deliver the same meaning (Tremblay, 2005 p.144). For this purpose, guidance and examples were provided before engaging with the task.

### 4. Results

A one-way analysis of variance (ANOVA) between groups was conducted to compare the results from Grammaticality Judgment Task 1 (GJT1). Looking at table (1) below, the analysis shows no significant results between the three groups: guided-planning group, semi-guided planning group, and no-planning group [F (2, 33) = 0.512, p = 0.604]. However, the participants from the guided-planning group showed better performance in their initiative judgment on task items (M = 4.33, SD = 1.155) than participants from semi-guided group (M = 3.92, SD = 1.832) and no-planning group (M = 3.75, SD = 1.288). In other words, the null hypothesis of meeting the assumption is true as no significant results emerged between the three groups from the Grammaticality Judgment Task 1 (GJT1).

<table>
<thead>
<tr>
<th>Group</th>
<th>Mea</th>
<th>Std. Deviation</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guided-planning group</td>
<td>4.33</td>
<td>1.155</td>
<td>0.604</td>
</tr>
<tr>
<td>Semi-guided planning group</td>
<td>3.92</td>
<td>1.832</td>
<td></td>
</tr>
<tr>
<td>No-planning group</td>
<td>3.75</td>
<td>1.288</td>
<td></td>
</tr>
</tbody>
</table>

The same analysis was repeated on the delayed test Grammaticality Judgment Task 2 (GJT2), which was given to the
same participants of the three groups eight weeks later. This time, the results yielded a slight significant change between the groups \(F(2, 33) = 3.317, p = 0.049\). Again, the participants from the guided-planning group showed better performance in their initiative judgment on task items (M = 4.83, SD = 1.115) than participants from semi-guided group (M = 4.33, SD = 1.557) and no-planning group (M = 3.58, SD = 0.793).

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guided-planning group</td>
<td>4.83</td>
<td>1.115</td>
<td>0.049</td>
</tr>
<tr>
<td>Semi-guided planning group</td>
<td>4.33</td>
<td>1.557</td>
<td></td>
</tr>
<tr>
<td>No-planning group</td>
<td>3.58</td>
<td>0.793</td>
<td></td>
</tr>
</tbody>
</table>

A Post Hoc Test and a Complex Contrast Test were employed on Grammaticality Judgment 2 (GJT2) to find out which group caused the significance. The Tukey HSD Post Hoc Test was selected for this purpose to determine which group made the difference in the results. The guided-planning group was compared to the other two groups (semi-guided planning and no planning) and the results from the complex contrast test showed that guided-planning group was significantly different in terms of their performance in the delayed grammaticality judgment task at \(p = 0.047\). However, the Tukey HSD test showed nothing significant between the guided-planning group and the semi-guided planning group \(p = 0.568\), and between the semi-guided planning group and the no-planning group \(p = 0.288\). Nonetheless, the statistical significance occurred between guided-planning group and no-planning group \(p = 0.039\). See table (3) below.

<table>
<thead>
<tr>
<th>(I) Group</th>
<th>(J) Group</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP Group</td>
<td>SGP Group</td>
<td>.500</td>
<td>.489</td>
<td>.568</td>
<td>-.70</td>
</tr>
<tr>
<td>NP Group</td>
<td>GP Group</td>
<td>1.250*</td>
<td>.489</td>
<td>.039</td>
<td>.05</td>
</tr>
<tr>
<td>SGP Group</td>
<td>NP Group</td>
<td>-.500</td>
<td>.489</td>
<td>.568</td>
<td>-1.70</td>
</tr>
<tr>
<td>NP Group</td>
<td>GP Group</td>
<td>-1.250*</td>
<td>.489</td>
<td>.288</td>
<td>-.45</td>
</tr>
<tr>
<td>NP Group</td>
<td>SGP Group</td>
<td>-.750</td>
<td>.489</td>
<td>.039</td>
<td>-2.45</td>
</tr>
</tbody>
</table>

When looking at table (4) below, we can see the gain scores for the guided planning group and the semi-guided planning group are convergent, guided-planning group score 6 points more in the delayed test where the semi-guided planning group scored 5 points. However, the gain score of the no-planning group is very disappointing at it came to (-2). In other words, the no-planning group performance in the grammaticality judgment delayed test, as expected, had fallen back.

<table>
<thead>
<tr>
<th>Group</th>
<th>GJT1</th>
<th>GJT2 (Delayed)</th>
<th>Gain Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guided Planning</td>
<td>52</td>
<td>58</td>
<td>6.00</td>
</tr>
<tr>
<td>Semi-Guided Planning</td>
<td>47</td>
<td>52</td>
<td>5.00</td>
</tr>
<tr>
<td>No-Planning</td>
<td>45</td>
<td>43</td>
<td>-2.00</td>
</tr>
</tbody>
</table>

In summation, there were no statistically significant differences among the three groups in the first GJT \(p = 0.604\). The delayed GJT, however, showed a slight significant change in the performance of the three groups \(p = 0.047\), which was expected as the delay time was only eight weeks and the same procedure for guidance and planning was conducted.

5. Discussion

The objective of this paper is to test the validity of GJ on Saudi EFL learners with a focus on the past verb forms. The test included 20 items: 10 control items, and 10 experimental items. The test items were presented in a counterbalanced order. The instructions were very specific and clear, and the participants were allowed 10 minutes to produce their judgment. The results revealed that the participants failed to intuitively determine whether a sentence was correct or deviant from a grammatical perspective. It has also shown that participants lack the ability to assign the correct temporal reference when reporting an incident in the past tense.

The results showed that the three groups did not make any statistical significance in first encounter of the GJT although two groups were allowed to plan their answers before engaging with the test and they were provided with guidance. This indicates that the participants were unable to identify grammatically deviant from well-formed sentences focused
on the past verb form. The second encounter, however, revealed a slight improvement in the performance of the first and second groups and a fall back in the performance of the third group. The improvement pinpoints that repeating the same test with the same planning and guidance procedure results in better outcomes.

The study focused on the three aspects in the past tense: simple, progressive and perfect to measure the participants understanding those aspects. The participants demonstrated good understanding of the past simple and were able, generally, to make clear judgments about the ungrammatical sentences. However, they showed some difficulty in recognizing deviant sentences in the past progressive form. The past perfect form was even harder for the participants to identify the ungrammatical sentences in this form.

Despite the slight statistical significance in the delayed test, the results indicated that guided-planning and semi-guided planning did not make any significant change in the informants’ performance, and the results were, generally speaking, convergent. Although the sample was relatively small, the results indicate that Saudi EFL learners face problems when using the past tense in English.

The target from using the grammaticality judgment task was to measure the intuitive judgment of the participants to the extent that they believed the sentences to be grammatically acceptable. However, the results show that the participants from the three groups were unable to make that clear judgment. That said, the two groups were given guidance and a chance to plan their answers. The same test was repeated eight weeks later on the same groups, and the results this time came with a slight statistical significant change (see table 2). Although the null hypothesis was rejected on the delayed test p. = 0.049, the performance of the guided-panning group and semi-guided planning was approximate, and the significance was made by the performance of the no-planning group (see table 4).

6. Conclusion

GJT is one of the widely used tools for Metalinguistic data in SLA research. Despite the criticism, GJT is still an essential tool for gathering metalinguistic data in the area of SLA. The finding of this study comes in alignment with the concerns about the validity of GJT in measuring grammatical awareness. This study makes a potential contribution to the literature on investigating the validity of GJT on Saudi EFL learners. The outcomes of this study suggest that design of the GJT plays a major role in determining the reliability of the collected data. The complexity of the experimental test items used in this study affected the performance of the participants and maybe led them to reject those items.

Based on the results of this study, it is suggested that researchers take extra caution when using GJT as a tool for the data collection. GJT can be more effective when designed carefully and the test items are less complex. A true-false-like test is not efficient for GJT as it might push L2 learners to make judgments about the grammaticality of a sentence based on random guessing. The test items in the GJT used in this study focused only on the past verb forms. This is believed to be another element that affected the results of the GJT and gave unrealistic readings for the participants’ grammatical awareness. Hence, it is important to adopt a diversity of verb forms in order to get a clearer idea about the L2 learner’s awareness of grammar. Having said that, the participants were relatively able to identify the simple past form, but they struggled with the past progressive and past perfect forms.

References


**Appendix A**

**Oral instructions:**
The following task comprises a group of sentences some of which are not grammatically correct. You are required to read the sentences carefully, and check the box next the sentence you think is ungrammatical. The following verb forms are used in formulating the sentences: the past simple, the past progressive, and the past perfect. Please, do not use random checking for the sentences.

**Appendix B**

**Kindly read the following sentences carefully, then check the box next to the ungrammatical sentence:**

**Example:**

☐ I was happy yesterday.

☒ She is watching TV last night.

**Task:**

1. ☐ I worked every day last week.
2. ☐ He called me while I was eating.
3. ☐ They studied when the electricity went off.
4. ☐ I had just put the phone down, when the doorbell rang.
5. ☐ She had not been studying for the exam.
6. ☐ By the time she found an umbrella, it had raining for ten minutes.
7. ☐ He plays with his friends yesterday.
8. ☐ She didn’t studied hard for her test.
9. ☐ We were playing football when it started to rain.
10. ☐ He had studied English before he come to the UK.
11. ☐ The man had been sleeping before the storm happened.
12. ☐ He knew London so well because he had visited the city several times.
13. ☐ A: What were you doing when you broke your leg?
    B: I drive my bicycle.
14. ☐ They went to cinema last Thursday.
15. ☐ A: Was she eating when you arrived home?
    B: No, she had already eaten before I arrived home.
16. ☐ A: Did you go to the party last night?
    B: No, I have work to do!
17. ☐ I didn't have any money because I have lost my wallet.
18. ☐ While I am writing the email, the computer suddenly went off.
19. ☐ She gained weight because she have been overeating.
20. ☐ He had been studying at the university before he left the UK.