

THE VALIDITY OF GRAMMATICALITY JUDGMENT TASK ON SAUDI EFL LEARNERS

Sami Alanazi, PhD Researcher Applied Linguistics
University of Central Lancashire, UK

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: GJT, SLA, EFL

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 investigation focused on the past verb forms and aspects. The paper is part of a larger study investigating the acquisition of past verb forms by Saudi EFL learners. Furthermore, it will bridge the gap in the literature by focusing the study on Arabic speaking learners.

Background

The use of Grammaticality Judgment Task (GJT) goes back to the late 70's and early 80's (Bialystok, 1979; Gass, 1983; among others). 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. Alternatively stated, the one means of objectifying language is to declare if a provided sentence is suitable or not. Grammaticality judgments comprise intricate behavioral operations that have to be applied carefully and with complete comprehension of their restrictions (Chaudron, 1983; Cowan and Hatasa, 1994; R. Ellis, 1991, 2004, 2005; Gass, 1994; Goss, Ying-Hua, and Lantolf, 1994; Loewen, 2009; Mandell, 1999).

A considerable amount of research has been carried out inside the arena of second language employing grammaticality judgments. For instance, the grammaticality judgment task is mainly a decoding task within which two stages are concerned in resolving to refuse or acknowledge a provided sentence. Primarily, learners have to translate the sentence in some manner. Secondly, they have to establish if it conforms to the sequences of English as signified by their interlanguage. Alternatively stated, a learner has to endeavor to match the sentence using an internalized linguistic structure.

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 & Kaplan, 1998; Gass, 1994; Goss, Ying-Hua, & Lantolf, 1994; Leow, 1996; Mandell, 1999). These researches have been specifically targeted at exploring the legitimacy and dependability of GJTs.

The main construct validity matter concerns what is quantified by a GJT. What sort of awareness do learners utilize when they judge a sentence's grammaticality: implicit knowledge, explicit knowledge or some sort of combination of the two? As observed by Birdsong (1989), "metalinguistic information [from a GJT] are comparable to cheap hot dogs: they are made up of meat but considerable other ingredients as well" (p.69). Unexpectedly, this is a matter that numerous SLA researchers utilizing GJTs are unable to address. Nonetheless Sorace (1996) clearly accepts the challenge:

"It may comprise an additionally intricate job [than is true of native-speaker judgments] to resolve on the sort of norm consulted by learners within the procedure of making a judgment, especially within a learning setting that encourages the growth of metalinguistic awareness. It is a challenge to see whether subjects disclose their thoughts or what their thoughts should be". (p. 385)

Alternatively stated, it may be theorized that when learners are requested to determine the grammaticality of a sentence quickly, they may be more inclined to depend on implicit awareness. However, if they are provided with time, they are capable of acquiring controlled access to explicit awareness. Sorace, too, suggests that a timed process is required to ascertain that the assessment draws on tacit as opposed to metalinguistic awareness. Ellis and Han's (1998) research reinforces this standpoint. They established that measures taken from a time and untimed GJT version of the same GJT factored out individually. Within a principal-components evaluation, the timed GJT loaded on the similar aspect as an oral production assessment, while the untimed GJT loaded on the similar factor as a metalingual-comments score. Ellis and Han noted these two factors as explicit and implicit awareness correspondingly. DeKeyser (2003)

nonetheless offers a cautionary word, observing that time pressure does not assure a measure of implicit awareness. As observed previously, it is feasible that several learners may develop comparatively automatized explicit awareness, which may also be appraised subject to time pressure. Additionally, it is not true that explicit awareness will be implemented by learners if they have the time to do this. They may still opt to depend on their implicit awareness. Certainly, they may be required to, (or instead, to guess) if they do not have the required explicit awareness to assess the grammaticality of a specific sentence. R. Ellis (1991b) and Goss et al. (1994) established that although there may have been the chance to contemplate a judgment, learners at times selected to answer immediately. At best, we may then state that an immediate judgment may more likely mirror implicit awareness, and a deferred judgment may indicate explicit awareness.

Additionally, there is the question of the manner in which to operationalize timed and untimed judgments (i.e., the period of time provided for implementing a judgment). A GJT possibly concerns three main processing functions:

1. semantic processing (i.e., comprehending the significance of a sentence)
2. noticing (i.e., seeking to determine if something is officially wrong with a sentence)
3. reflecting (i.e., contemplating what is wrong regarding the sentence, and probably, the reason for its incorrectness)

An additional challenge associated with the judgment of sentences as being ungrammatical or grammatical is associated with whether the learners really judge the particular structures intended for them to judge by researchers, or alternative structures included within the test sentences. This challenge may be addressed if learners are requested to show or rectify what they believe to be ungrammatical within the sentences. Nonetheless, it is unclear if this improves the legitimacy of a GJT as a measure of explicit awareness. In researches concerning L1 metalingual awareness, the capability to repair sentences at a considerably early age (4years) is observed as mirroring a tacit awareness of the rules of language as opposed to a conscious knowledge. Gombert (1992), states that it mirrors “episyntactic” as opposed to “metasyntactic” conduct. Indeed, children subsequently apply conscious awareness to rectify ungrammatical sentences, but obviously the capability to carry out such a function utilizing tacit awareness does not

vanish. Time may once more be anticipated to comprise an essential aspect; requested to show or rectify a mistake online, L2 learners may be anticipated to depend more on their implicit awareness, while if provided sufficient time, they may have the chance to utilize their explicit awareness.

The dependability of the GJTs employed in specific research has additionally been questioned. Birdsong (1989) highlights the hazards of reaction bias (e.g., an overall inclination to deem sentences as ungrammatical). R. Ellis (1991b) documents three studies within which the similar GJT was dispensed to L2 learners within a single week. The learners altered 22.5%, 31.0%, and 45% of their judgments from one test to the next. Ellis asserts that the GJTs could have been undependable, as the awareness of the L2 learners was not certain, resulting in them inconsistently applying a different set of techniques for providing judgment. Only one of these techniques concerned applying explicit awareness in the state of pedagogical regulations of differing precision. Ellis' translation has been supported by alternative researches that have explored the manner in which learners reach judgments (e.g. Goss, Ying-Hua and Iantolf, 1994). Additionally, learners of L2 have been seen to utilize a greater range of tactics contrasted to native speakers when completing a GJT (Davies and Kaplan, 1998).

In conclusion, as stressed by Birdsong (1989), among others, a GJT does not offer an immediate window for observing the linguistic aptitude of learners, but rather comprises a performance that is inclined by a variety of aspects. The possibility of a GJT offering a measure of explicit awareness may be raised if (a) learners are provided time to assess sentences and to rectify ungrammatical sentences, (b) reactions of learners to the ungrammatical sentences on the assessment (or the sentences that have been considered ungrammatical by the learners) are contemplated individually from their reactions to the grammatical sentences, and (c) the uncertainty of learners in assessing individual sentences is considered.

Methodology

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 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 also given guidance and time to plan their answers prior to the task 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.

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 served as a comparison point for the experimental items (Bullock et al., 2005). In contrast, the verbs in the experimental items were manipulated because they were the focus of the study.

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 I travelled to meet a group of participants. I gave the participants 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.

Results

A one-way analysis of variance (ANOVA) between groups was conducted to compare the results from Grammaticality Judgment Task 1 (GJT 1). 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).

Group	Mean	Std. Deviation	Sig.
Guided-planning group	4.33	1.155	0.604
Semi-guided planning group	3.92	1.832	
No-planning group	3.75	1.288	

Table (1) shows the One-way ANOVA results for the GJT

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 ($M = 3.58, SD = 0.793$).

Group	Mean	Std. Deviation	Sig.
Guided-planning group	4.83	1.115	0.049
Semi-guided planning group	4.33	1.557	
No-planning group	3.58	0.793	

Table (2) shows the One-way ANOVA results for the delayed GJT

A Post Hoc Test and a Complex Contrast Test were employed on Grammaticality Judgment 2 (GJT2) to find out which group caused the significance. I selected the Tukey HSD Post Hoc Test 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.

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
GP Group	SGP Group	.500	.489	.568	-.70	1.70
	NP Group	1.250*	.489	.039	.05	2.45
SGP Group	GP Group	-.500	.489	.568	-1.70	.70
	NP Group	.750	.489	.288	-.45	1.95
NP Group	GP Group	-1.250*	.489	.039	-2.45	-.05
	SGP Group	-.750	.489	.288	-1.95	.45

Table (3) Tukey HSD Post Hoc Test to determine which group caused the significance

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 as it came to (-2). In other words, the no-planning group performance in the grammaticality judgment delayed test, as expected, had fallen back.

Group	GJT1	GJT2 (Delayed)	Gain Score
Guided Planning	52	58	6.00
Semi-Guided Planning	47	52	5.00
No-Planning	45	43	-2.00

Table (4) shows the gain score from the delayed GJT

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 of guidance and planning was conducted

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.

Despite the slight significant change 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-planning group and semi-guided planning was approximate, and the significance was made by the performance of the no-planning group (see table 4).

Conclusion

GJT is one of the widely used tools for Metalinguistic data for SLA research. However, the controversy about GJT is escalating and a great amount of research was conducted to test the validity of this tool (Ellis, 1991; Gass, 1994; Sorace, 1996; White, 2003; Tremblay, 2005). As Tremblay (2005) points out, a concern arises in the literature is the extent to which GJT actually reflects grammatical competence. The finding of this study comes in alignment with this concern and questions the validity of GJT

in measuring grammatical awareness as shown by the statistical significant differences in performance between the two encounters of GJT. Schachter (1989) suggest that learners reject sentences not because they are ungrammatical but because they pose processing problems for them. By the same token, Ellis (1991) suggests that learners “may base their judgment on their existing interlanguage or on what they think the target language norms are” (p.164). Given these suggestions, the complexity of the experimental test items affected the performance of the participants and maybe led them to reject those items.

To sum up, this study makes a potential contribution to the literature on investigating the validity of GJT on Saudi EFL learners. Despite the criticism, GJT is an essential tool for gathering metalinguistic data in the area of SLA. However, as highlighted by Ellis (1991) and Tremblay (2005) among others, careful design and administration of GJT is required to make efficient and reliable.

References:

- Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic accept- ability. *Language*, 72, 32–68.
- Bialystok, E. (1979). Explicit and implicit judgments of L2 grammaticality. *Language Learning*, 29, 81–103.
- Birdsong, D. (1989). *Metalinguistic performance and interlinguistic competence*. New York: Springer-Verlag.
- Bullock, G., Omaki, A., Schulz, B., Schwartz, B. D., & Tremblay, A. (2005). Did they say in Interlanguage that CP is missing, or did they say that in Interlanguage CP is missing? Paper presented at the 8th Bi-Annual Generative Approaches to Language Acquisition Conference, Università degli Studi di Siena, September 9.
- Chaudron, C. (1983). Research on metalinguistic judgments: A review of theory, methods, and results. *Language Learning*, 33(3), 343–377.
- Cowan, R., & Hatasa, Y. A. (1994). Investigating The Validity and Reliability of Native Speaker and Second-Language Learner Judgments about Sentences. In E. Tarone, S. M. Gass, & A.D. Cohen (Eds.), *Research Methodologies in Second-Language Acquisition* (pp. 287-302). Hillsdale, NJ: Lawrence Erlbaum.
- Davies, W., and T. Kaplan. (1998). Native Speaker vs. L2 Learner Grammaticality Judgments. *Applied Linguistics* 19: 183-203.

- DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Oxford: Blackwell.
- Ellis, R. (1991). Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition*, 13(2).
- Ellis, R. (2004). The Definition and Measurement of L2 Explicit Knowledge. *Language Learning*, 54(2), 227–275.
- Ellis, R. (2005). Measuring Implicit and Explicit Knowledge of A Second Languages: A Psychometric Study. *Studies in Second Language Acquisition*, 27.
- Gass, S. (1983). The development of L2 intuitions. *TESOL Quarterly*, 17, 273–291.
- Goss, N., Ying-Hua, Z., & Lantolf, J. (1994). Two heads may be better than one: Mental activity in second-language grammaticality judgments. In E. Tarone, S. M. Gass, & A. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 263–286). Mahwah, NJ: Erlbaum.
- Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 94–112). Bristol, UK: Multilingual Matters.
- Mandell, P. (1999). On the reliability of grammaticality judgment tests in second language acquisition research. *Second Language Research*, 15, 73–99.
- Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistics methodology*. Chicago: The University of Chicago
- Tremblay, A. (2005). On the use of grammaticality judgments in linguistic theory: Theoretical and methodological perspectives. *Second Language Studies*, 24, 129-167.
- Sorace, A. (1996). The use of acceptability judgments in second language acquisition research. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 375–409). San Diego, CA: Academic.
- White, L. (2003). *Second language acquisition and Universal Grammar*. New York: Cambridge University.