NATIVE VS NON-NATIVE STRATEGIES IN RENDERING GRAMMATICALITY JUDGMENTS*

María del Pilar García Mayo
Universidad del País Vasco

ABSTRACT

Grammaticality judgment (GJ) tasks have played an important role in the development of theoretical linguistics, but the study of their nature with learners of a second language (L2) has not received much attention until quite recently. In this paper we examine dyad speak-aloud protocols and individual judgments by 36 Spanish subjects studying English as an L2 and rendering GJs in Spanish and in English. Our findings support those of previous research and show that strategies used in judging sentences in the subjects' first and second languages are not the same. In line with Gass (2001) we suggest that GJs should be used as a data-gathering technique but in combination with other methods in order to obtain more reliable results.

KEY WORDS: Grammaticality judgments, learner strategies, L1 and L2

RESUMEN

Las tareas que hacen uso de los juicios de gramaticalidad han desempeñado un importante papel en el desarrollo de la lingüística teórica pero hasta hace poco tiempo no se les ha prestado demasiada atención cuando las utilizan los aprendices de una segunda lengua (L2). En este trabajo analizamos la producción hablada de parejas y los juicios individuales de 36 informantes españoles que estudian inglés como L2 y ofrecen juicios de gramaticalidad en español e inglés. Nuestros resultados corroboran los de investigaciones previas y demuestran que las estrategias utilizadas para juzgar frases en la primera y la segunda lengua no son las mismas. Sugerimos, de acuerdo con Gass (2001), que los juicios de gramaticalidad deben utilizarse como técnicas de obtención de datos pero en combinación con otros métodos para que los resultados sean más fiables.

PALABRAS CLAVE: juicios de gramaticalidad, estrategias del aprendiz, L1 y L2

1. INTRODUCTION

Grammaticality judgment (GJ) tasks have played an important role in the development of theoretical linguistics, but the study of their nature with learners of a second language (L2) has not received much attention until quite recently (Davies and Kaplan, 1998; Ellis, 1990; Hedgcock, 1993; García Mayo, 1999, 2000, 2003;

Gass, 1994; Munnich *et al.*, 1994; Murphy, 1997). Davies and Kaplan (1998: 183) point out that there are two assumptions underlying the use of GJ tasks: the first one, accepted within much of formal linguistics, is that GJs can provide data representative of a speaker's competence or, as Ellis (1991:163) has put it, "grammaticality judgments provide the best way of studying the 'mental structures and processes that make learning possible [Bley-Vroman, Felix and Ioup 1988]' because it is believed that they obviate the need for the learner to access the processing systems responsible for using the underlying grammar in actual performance". The second, a very important assumption, is that rendering GJs in a second language is essentially the same activity as doing so in one's first language.

It is this second assumption that is addressed in this paper by examining dyad speak-aloud protocols and individual judgments provided by 36 Spanish speakers studying English as a foreign language rendering GJs in Spanish and in English. Our findings support those by Davies and Kaplan and show that the strategies used in rendering GJs in L1 and L2 are not the same. We believe that this finding should not discourage the use of GJs as a data-gathering technique. They just point to the often-mentioned need to use GJs in combination with other methods in order to obtain more reliable results.

2. REVIEW OF THE LITERATURE

It has been a standard practice of both theoretical linguists and second language (L2) acquisition researchers to support theoretical claims by means of the results obtained from GJs tasks. In the second language acquisition (SLA) research area there are basically two positions concerning GJs. Some researchers (Bley-Vroman *et al.*, 1988; Chaudron, 1983; Gass, 1994; White, 1989) consider that this elicitation technique provides valid data, whereas other researchers (Birdsong, 1989; Ellis, 1990, 1991; Goss *et al.*, 1994; Kellerman, 1985) argue that the results obtained from GJs should be considered with great caution because of the possible influence of the L1 when making judgments or the metalinguistic knowledge of the L2.

As Sorace (1996: 385) points out:
It can be a [...] complex task [...] to decide about the kind of *norm* consulted by learners in the process of producing a judgment, particularly in a learning environ-

MARÍA DEL PILAR GARCÍA MAYO **184**

ment that fosters the development of metalinguistic knowledge. It is difficult to tell whether subjects reveal what they think or what they think they should think.

It is for that reason that serious questions have been raised about the *validity* of judgments, that is, the relationship between judgments and the state of knowledge they are claimed to reflect (grammatical competence) and the *reliability* of judgments, the degree of consistency among the judgments produced by different subjects (intersubject consistency) or by the same subject (intrasubject consistency) in different replications of a test. Sorace (1996: 376-377) mentions that one of the arguments raised against the supposed validity of GJs is that they may be affected by factors that are extralinguistic in nature and cannot be isolated. These factors will provide spurious intuitions, which will differ from genuine intuitions originating from the subjects' internalised grammar (Botha, 1973). Some of the most relevant extralinguistic factors are:

(I) parsing strategies (Snow, 1974): difficulty in parsing may be responsible for the rejection of perfectly grammatical sentences and ease of parsing for the acceptance of ungrammatical ones.
(II) context and mode of presentation (Snow, 1974): a sentence of dubious grammaticality is more likely to be judged as ungrammatical if placed after a set of clearly grammatical sentences, or as grammatical if placed after a set of clearly ungrammatical sentences.
(III) pragmatic considerations: subjects tend to prefer a reading that represents the most frequent interpretation of a sentence and requires fewer assumptions about previous discussion.

There are a number of ways in which GJs have been used. Gass (1994) summarizes them in three main areas: (I) researchers differ in whether or not they ask learners to correct sentences that are judged ungrammatical (eg. Munnich *et al.*, 1994); (II) sometimes learners judge individual sentences, some others they are asked to provide preference judgments (i.e. select the more appropriate sentence among the ones provided); (III) learners are given a number of possible responses to choose from (responses may be dichotomous —a sentence can be either grammatical or ungrammatical; or there may be a range of possibilities that include the degree of confidence a learner has in making responses). Researchers may also vary widely in the number of sentences subjects are asked to give judgments about, ranging from 30 to 40 to more than 200.

Recently, speak-aloud protocols when administering GJs tasks have been used as an innovation that may help researchers to discover what underlies standard judgments of grammaticality[1]. In what follows we briefly summarize the content of

---

[1] A distinction is normally made in the literature between standard GJ tasks in which subjects are allowed as much time as necessary to complete the task and may be asked to correct sentences when considered ungrammatical versus a timed GJ task (Sorace's (1996) Magnitude Estimation).

three previous studies employing speak-aloud protocols. The first study was carried out by Ellis (1991); the subjects of his study were 21 native speakers of Chinese studying in postgraduate courses in London. Ellis defines them as "*sophisticated and advanced —well suited to a grammaticality judgment test*" (op. cit.: 172). Once the subjects had completed a GJ task with 40 sentences (the target structure was dative alternation), eight students were selected for an individual think-aloud task, which was carried out one week after the original test. Besides isolating seven types of strategies used by these L2 learners[2], Ellis concludes that learners in this study displayed considerable inconsistency in their judgments and that they appear to rely extensively on intuition in order to make a judgment.

A second study was carried out by Goss et al. (1994), who tested 52 native speakers of English who were learning Spanish. Besides the individual speak-aloud protocol used by Ellis, they chose a group speak-aloud procedure (in dyads or triads). Within a sociocultural view of mind, the authors consider it essential to compare individual and group performance in order to determine if there are any marked differences. The point they want to establish is that "the discourse produced in the joint format not only informs us about how learners respond to judgment tasks collectively but also represents what is going on in the mind of individual learners from the same population when they respond to the same task in the individual format" (*op. cit.*: 268).[3] Goss et al. report no significant difference between group and individual responses and identify three main strategies used by the L2 learners in making GJs: translation, recall of explicit metaknowledge and feel. The elementary Spanish L2 learners in their study relied on translation and recall of explicit metaknowledge; the more advanced learners also rely on translation, explicit metaknowledge and memory to make more accurate judgments.

To our knowledge, the most recent study on strategies in rendering GJs is the one by Davies and Kaplan (1998) who tested 37 native speakers of English who were learning French as their L2. They gathered data both from dyads and individuals and their main conclusions were that (I) the general response patterns of the dyads matched that of the individuals, and (II) the strategies used by the English native speakers to judge grammaticality in English and French differed in at least two important aspects (a) the number of strategies employed per sentence, and (b) the particular strategies employed in each context.

3. THE PRESENT STUDY

The present study is modeled on Davies and Kaplan (1998) in its overall design but our subjects are bilingual (Basque/Spanish) learning English as a foreign

MARÍA DEL PILAR GARCÍA MAYO   **186**

---

[2] Those were: Feel, rehearse (learners repeated a test sentence or part of it), rehearse alternative version, try to access explicit knowledge, use of analogy (comparison with a previous test sentence), evaluation of a sentence and guessing.

[3] For recent work supporting the idea that learning processes surface during collaborative work in problem-solving activities see Swain (1998) and Swain and Lapkin (1995; 2001).

language. The rationale for the replication was that different results could be obtained, as our subjects had been exposed to the L2 for a longer period of time and they have two languages as their L1s. The goals of the study are (I) to identify the strategies used and determine whether they are the same identified by Davies and Kaplan, (II) to determine whether those strategies are different from the ones the subjects use in judging grammaticality in one of their L1s (Spanish). The rest of this section describes the subjects that participated in the study, the GJs tasks and procedures used in data collection and the findings.

3.1. THE SUBJECTS

The subjects in this study were 36 Basque/Spanish bilinguals studying English as a foreign language in an institutional setting. Their mean age was 16.3 at time of testing and they had been exposed to the foreign language for 6 years (approximately 594 hours). As a control group, we tested 10 native speakers of English, North American college students who had come to a major Spanish university to study Spanish as members of the USAC (University Study Abroad Consortium). Their mean age was 21.5.

3.2. DATA COLLECTION PROCEDURES

Out of the 36 subjects , 18 were tested in 9 dyads and 18 individually. The subjects were tested on 12 Spanish sentences and 12 English sentences (see Appendix). Seven of the Spanish sentences were ungrammatical and five grammatical. The structures chosen were representative of those elicited from native speakers of Spanish by researchers working within a generative framework (see Demonte Barreto, 1989). That is, they were not excessively simple because some discussion needed to be generated in order to determine the strategies used when judging grammaticality in the subjects' L1. As for the English sentences, eight were ungrammatical and four grammatical. They were selected because the sentences feature typical problems for Spanish learners of L2 English, some of them explicitly mentioned in the L2 classroom (null subjects, subject/verb inversion) and one that is not (extraction of wh-subjects from subordinate clauses). The pieces of paper provided to both the dyads[4] and the individual learners contained the 12 sentences in Spanish and the 12 in English with the choices 'correct', 'incorrect' and 'not sure'. Subjects were told to decide what they thought of each sentence and circle the relevant choice. No time limit was given because interaction between the two members of the dyad was

---

[4] Following Goss et al. (1994) and Storch (1998), only one copy of the test sentences was given to each dyad in order to ensure that the two participants would contribute to the response pattern.

an issue at stake here so that we could determine the type of strategy used to arrive at the judgment. The sessions for the dyads were tape recorded and later transcribed. The subjects working individually simply answered on the pieces of paper provided. The English native speakers were tested in the same way as described for the dyads.

4. FINDINGS

4.1. RESULTS FROM GJS BY DYADS AND INDIVIDUALS

We first provide information on the data obtained from the judgments (individual and in dyads) of the Spanish and English sentences. Table 1, with the results for the sentences in Spanish, shows that the performance of dyads and individuals matches well. Table 2 features the results of a Mann-Whitney test run on these data indicating that individual judgments were not significantly different from the ones offered by the dyads in the speak-aloud protocol.

TABLE 1. JUDGMENTS OF SPANISH SENTENCES FOR DYADS AND INDIVIDUALS							
		DYADS (N=9)			INDIVIDUALS (N=18)		
SENTENCE	JUDGMENT	CORRECT	INCORRECT	NOT SURE	CORRECT	INCORRECT	NOT SURE
1	*	2	6	1	6	10	12
		22%	67%	11%	33%	56%	11%
2	OK	9	0	0	18	0	0
		100%			100%		
3	OK	6	2	1	14	4	0
		67%	22%	11%	78%	22%	
4	*	0	9	0	0	18	0
			100%			100%	
5	*	0	9	0	0	18	0
			100%			100%	
6	*	0	9	0	0	18	0
			100%			100%	
7	*	0	9	0	0	18	0
			100%			100%	
8	OK	9	0	0	18	0	0
		100%			100%		
9	*	0	9	0	0	18	0
			100%			100%	

10	OK	7	2	0	14	4	0
		78%	22%		78%	22%	
11	*	0	9	0	0	18	0
			100%			100%	
12	OK	8	0	1	18	0	0
		89%		11%	100%		

TABLE 2. MANN-WHITNEY STATISTICS FOR OVERALL INDIVIDUAL VS. OVERALL DYAD JUDGMENTS	
SENTENCE	*p*-VALUE
1	0.5765
2	1.0000
3	0.6616
4	1.0000
5	1.0000
6	0.3358
7	1.0000
8	1.0000
9	1.0000
10	1.0000
11	1.0000
12	0.1817

Table 3 indicates that the pattern of results is similar for the English sentences. That is, the percentages of 'correct,' 'incorrect' and 'not sure' answers are roughly equivalent but, as Table 4 shows, there are two exceptions: sentence 3 (*Who do you think will win the prize?*) and sentence 10 (*Who did you say that arrived late?*), which feature extraction of wh-words functioning as subjects of the embedded clause. Curiously enough, this is the type of sentence that has been shown to be clearly dissociated from the other properties first argued to be encompassed by the *pro-drop* parameter (García Mayo, 1998, 2003; Liceras, 1989; White, 1985). The Mann-Whitney statistics for these two sentences are significantly different. Except for those two sentences, the response pattern of the dyads matches that of the individuals for the English test sentences. Therefore, our results support Davies and Kaplan's first finding, that is, the performance of dyads matches well with that of the individuals.

TABLE 3. JUDGMENTS OF ENGLISH SENTENCES FOR DYADS AND INDIVIDUALS							
		DYADS (N=9)			INDIVIDUALS (N=18)		
SENTENCE	JUDGMENT	CORRECT	INCORRECT	NOT SURE	CORRECT	INCORRECT	NOT SURE
1	*	1 11%	7 78%	1 11%	7 39%	10 56%	1 5%
2	*	2 22%	6 67%	1 11%	3 17%	14 78%	1 5%
3	OK	7 78%	1 11%	1 11%	6 33%	8 44%	4 22%
4	OK	6 67%	2 22%	1 11%	11 61%	0	7 39%
5	*	3 33%	5 55%	1 11%	7 39%	8 44%	3 17%
6	*	3 33%	5 55%	1 11%	3 17%	13 72%	2 11%
7	*	9 100%	0	0	14 78%	3 17%	1 5%
8	OK	5 55%	2 22%	2 22%	11 61%	5 28%	2 11%
9	*	2 22%	6 67%	1 11%	6 33%	10 56%	2 11%
10	*	7 78%	1 11%	1 11%	8 44%	9 50%	1 5%
11	*	2 22%	6 67%	1 11%	2 11%	15 83%	1 5%
12	OK	8 89%	1 11%	10	15 83%	3 17%	0

TABLE 4. MANN-WHITNEY STATISTICS FOR OVERALL INDIVIDUAL VS. OVERALL DYAD JUDGMENTS	
SENTENCE	p-VALUE
1	0.2556
2	0.5914
3	0.0360*
4	0.9038
5	0.6730

6	0.3701
7	0.1447
8	0.9533
9	0.5765
10	0.0754**
11	0.3568
12	0.7383

4.2. Identification of strategies

As mentioned above, the tape-recorded sessions of the nine dyads were transcribed verbatim. The author, with the help of a doctoral student, identified the strategies used by the members of the dyads, which coincided with the ones previously identified by Davies and Kaplan for their subjects (1998:190-91). Inter-rater agreement was 96%. In what follows we provide a brief definition of each of the strategies used in judging grammaticality in L1 and L2 and we illustrate them with examples from our data base.

Strategies used in judging L1 grammaticality other indications that some other strategy might be involved and where subjects

(I) *Feel:* The subjects provide an intuitive response ("no pauses or hesitations or stated that the sentence 'sounded' or 'looked' good" (Davies and Kaplan, 1998: 202, note 6).
  A. Carlos fue recomendado que fuera   a ver a un psiquiatra
    Carlos was recommended that go (3rd sg.subjunctive) to see a psychiatrist
    Carlos was recommended to see a psychiatrist
  B. Está   mal
    be (3rd. sg.) bad
    It is bad
  A. Sí, suena   mal.
    yes, sound (3rd. sg.) bad
    Yes, it sounds bad
(II) *Meaning-based*: The subjects make an attempt to figure out what the sentence means:
  A. Creo   que no lo conozco
    Believe (1st sg.) that no it know (1st sg.)
    I believe I don't know him
  B. Eso está bien ¿conoces   a Juan? y me   dices
    that is   good know (2nd sg.) Juan and me (acc.) say (2nd sg.)

tú,   no, creo   que no lo conozco.
you, no, believe (1st. sg.) that no him know (1st sg.)
That is good. Do you know Juan? And you tell me, no, I believe I don't
know him.

(III) *Repair*: The subjects attempt to repair a perceived problem with the sentence:
A. ¿Cómo lamentas   que tu hijo se haya   portado?
how   regret (2nd. sg.) that your son refl. have (3rd sg. subj.) behave (p.p.)
¿Cuánto lamentas...
how much regret (2nd. sg.)
How do you regret that your son has behaved himself? How much do
you regret...
B. Sí, o lamentas   sólo, quitando el cómo¿no? Está   mal.
yes, or regret (2nd. sg.) alone, drop (ger.) the how no is (3rd. sg.) bad
Yes, or just regret, dropping the how right? It is bad

(IV) *Learned:* The subjects make reference to a grammatical rule or other explicit
knowledge related to the sentence:[5]
A. Creo que no lo conozco...   *lo*...   no porque *lo* es una cosa no...
believe that no him know (1st. sg.) him/it no because him/it is a thing
no...
I believe I don't know him... him no because him is a thing no...
B. No, pero *lo* se puede poner con hombres sí, o con hombres no, o algo
así...
no but him reflex can put with men yes or with men no, or some-
thing like that...
No, but him can be used with men, yes, or with men no... or some-
thing like that...
A. Bien, entonces...
Well, then...
B. Hombre, se tendría   que poner *le*
man,   tener (cond.) that put le
Well, one should use *le*
A. Pues no estamos   seguros, no
well no be (1st. pl.)   sure   no
Well, we are not sure, no

---

[5] This example is difficult to translate. Lo is a clitic pronoun which in standard Spanish is
used to refer to both masculine animate and inanimate entities functioning as direct objects in the
sentence. As pointed out by Landa and Franco (1999:572) "The leísmo found in Basque Spanish is
half-way between the leísmo of educated speech that replaces the lo forms that refer to animate
masculine entities by le and the "radical" leísmo that substitutes all the etymological accusative clitics
by the etymological dative le. In other words, in Basque Spanish only I[ndirect] O[bject] and D[irect]
O[bject] clitics that refer to animate entities are rendered by the clitic le."

Strategies in rendering L2 grammaticality

(I) *Feel*
      A. Yesterday I met a girl but I forgot her name
      B. Bien
          fine
      A. Sí
          Yes

(II) *Meaning based*
      A. The postman came. Have arrived three letters... the postman came y
              después... (and then...)
      B. he has arrived... y (and) arrived encima (besides) three letters.
      A. no tiene   mucho sentido, no, y trajo   tres cartas...
          no have (3rd. sg.) much sense no and bring (3rd. sg.) three letters
          It does not make much sense, that is to say, and he brought three letters.
      B. porque si no no entiendo   bien esto.
          because if no no understand (1st. sg.) well this
          because if that is not the case I do not understand this properly

(III) *Repair*
      A. walked the boy very far ... incorrect order
      B. the boy walked very far
      A. ok

(IV) *Learned*
      A. I wonder where Ana is going... está   bien
                                be (3rd. sg.) good
                                it is good
      B. Sí
          Yes
      A. porque es   una pregunta indirecta
          because be (3rd. sg.) a   question indirect
          because it is an indirect question

Three other strategies were identified in judging L2 grammaticality. They are illustrated in what follows:

(V) *Translation*: The subjects translate the L2 sentence into the L1 and base their judgment on this translation:
      A. Who do you say that arrived late?.. a ver, ¿quién has dicho   que
                                  see who have (2nd. sg.) say (p.p.) that
          ha   venido?
          have (3rd. sg.) come (p.p.)
      B. ok, correct

(VI) *Analogy*: The subjects seem to judge the sentence based on some perceived similar structure:

A. Walked the boy very far... sí , otra vez en el orden está   mal ahí ¿no?[6]

                  yes again in the order be (3rd. sg.) bad there no

B. sí, the boy walked very far.

   yes

(VII) *Guess*: The subjects provide some verbal clue that they are guessing whether the sentence is grammatical or not:

A. I wonder where Ana is going... pues

                       then

B. el orden no estamos   seguros...

   the order no be (1st. pl.)   sure

   we are not sure about the order

A. where is Ana going...

B. sí, parece   que sí

   yes seem (3rd. sg.) that yes

   Yes, it seems so

Once the strategies were identified in the dyads' interaction, we had to compare the use of these strategies in both the L1 and the L2 context in order to determine whether those uses differed or not. This involved the independent scoring of the verbatim transcripts and the noting of each instance of the seven identified strategies. Table 5 shows the results:

TABLE 5. STRATEGIES USED TO JUDGE THE SPANISH AND ENGLISH SENTENCES						
	SPANISH NSs				ENGLISH NSs	
	SPANISH		ENGLISH		ENGLISH	
Feel	83	76.8%	45	41.6%	48	80%
Repair	40	37%	47	43.5%	15	25%
Learned	11	10.18%	19	17.59%	6	10%
Meaning	14	12.9%	23	21.29%	0	0
Translation	0	0	22	20.37%	0	0
Analogy	0	0	1	0.9%	0	0
Guess	0	0	10	9.25%	0	0
Single strategy	83	76.8%	55	50.9%	49	81.6%

---

[6] The learner refers to sentence 2, which featured the same problem.

What we find when looking at the results above is that the particular strategies to render grammaticality judgments are different depending on whether the sentence is in the subjects' L1 or in the L2 they are learning, thus supporting Davies and Kaplan's (1998) findings. First of all, for the subjects in this study, 76.8% of the sentences in Spanish were judged by using a single strategy; this percentage drops in a significant way to 50.9% (p.<.05) when the subjects judge the English sentences. Secondly, we find a wider variety of strategies used when judging grammaticality in the L2 (recall that three additional strategies were identified). Thirdly, the rate of usage of the strategies is different in the two contexts. In the L1 context, the subjects use the *feel* strategy in 83 instances (76.8%), the highest number compared to the other strategies used. By contrast, the subjects use the *feel* strategy only in 45 instances (41.6%) when judging the L2 sentences (significant difference p.<.05). In this context, it is the *repair* strategy the one that is more frequently used but *meaning-based* and *translation* are used frequently as well. The differences between the use of the *learned* and *meaning-based* strategies is significant (although only at p.< .10) between the two contexts, although the differences are not as marked as the ones observed by Davies and Kaplan. The reason for this difference in the use of learned rules could be that the North American subjects in their study have very little metalinguistic knowledge of their own language but much more about the foreign language (French) they were studying in a formal classroom setting. Our subjects, however, are familiar with metalinguistic explanations both about Spanish and about English.

The last point that needs to be made regarding the results presented in Table 5 is that the North American students who acted as a control group judging the English sentences feature a response pattern very similar to the one by our subjects when judging the sentences in Spanish (specifically *feel* Spanish L1: 76.8% and English L1: 80%; *learned* Spanish L1: 10.18% and English L1 10%). These results suggest that the different grammatical structures included in the sentences are not responsible for the differential use of strategies by our subjects in the L1 and L2 context, that is, the results cannot be claimed to be an artifact of the different type of sentence used to study GJs in L1 and L2.

5. CONCLUSION

This study has presented the results obtained from the analysis of grammaticality judgments rendered by the same group of subjects in Spanish (L1) and English (L2), both individually and in dyads. We have shown that the response pattern shown by individuals and dyads in L1 and L2 matches well. We have also shown that different strategies were used in the two contexts when providing information about grammaticality. These results support previous work by Davies and Kaplan (1998), who had also used the same set of subjects to respond to grammaticality judgment tasks individually and in dyads.

Obviously, much work needs to be done. A follow-up study in which dyads provide grammaticality judgments in Basque needs to be designed and learners

with higher levels of proficiency in the language need to be tested. What the present study has shown is that, as any other data-gathering technique, GJ tasks are not without limitations (Gass, 2001) and future work should addressed both design and administration issues if we want them to be a valid and reliable measure of L2 learner competence. We are not claiming that GJs tasks should not be used in L2 research: what we want to stress is the idea that researchers should not only improve on their GJ design but also use them in combination with other data-gathering techniques (García Mayo, 2003).

APPENDIX

A. Spanish sentences
    1. *Carlos fue recomendado que fuera a ver un psiquiatra.
    2. ¿Dónde dices que viste a María?
    3. Ana fue obligada a cumplir sus responsabilidades.
    4. *¿Qué coche le prestó Juan dinero a su hijo para que pudiera comprar?
    5. *Juan nos dijo salir inmediatamente.
    6. *¿Cómo lamentas que tu hijo se haya portado?
    7. *Nos fuimos sin despedirse.
    8. ¿Quién dices que vendrá a cenar?
    9. *Yo creo a Mario ser un buen arquitecto.
    10. No sé lo que compró Ana.
    11. *¿Qué libro te interesó la noticia de que mi jefe compró?
    12. Creo que no lo conozco

B. English sentences
    1. *We will be late for school if don't take this bus
    2. *Slept the baby for three hours.
    3. Who do you think will win the prize?
    4. It seems that Patricia is sad.
    5. *My sister is tired because came home late last night.
    6. *The postman came. Have arrived three letters.
    7. *Which movie do you think that will be on television this evening?
    8. I wonder where Ana is going.
    9. *Is raining very hard today.
    10. *Who did you say that arrived late?
    11. *Walked the boy very far.
    12. Yesterday I met a girl but I forgot her name.

WORKS CITED

BIRDSONG, David. *Metalinguistic Performance and Interlanguage Competence*. New York: Springer, 1989.

BLEY-VROMAN, Robert, Sasha FELIX, & Georgette IOUP. "The Accessibility of Universal Grammar in Adult Language Learning." *Second Language Research* 4 (1988): 1-32.

BOTHA, Richard P. *The Justification of Linguistic Hypotheses*. The Hague: Mouton, 1973.

CHAUDRON, Craig. "Research on Metalinguistic Judgments: A Review of Theory, Methods and Results." *Language Learning* 33.3 (1983): 343-377.

DAVIES, William.D. & Tamara I. KAPLAN. "Native Speaker vs L2 Learner Grammaticality Judgments." *Applied Linguistics* 19.10 (1988): 343-203.

DEMONTE BARRETO, Violeta. *Teoría sintáctica: de las estructuras a la reacción.* Madrid: Síntesis, 1989.

ELLIS, Rod. "Grammaticality Judgments and Learner Variability." *Variability in Second Language Acquisition*. Ed. H. Burmeister & P. Rounds. Eugene: Department of Linguistics, University of Oregon at Eugene, 1990. 25-60.

—— "Grammaticality Judgments and Second Language Acquisition." *Studies in Second Language Acquisition* 13 (1991): 161-186.

GARCÍA MAYO, María del Pilar. "Age, Length of Exposure and Grammaticality Judgments in the Acquisition of English as a Foreign Language." *Age and the Acquisition of English as a Foreign Language.* Ed. M.P. García Mayo & M.L. García Lecumberri. Clevedon: Multilingual Matters , 2003. 94-114.

—— "Grammaticality Intuitions of Bilingual and Monolingual Basque EFL Learners." *Estudios de Lingüística Aplicada* 29 (1999): 71-82.

—— "Grammaticality Judgments, Metalinguistic Awareness and the Age Factor in EFL." *Proceedings of the XXII International AEDEAN Conference*. Ed. P. Gallardo & E. Llurdá. Lleida: Universitat de Lleida, 2000. 561-566.

—— "The Null Subject Parameter in Adult Second Language Acquisition." *Atlantis* 20.1 (1988): 47-58.

GASS, Susan M. "Innovation in Second Language Research Methods." *Annual Review of Applied Linguistics* 21 (2001): 221-232.

—— "The Reliability of Second Language Grammaticality Judgments." *Research Methodology in Second Language Acquisition*. Ed. E. Tarone, S.M. Gass & A Cohen. Mahwah, N.J.: Erlbaum, 1994. 303-322.

Goss, Nancy, Zhang Ying-Hua, & James P. Lantolf. "Two Heads May Be Better Than One: Mental Activity in Second Language Grammaticality Judgments." *Research Methodology in Second Language Acquisition*. Ed. E. Tarone, S.M. Gass & A. Cohen. Mahwah, N.J.: Erlbaum, 1994. 263-286.

Hedgcock, John. "Well-Formed vs Ill-formed Strings in L2 Metalinguistic Tasks: Specifying Features of Grammaticality Judgments." *Second Language Research* 9 (1993): 1-21.

Kellerman, Eric. "Dative Alternation and the Analysis of Data: A Reply to Mazurkewich." *Language Learning* 35.1 (1985): 91-106.

Landa, Alazne & Jon Franco. "Converging and Diverging Grammars." *ASJU* 33.2: 569-581.

Liceras, Juana M. "On Some Properties of the 'Pro-Drop' Parameter: Looking for Missing Subjects in Non-Native Spanish." *Perspectives on Second Language Acquisition*. Ed. S.M. Gass & J. Schachter. Cambridge: Cambridge UP, 1989. 109-133.

Munnich, Edward. Suzanne Flynn & Gita Martohardjono. "Elicited Imitation and Grammaticality Judgment Tasks: What They Measure and How They Relate to Each Other." *Research Methodology in Second Language Acquisition*. Ed. E.Tarone, S.M. Gass & A. Cohen. Mahwah, N.J.: Erlbaum. 1994. 227-243.

Murphy, Victoria A. "The Effect of Modality on a Grammaticality Judgment Task." *Second Language Research* 13.1 (1997): 34-65.

Snow, Catherine. "Linguists as Behavioural Scientists: Towards a Methodology for Testing Linguistic Intuitions." *Linguistics in the Netherlands 1972-1973*. Ed. A. Kraak Assen: Van Gorcum, 1974.

Sorace, Antonella. "The Use of Acceptability Judgments in Second Language Acquisition Research." *Handbook of Second Language Acquisition*. Ed. W.C. Ritchie & T.K. Bathia. New York: Academic, 1996. 375-409.

Storch, Neomy. "Comparing Second Language Learners' Attention to Form across Tasks." *Language Awareness* 7.4 (1998): 176-191.

Swain, Merrill. "Focus on Form through Conscious Reflection." *Focus on Form in Classroom Second Language Acquisition.* Ed. C. Doughty & J. Williams. Cambridge: Cambridge UP, 1998. 64-81.

Swain, Merrill & Sharon Lapkin. "Problems in Output and the Cognitive Processes They Generate: A Step Towards Second Language Learning." *Applied Linguistics* 16 (1995): 371-391.

—— "Focus on Form through Collaborative Dialogue: Exploring Task Effects." *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*. Ed. M. Bygate, P. Skehan & M. Swain. London: Longman, 2001. 99-118.

White, Lydia. "The Pro-Drop Parameter in Adult Second Language Acquisition." *Language Learning* 35 (1985): 47-62.

—— *Universal Grammar and Second Language Acquisition*. Amsterdam: John Benjamins, 1989.