

## Grammar Sections of EFL Classroom Tests in Secondary Schools from Tucumán

Jorge Alejandro Abboud

Universidad Nacional de Tucumán, Argentina

[jorge.abboud@filo.unt.edu.ar](mailto:jorge.abboud@filo.unt.edu.ar)

(first received: March 26, 2023; accepted May 1, 2023)

### Abstract

One interesting approach to gain an insight into the treatment of grammar in our classrooms consists in analyzing the test tasks we design to assess it. This article presents a study on the grammar sections of 60 EFL written tests used in secondary schools from Tucumán. The study involves an analysis of the design (based, mainly, on Bachman and Palmer, 1996) and the tasks included (based, partly, on Celce-Murcia, 2005; Larsen-Freeman, 2003, 2009, and Purpura, 2004). The analysis of the design of the grammar section considers its inclusion, relative salience and location. In turn, the tasks are classified according to context and type of response. Results show certain inconsistency in the design of test sections and a tendency to present decontextualized grammar tasks that elicit explicit knowledge about form. These findings may serve as a basis on which to make informed decisions regarding the teaching and testing of grammar.

*Keywords:* EFL classroom tests, grammar sections, grammar test tasks

### Resumen

Un enfoque interesante para obtener una idea del tratamiento de la gramática en nuestras clases consiste en analizar las tareas que diseñamos para evaluarla. Este artículo presenta un estudio sobre las secciones de gramática de 60 pruebas escritas de inglés utilizadas en instituciones de educación secundaria de Tucumán. El estudio abarca un análisis del diseño (basado, principalmente, en Bachman y Palmer, 1996) y las tareas incluidas (basado, en parte, en Celce-Murcia, 2005; Larsen-Freeman, 2003, 2009 y Púrpura, 2004). El análisis del diseño de la sección de gramática considera su inclusión, su prominencia relativa y su ubicación. Por su parte, las tareas se analizan según el contexto y el tipo de respuesta. Los resultados muestran cierta inconsistencia en el diseño de las secciones y una tendencia a presentar tareas gramaticales descontextualizadas que demandan conocimiento explícito sobre la forma. Estos hallazgos pueden servir como base para la toma de decisiones razonadas en torno a la enseñanza y la evaluación de la gramática.

*Palabras clave:* pruebas de inglés lengua extranjera, secciones de gramática, tareas evaluativas de gramática

### Introduction

Classroom tests cannot be analyzed in isolation from the context in which they were designed. As Bachman and Damböck (2018) put it, “[i]f you want to understand classroom-based language assessment, you need to think about it in the context of the language classroom” (p. 25). Out of the many possible characteristics of the courses from which the tests under analysis in this paper were taken, the students’ level of proficiency is key to this study. Although the courses belong to state and private secondary schools, the students share the same level, within A1 and A2 according to the Common European Framework (CEFR). The fact that they have the same level may be explained mainly by the hours of instruction the schools selected devote to this subject, either 3 or 4 weekly 40-minute lessons. Given the purposes of this study, which delves into the analysis of grammar sections in tests, no further distinctions between these two types of schools are made.

Classroom tests also imply a close relationship between teaching and testing. As this article focuses on grammar sections, the main objective of this work is to trace some features of these test sections and to try to establish certain connections between our teaching and testing practices regarding grammar. The findings, then, could shed some light on how grammar is treated in our classrooms

The treatment of grammar has evolved with the advent of new theories about the nature of language and language acquisition. As a result, the notion of ‘grammar’ has been the field of many linguistic and pedagogical battles over decades. This gets even more bellicose in English Language Teaching (ELT) pedagogy, where we have witnessed radical changes: from classical perspectives that place grammar at the core of any second/foreign language (L2) course to extreme communicative versions that completely ignore it. Similarly, the testing of grammar has developed together with the methods to teach it, depending on the changes in the definition of what grammar is. Regarding the different techniques used to test grammar over time, Purpura (2004) mentions reciting rules, providing the translation of a passage, choosing the correct option, supplying a correct word or phrase, constructing sentences, revising the accuracy of a sentence/utterance, among others. Many of these methods co-exist nowadays, as will be observed in the results.

Despite the changing attitudes towards the role of grammar in ELT, its status has drawn little attention on part of local researchers. As a brief overview of works that give hints on the position of grammar in our teaching and testing practices, we will refer to a study that focuses on textbooks and one that deals with tests.

López Barrios and Villanueva de Debat (2011) present a retrospective study on textbooks produced in Argentina from 1886 to 1999. Although their interest lies in the shift from the Grammar-Translation Method to Communicative Language Teaching (CLT) in accordance with the demands of varying curriculums, we can have an outlook over the role of grammar in the books they examine. From their analysis, we can trace the shift that accompanies the evolution of the notion of grammar over the last 130 years. In the traditional, classical Grammar-Translation method, grammar was almost a synonym of language, as can be inferred from the (sometimes derogatorily used) name given to the method and from the title of the 1886 textbook, *Gramática Inglesa*, which has the word ‘grammar’ in it instead of ‘language,’ for instance. As the authors suggest, gradually over many years, the move goes on towards a more skills-based, implicit-knowledge treatment of grammar, though never directly or linearly, in the more communication-based books.

Despite this shift in orientation in English as a Foreign Language (EFL) materials produced in Argentina, Abboud (2021) reveals the centrality of grammar as the object of assessment in tests. Although his study focuses on the reading section of secondary school tests in Tucumán, some relevant data on grammar can be found. The results show that the grammar section is present in 98% of the tests analyzed, which reflects the central importance this section has in relation to the others, since vocabulary and reading sections are present in 68% and writing in 36%. These findings on tests may serve to have a brief glimpse at some relevant aspects of the treatment of grammar.

Given the local context, research into the testing of grammar seems essential in Argentina. This article aims to start filling this research gap by presenting an analysis of the grammar section of written tests in terms of its design and its tasks. The analysis of the design covers its inclusion, its location, and its salience. The tasks, in turn, are analyzed according to context and type of response. The overall framework of implicit and explicit knowledge of the dimensions of grammar, which contributes to the identification of the object of assessment, enriches the qualitative analysis of the samples provided and leads to establish coherent connections between teaching and testing.

### Theoretical framework

Since this study analyses grammar sections in written tests, it is assumed that it does not cover all the ways grammar can be measured. What do we mean by this? When we score students’

performance in any skill, we may take into account students' grammatical knowledge as well, which means that it is integrated into reading, writing, listening and/or speaking. In such cases, we consider grammatical knowledge as one component of a larger construct. For instance, if we deem it as a type of knowledge that is necessary for students to write a narrative text, then we may include grammar as one aspect in a scale, together with, say, vocabulary and discourse organization, with specific points assigned to each. As Purpura (2004) puts it, "[I]n many assessment contexts today, knowledge of grammar may be inferred from the ability to use grammar correctly while reading, writing, listening to or speaking the L2" (p.4).

The distinction that is made here between sections focused on grammar and grammar as integrated into other skills brings along the concepts of explicit and implicit knowledge. Drawing on Ellis's (2009) distinction between these two types of linguistic knowledge, we can briefly state that implicit knowledge is tacit, intuitive, and procedural, whereas explicit knowledge is conscious and declarative. This means that when students give explanations about how language works, provide some grammatical rule, choose one form over others and are conscious of it, they may be demonstrating their explicit linguistic knowledge. In contrast, when students use forms and structures with some degree of automaticity during their linguistic performance, without devoting considerable attentional resources to them, they may be providing evidence of their implicit linguistic knowledge.

As Ellis (2009) points out, "implicit knowledge is only evident in learners' verbal behavior" (p.13). This idea is key to this study. Since students' verbal behavior is not always clearly seen in grammar sections of tests due to the type of responses (as discussed below), then it can be said that the tasks analyzed here measure explicit knowledge of grammar almost exclusively. This is so because students generally have to show that they know the forms (the past participle of a verb, for instance), but not that they know how to use these forms meaningfully and appropriately. This difference is better explained with reference to Larsen-Freeman's (2003) notion of the dimensions of grammar. It is beyond the scope of this paper to analyze each task in the corpus under her framework. However, it is worth mentioning that, in her view, grammar has three dimensions: form, meaning and use. *Form* deals with phonology, graphology, morphology and/or syntax; *meaning* falls under the realm of semantics. Finally, *use* ("when/why do we use it?", Larsen-Freeman, 2003, p.38), which is related to pragmatic and discourse levels, depends directly and exclusively on context and it is more clearly relatable to implicit knowledge as evidenced in performance.

Within this frame, it is clear that, since this study deals with test sections that focus on grammar, we leave out the measurement of grammatical knowledge as integrated into other skills. Thus, for the present purposes, the analysis of this section includes the design and the tasks. The focus on the design considers the relative importance of sections, their salience, and their sequence (based on Bachman and Palmer, 1996). The two variables proposed by Heaton (1990), importance and time, serve as a guide to establish direct connections to teaching practices. In turn, the analysis of the test tasks considers them as activities that involve "individuals in using the language for the purpose of achieving a particular goal or objective in a particular situation" (Bachman and Palmer, 1996, p. 44) and that yield information for teachers to infer students' language ability in the specific area they intend to measure. The precision of the information, as well as the quality of the inferences it allows us to make, depends on a myriad of factors, including the characteristics of the tasks, the alignment with the curriculum in general and the objectives of the course in particular, their validity and reliability, among many others. Analysing these factors is, of course, beyond the scope of this study. The analysis of tasks offered here follows a classification regarding the type of response (based mainly on Bachman and Palmer, 1996) and context (based, partly, on Celce-Murcia, 2005; Larsen-Freeman, 2003, 2009 and Purpura, 2004). By context we mean the input in which the items are immersed in each task. For us to consider that tasks are contextualized, items should be presented in passages that cover two or more sentences, or accompanied by images that

provide contextual communicative clues. The importance of context is developed in section 1.2.1 below.

### **The design of the grammar section**

In this article, the term “section” refers to a task or a series of tasks grouped according to the content, the component(s), or the skill(s) they are intended to assess. In the field of language testing, some other terms denote the same meaning, such as “part” (Bachman and Palmer, 1996) and “paper” (UCLES, 2018).

Following the framework of language task characteristics proposed by Bachman and Palmer (1996), the current analysis is based on some aspects of the characteristics of the test rubrics, namely: *relative importance*, *salience*, and *sequence* of the sections.

*Relative importance* deals with the weight tasks or sections are given in relation to others in the test. Typically, it focuses mainly on the length and the score assigned to each section and task. Due to data collection limitations, the assignment of scores is not considered in this paper. The relative importance here takes into account whether the sections are included in the tests. Since this inclusion is crucial in the link between teaching and testing, it is worth considering Heaton’s (1990) variables in his proposal of a simple method to design a balanced test. He recommends making lists of skills and components covered in class and the assignation of a percentage to each of them according to two variables:

- the time devoted to teaching and practicing this particular skill or component;
- the importance this specific skill or component has in this course or part of course.

These two variables, time and importance, are the basis for us to decide which sections to include in the test, their length, the scores assigned to each, among other aspects. For instance, if reading is important in our course and if it takes up, say, 30% of our lessons, it is only logical for us to include in our tests a reading section whose length and score reflect these two variables. In this way, our tests are supposed to mirror our teaching practices.

*Salience* refers to the way in which the different sections are visually discernible from one another. Basically, in this paper we consider the use of headings that indicate the beginning of a section. This feature could influence students’ perception of the test, since the salience of sections may lead them to “adopt differing response strategies, depending on a number of factors” (Bachman, 1990, p.120). The use of headings to introduce each section may help students understand the objective of each task in it.

The *sequence* of the sections deals with the order in which they are presented in the test. If the sections are not related to one another, two typical ways to sequence them may be: from easy to more demanding, or randomly. Whichever is chosen, “the sequence in which items are presented in a test reflects, to some degree, the test designer’s intention, and introduces an element of control on the test takers’ responses” (Bachman, 1990, p.121). However, we should bear in mind that students usually have the possibility of choosing what to do first and what to do next.

### **The tasks in the grammar section**

#### ***Tasks according to context***

Purpura (2004) emphasizes the importance of context in allowing the association of form with meaning as a pre-requisite to teach and test grammar. This idea of context implies a move beyond the sentence-level towards extended discourse. Following the same line, Celce-Murcia (2005) states that “most ESL/EFL teachers tend to view ‘grammar’ as an exclusively sentence-level phenomenon” (p.174). She goes on to explain that the majority of grammar rules operate beyond sentence boundaries. According to the author, there are only but a few rules that can be resolved within a sentence, like subject-verb agreement, determiner-noun agreement, use of gerunds after prepositions and reflexive pronominalization at clause level. The rest, from tense-aspect-mood choice up to clefts, depend on contexts larger than a sentence.

In turn, Bardovi-Harlig (2005) argues in favor of the use of texts when dealing with grammar, “the use of texts as input is not only methodologically desirable, but acquisitionally necessary” (p.186). According to the author, the use of text as context “offers the opportunity to discover form-

meaning-use associations” (p.199) that would not be always possible without a context. In the same vein, and in direct connection to the teaching of grammar, Purpura (2004) suggests that form-meaning connections “trigger acquisitional processes” (p.43).

Thus, bearing in mind the importance of context, the tasks in the present analysis will be classified according to the input in which the items are immersed. Three groups can be identified:

- contextualized tasks: items are presented within a text, a passage or with images providing a meaningful context;
- sentence-level tasks: items are presented within the boundaries of an isolated sentence;
- word-level tasks: items are presented as one isolated word.

#### ***Tasks according to type of response***

Bachman and Palmer (1996) propose an exhaustive framework for describing task characteristics. When dealing with the characteristics of the expected response, they distinguish three types: selected, limited production and extended production. A *selected response* is typically exemplified with multiple-choice tasks, in which students signal their option among different alternatives. *Limited production responses* involve one word or phrase up to one sentence or utterance. Finally, an *extended production response* consists of two or more sentences or utterances. For the purposes of the present study, we consider it necessary to include a category between limited and extended production to better categorize the data obtained: *moderate production*. Thus, the categories in this study include the following:

- *Selected response*: as described by Bachman and Palmer (1996).
- *Limited production*: one word or phrase.
- *Moderate production*: longer than a phrase, it could be a clause, up to one sentence or utterance.
- *Extended production*: as described by Bachman and Palmer (1996), although it must be noted that it was not found in the data obtained.

As will be seen in the analysis, this *moderate production* category serves the purpose of signaling out many different tasks analyzed.

## **Methodology**

### **Data collection**

The data are derived from a corpus of 60 written achievement tests (end-of-term and end-of-year tests) used in secondary schools in Tucumán. It was collected over a period of four years, from 2016 to 2019.

The courses selected belong to either state or private schools from the Gran San Miguel de Tucumán area. The courses fall within A1 and A2 levels, according to the CEFR. As for the hours of instruction, they have 3 or 4 weekly 40-minute lessons. No significant difference was found in the data obtained from these two types of schools, so no further distinction between them is made in the analysis and discussion of the results.

## **Results and discussion**

### **The inclusion of a grammar section**

The written tests in the corpus do not always present easily identifiable, clear-cut sections, as will be discussed in 3.2 below. However, an analysis of the tasks allows for a rudimentary systematization of the data. In this line, then, the tasks found in the tests can be grouped into five sections: grammar, vocabulary, reading, writing, and listening. Table 1 shows the sections and tests that include each of them expressed in terms of quantity and percentage.

Table 1. The number and percentage of tests that include each section.

Section	Number of tests including the section	Percentage of tests including the section
Grammar	60	100%
Reading	42	70%
Vocabulary	40	67%
Writing	22	37%
Listening	1	1%

Sections in Table 1 are presented in decreasing order according to the quantity of tests that include them. Out of the 60 tests analyzed, all of them include a grammar section (100%); 42 tests have a reading section (70%); 40 include a vocabulary section (67%); 22 tests have a writing section (37%) and finally, only 1 test includes a listening section (1%). These percentages show the prominence grammar is given in relation to the other components. Consequently, it becomes clear that research into the assessment of grammar is essential.

If Heaton's (1990) variables are applied to these data, it is evident that teachers tend to give much more importance and time to grammar in the classroom, since it is present in all the tests. Next, it seems that reading and vocabulary are given rather moderate importance and time, with 70% and 67%, respectively. In turn, writing does not draw much attention in class, judging from the percentage of tests including it, i.e., 37%. In relation to listening, it should be further studied, since this skill is almost absent from the data. It must be remembered that this study analyzes written tests and listening is generally reported to be tested in the speaking section of tests.

These results are similar to those in Abboud (2019), in relation to the percentage and implied importance of the sections. This may indicate that there seems to be a tendency to particular testing and, presumably teaching, habits. It is worth mentioning that in the present study, as well as in Abboud (2019), there is no significant difference between public and private schools. This may be explained by the fact that two important characteristics are shared by the schools selected in both studies: the students' level of proficiency and the hours of instruction.

#### **The relative location of the grammar section**

In the tests under scrutiny, the grammar section tends to be located towards the beginning of tests. This is shown in Table 2. The most frequent location of this section is 2<sup>nd</sup> place, in 43% of the tests in the corpus. The second most common location is 1<sup>st</sup> place, in 22% of the tests. Next, the third place is found in 17% and the fourth in 3%. The first two most frequent locations of the grammar section cover 65% of all the tests analyzed. In the majority of cases in which grammar comes in second place, the first section is devoted to reading.

Table 2. The number and percentage of tests according to the location of the grammar section.

Location of grammar section	Number of tests	Percentage of tests
1st place	13	22%
2nd place	26	43%
3rd place	10	17%
4th place	2	3%
Multiple places	9	15%
TOTAL	60	100%

Interestingly, in nine tests (15%), the grammar section is split in different locations. Such splitting is possible because these tests present some grammar task(s), followed by a task or tasks testing some other content, and then a grammar task is included again. These multiple locations include four tests in which the grammar section appears in first and third place; three tests where

it is found in second and fourth place; one test where this section is presented in third and fifth place, and finally, there is one test in which the grammar section is split in 3 places: first, third and fifth.

In relation to the other sections, 74% of the tests that include a reading section present it in first place, 19% in second and 7% in third position. Regarding the vocabulary section, out of the 40 tests that include it, 48% place it in second order, 23% in third and 20% in fourth position. As for the writing section, 96% of the tests have it as the last one, and 4% as the last-but-one. Finally, the only test that includes a listening section places it in first position.

The results of the location of sections show that there is a tendency towards a prototypical sequence in written tests in our context. As a general outline, it can be said that tests start with a reading section followed by a vocabulary and/or a grammar section in a random order, and, finally, the writing section if any. This prototypical sequence may be considered acceptable for a number of reasons, two of which will be discussed here. Firstly, there seems to be agreement in that tests should grow in complexity from beginning to end (Bachman, 1990). Following this line in relation to skills, comprehension preceding production looks like a coherent sequence. Regarding response formats of tasks, a sequence that goes from selection to extended production would flow naturally as well.

Secondly, if the rest of the sections are related to the reading material, it is necessary to place the reading section at the beginning in order to provide a certain degree of context for the rest of the tasks. Nevertheless, it should be pointed out that the majority of the tests analyzed do not present sections that are topically related to one another. A final consideration about the location of sections is related to the physical administration of the tests. Generally, we hand out the complete set of copies to students at the beginning of the session, which makes it possible for students to move backwards and forwards from section to section without following any pre-determined order.

### The relative salience of the grammar section

As introduced earlier, in this paper *salience* refers to the inclusion of a heading that indicates the beginning of a section, which contributes to the visual distinction between one section and another.

Table 3. The number of tests with headings for section and their relative frequency.

Section	Number of tests with headings	Relative frequency of headings
Writing	9	.41
Reading	13	.31
Vocabulary	7	.18
Grammar	9	.15
Listening	-	-
All the sections	9	.15

Table 3 shows the number of tests with headings for each section and the relative frequency this represents. The relative frequency is a value that expresses how often something happens within a certain number of events. Let us take writing and grammar as examples here. Nine tests include headings for writing and nine tests include headings for grammar. Given that they both have the same number of tests, why is it that writing gets a much higher relative frequency (.41) than grammar (.15)? This might be accounted for by virtue of the fact that the writing section appears only in 22 tests (Table 1), whereas grammar is present in all the tests under analysis (60). This implies that 9 times in 22 tests is not the same as 9 times in 60 tests. That is what relative frequency makes clear.

Taking into account bare numbers, the section with more headings is reading, in 13 tests. Next, grammar and writing have headings in 9 tests each. In turn, the vocabulary section has headings in 7 tests. The listening section is not marked by a heading. Finally, only 9 tests out of 60 have headings for all the sections they include. However, the value that is more relevant here is the relative frequency, which is the criterion for the ordering of sections in Table 3. From this value, it is easily observable that grammar has the lowest frequency: it is included in all the tests, but it just has a .15 value in relation to the inclusion of headings.

These results show some inconsistencies in the use of headings, especially when it comes to grammar, which may be detrimental to our students. Following Bachman (1990), if headings are included, they could serve as a guide for students to use the appropriate strategies for each skill or component. Connecting this aspect to our lessons, this type of salience is even more beneficial for students who are used to paying attention to the different language focuses during lessons. Take, for example, the headings that textbooks use to mark out the moments in lessons. They generally refer to the skills and/or components the tasks mainly focus on. It may be advisable to draw students' attention to the different moments and the varying focuses of each in the lessons. If students are aware of these shifts in class, the use of headings in tests would have positive results since this technique directs their attention to where it is necessary.

### **The tasks in the grammar section**

In this study, the tasks are grouped according to the context in which the items are immersed. This is shown in Table 4 below.

Table 4. Number and percentage of tasks according to the type of context.

<b>Type of context</b>	<b>Number of tasks</b>	<b>Percentage of tasks</b>
Sentence-level	100	62%
Contextualized	58	36%
Word-level	3	2%
TOTAL	161	100%

The majority of tasks, 62%, test students' grammatical knowledge at sentence-level. Contextualized tasks, in turn, represent 36%. Finally, word-level tasks are minimally represented, 2%.

Following, a more specific classification is applied to each group of tasks (contextualized, sentence-level and word-level) according to their type of response. Although the tasks are not analyzed quantitatively in terms of grammar dimensions (*from, meaning, use*), the descriptions of the examples provided include a brief reference to this theoretical frame.

### **Contextualized tasks**

Three types of response are found within contextualized tasks. Table 5 shows the number of tasks together with their relative frequency.

Table 5. Contextualized tasks according to the type of response.

<b>Type of response</b>	<b>Number of tasks</b>	<b>Relative frequency</b>
Moderate production	21	.36
Limited production	20	.35
Selection	17	.29
TOTAL	58	1

Tasks that demand limited and moderate production have approximately the same relative frequency, .35 and .36, respectively. Selection tasks, in turn, show .29 of frequency. This indicates that there is no significant difference between these types of responses.



Regarding selection tasks, the following variations are found: multiple choice (12 tasks, .71), matching with elements in box (3 tasks, .18), underlining (1 task, .06), and transcribing (1 task, .06). This is shown in Table 6

Table 6. Number and relative frequency of variations of contextualized selection tasks.

Variation	Number of tasks	Relative frequency
Multiple choice (MC)	12	.70
Matching	3	.18
Underlining	1	.06
Transcribing	1	.06
TOTAL	17	1

A typical sample of a contextualized selection task is the multiple choice (MC) in Example 1 below. The items are immersed in a dialogue; however, there is no reference to the topic (UFOs) in the instruction, which could have contributed to the contextualization. This task falls within the selection type because students have to choose between two options that are alternatives in terms of tense/aspect. In this case, the dimensions that appear to be tested are form and meaning, since students have to choose the tense/aspect based on the meaning implied in the dialogue.

5) Circle the correct option to complete the dialogue. The first one has been done for you.

A reporter is interviewing Mike and Harriet.

Reporter: Mike and Harriet, tell me what you saw *were seeing*.

Harriet: Well, when we *drove/were driving* home last night, we *were seeing/saw* a strange object in the sky.

Mike: While we *were coming/came* down the hill into town, it just suddenly *appeared/was appearing* in front of us. We *were stopping/stopped* the car and *got/were getting* out.

Harriet: It *was/was being* a very clear night. The stars *were twinkling/twinkled*.

Mike: It *was being/was* a spaceship. It *seemed/was seeming* quite big. It *was having/had* some strange writing on the side. And a light *flashed/was flashing* on the top.

Harriet: While we *watched/were watching* it, it suddenly *was flying/flew* away and *disappeared/was disappearing*.

Example 1. A contextualized selection task.

As for the contextualized limited production type, the 20 tasks found respond to the same variation: completing with the correct form of certain elements in brackets. These elements are generally verbs and, to a much lesser extent, adjectives or adverbs. Example 2 illustrates this type.

**1. Read and complete with the correct form of the verbs in present simple. (23)**

Hi! My name <sup>1</sup>..... (**be**) Jordan. I <sup>2</sup>..... (**be**) 13 years old and I <sup>3</sup>.....(**live**) with my family in a small house in Scotland.

I <sup>4</sup>..... (**wake**) up at 6 o'clock and I <sup>5</sup>..... (**take**) a shower in the morning every day. Then, I <sup>6</sup>..... (**have**) breakfast with my parents and my brother. They <sup>7</sup>.....(**take**) us to school by car. My Mum <sup>8</sup>..... (work) in a Clinic in Glasgow, and my father, George, <sup>9</sup>.....(**work**) there too. They <sup>10</sup>..... (**be**) both doctors.

I <sup>11</sup>..... (**return**) home on foot at 12.30 with my little brother. Albert <sup>12</sup>.....(**be**) 7, and we <sup>13</sup>..... (**talk**) on our way home. When we <sup>14</sup>.....(**arrive**), Mum <sup>15</sup>.....(prepare) lunch. After that, Albert <sup>16</sup>.....(sleep) for a while, but I <sup>17</sup>.. .....(**do**) my homework.

In the afternoon, Albert and I <sup>18</sup>.....(**watch**) TV or we <sup>19</sup>..... (**listen**) to music. At 9 o'clock, we <sup>20</sup>.....(**have**) dinner together. Albert <sup>21</sup>.....(**go**) to bed at 10 o'clock, but I <sup>22</sup>.....(**stay**) for a while chatting with my friends in the computer. At 11, I <sup>23</sup>.....(**go**) to bed.

Example 2. A contextualized limited production task.

In Example 2, the items are immersed in a text whose topic is not explicitly mentioned in the instruction, just as in Example 1 above. This task demands limited production responses since just one word is needed for each item. The only dimension that seems to be tested here is form, because students' job is to provide the correct form of the verb in brackets. Meaning is not assessed since students do not need to know the meaning of the structure to complete the task. Similarly, use is not tested either because there is no choice: the instruction explicitly establishes which tense/aspect should be used (present simple). It is worth noting that 14 out of the 23 items in this task require the unmarked form of the verbs (e.g., item 3 "I live"), which means that just the transcription of the verbs in brackets is enough to complete most of the task successfully. This implies that these items may lack validity, since it is impossible to be sure that students actually demonstrate the type of knowledge supposedly tested.

Finally, contextualized moderate production tasks can be found in four basic variations, as shown in Table 7. These variations are: writing/completing sentence in response to a situation/image using a specific structure (13 tasks, .62); asking/answering/completing questions/sentences based on image/table (3 tasks, .14); writing questions from prompts with images (2 tasks, .1); asking questions to complete dialogue (2 tasks, .1); and translation (1 tasks, .04).

Table 7. Number and relative frequency of variations of contextualized moderate production tasks.

Variation	Number of tasks	Relative frequency
Writing/Completing sentence in response to a situation/image	13	.62
Asking/Answering/Completing questions/sentences based on image/table	3	.14
Writing questions from prompts with images	2	.1
Asking questions to complete dialogue	2	.1
Translation	1	.04
TOTAL	21	1

Example 3 below presents a sample where students have to complete dialogues by asking questions using the prompts provided.

5-) WRITE THREE QUESTIONS FOR EACH SITUATION. USE PRESENT PERFECT SIMPLE AND PAST SIMPLE.	
a-) Ride a horse	_____ ?
Yes, I have.	
Where	_____ ?
In the countryside.	
Did	_____ ?
Yes, I did.	
b-) Meet a famous person.	_____ ?
Yes, I have.	
Who	_____ ?
Martin Scorsese, a film director.	
Where	_____ ?
In a restaurant.	
c-) Go to a rock concert.	_____ ?
Yes, I have.	
Which group	_____ ?
I saw Oasis.	
Where	_____ ?
In a football stadium.	

Example 3. A contextualized moderate production task.

We can consider that Example 3 has some degree of contextualization granted by the situation presented verbally on which students have to base their questions. This task demands what we call moderate production responses, because it is halfway between limited and extended production. In this case, students must write complete or almost complete questions for each item. We may consider that this task tests mainly knowledge of forms and, presumably, meanings: students have to construct interrogative forms based on the meanings of the situations and of the verbs in each tense/aspect. However, there are some linguistic cues that point to a specific choice of tense/aspect, like “Yes, I have” or “Yes, I did”. Besides, the instruction explicitly mentions the two possible forms to use, leaving students with a limited range of choice to show their grammatical knowledge.

**Sentence-level tasks**

There are 100 sentence-level tasks distributed in three types of response. Selection tasks are the first most frequent (.41), while moderate production tasks are second (.38). There is no significant difference between these two types. Finally, limited production tasks are the least frequent (.21). The number of tasks and the relative frequency of each type are presented in Table 8.

Table 8. Sentence-level tasks according to the type of response.

Type of response	Number of tasks	Relative frequency
Selection	41	.41
Moderate production	38	.38
Limited production	21	.21
TOTAL	100	1

As regards sentence-level selection tasks, six variations are found, which are shown in Table 9 below.

Table 9. Number and relative frequency of variations of sentence-level selection tasks.

Variation	Number of tasks	Relative frequency
MC of correct form	20	.49
Ordering words to form sentences/questions	10	.24
Completing with correct option from box/instruction	7	.17
Matching of correct form	2	.05
MC on explicit knowledge	2	.05
TOTAL	41	1

Multiple choice of correct forms (.49) is the most common variation in this type. Next, ordering words to construct sentences or questions (.24) and completing with correct option from a box or from the instruction (.17) are the second and third, respectively. Matching of correct form and multiple choice on explicit knowledge are fourth and fifth (.05 each).

Two variations need further explanation, ordering words and completing with correct option from box. These are selection types because they do not require modifications; otherwise, they would be production types. For example, if the ordering item does not provide all the elements or if some elements need to be altered (like in “make all the necessary changes”), then it would be a production task. Similarly, if the options given in the box were to be modified (generally at morphological level), then it would require production as well (see Example 5).

A sample of a selection task that tests explicit knowledge is Example 4.

5.	<b>Finished or Unfinished</b>	<b>score:10/10</b>
a.	He gave students rock lessons.	_____
b.	He has already taught hundreds of people.	_____
c.	He started the school in 1998.	_____
d.	My students have never played in front of a real audience before.	_____
e.	When I was younger I played the violin.	_____

Example 4. A sentence-level selection task.

The items in Example 4 are presented in isolated sentences. This selection type can be considered a multiple-choice task in which the options are displayed in the instruction. Regarding the dimensions, this task seems to test explicit knowledge of form and meaning, since students need to understand the difference between past simple and present perfect in terms of “finished” or “unfinished.”

As for sentence-level limited production, the 21 tasks present two variations of completion: with or without the provision of the base form of the words to be used, as shown in Table 10.

Table 10. Number and relative frequency of variations of sentence-level limited production tasks

Variation	Number of tasks	Relative frequency
Completion with base form provided	17	.81
Completion without base form provided	4	.19
TOTAL	21	1

The variation that includes the base form (generally added in brackets) is much more common (.81) than the variation without it (.19). Example 5 illustrates a task with the base forms in brackets.

4) Complete the sentences using the Present Continuous forms of the verb in brackets.
1- I can't come to the cinema. I <u>'m doing</u> my homework. (do)
2- You can't go outside. You <u>aren't wearing</u> any shoes. (not wear)
3- My parents aren't here at the moment. They _____ golf. (play)
4- She can't answer the phone because she _____ her hair. (wash)
5- We don't want to leave the party. We _____ to a very interesting person. (talk)
6- My children _____ English at the moment. (learn)
7- They _____ this afternoon because they're both very tired. (not work)
8- You can't listen to music because the stereo _____ (not work)
9- I _____ out this evening because I've got a headache. (not go)
10- What _____ the children _____ in the garden? (do)
11- What's the matter? Why _____ you _____? (cry)
12- the sky is cloudy. Why _____ you _____ sunglasses? (wear)

Example 5. A sentence-level limited production task.

In Example 5, the items are immersed in isolated sentences, so there is no further context (a specific situation or a topic, for example) beyond sentence boundaries. Each item requires a limited production response, just the verb phrase. As for the dimensions, since the verbs are given in brackets in the place where they are to be used, and since the tense/aspect is established explicitly in the instruction, neither meaning nor use seem to be tested. Students have to provide the correct form of the present continuous of the verbs given, so form is the focus of this task.

Regarding sentence-level moderate production, seven variations can be observed, as Table 11 shows. Writing/completing sentences with certain structures is by far the most common (.55). Rewriting (keeping the same meaning) and changing to interrogative/negative structures are both second most frequent variations (.11 each). Making questions for isolated sentences are in third position (.08). Finally, the fourth position is shared by three variations: correcting grammar mistakes, answering personal questions, and answering questions based on symbols given (.05 each of them).

Table 11. Number and relative frequency of variations of sentence-level moderate production tasks.

Variation	Number of tasks	Relative frequency
Writing/completing sentences with certain structure	21	.55
Rewriting (keeping meaning)	4	.11
Changing to interrogative/negative structure	4	.11
Making questions for isolated sentences	3	.08
Correcting grammar mistakes	2	.05
Answering personal questions	2	.05
Answering based on symbol given	2	.05
TOTAL	38	1

A typical task that is absolutely meaningless, or rather contradictory, is illustrated in Example 6. In this task, changing affirmative sentences into negative ones has no other purpose than to negate the verbs. The result is not only meaningless, but also incoherent.

<b>A- Turn the following sentences into negative</b>	
1 I was at the recital last night.	_____
2 He was at the supermarket at 10 o'clock.	_____
3 We were absent last class	_____
4 They were in England last year.	_____
5 It was sunny yesterday	_____

Example 6. A sentence-level moderate production task.

Example 6 shows items that are presented at sentence-level and that demand moderate production, since students have to write a sentence. Concerning the dimensions tested, neither meaning nor use appears to be involved, so the focus may be on form because students must reproduce the sentences providing the negative versions of the verbs in each item. With a relative frequency of .11, tasks like this, which require incoherent production, should be a call for reflection on the teaching and testing practices that focus on forms regardless of meaning.

#### ***Word-level tasks***

The tests analyzed present 3 word-level tasks, all of which belong to the limited production type. They typically ask students to complete charts with the correct versions of base forms. Example 7 shows an instance in which students have to modify the verbs morphologically to add the *-ing* suffix.

1)- Write the -ing form of the verbs. (1.5p)	
1 write	writing
2 stop	_____
3 travel	_____
4 go	_____
5 dance	_____
6 read	_____
7 sit	_____

Example 7. A word-level limited production task.

Example 7 shows an extreme version of lack of context, since the items are the words that are to be modified. Only one word has to be produced for each, which is why it is considered limited production. Finally, since no student needs to understand the meaning of the *-ing* ending

in each case and since there is no choice regarding the use of it, this task may be said to test only knowledge of the form of the verbs with *-ing* suffix.

### Conclusions

The results obtained in this study on tests shed light on the treatment of grammar in A1 – A2 courses from secondary schools in Tucumán. We can now distinguish certain tendencies in the design of the grammar sections and tasks of our tests and can establish certain connections to our lessons. Using reverse engineering and applying Heaton’s (1990) variables, we may interpret the results of the inclusion of the grammar section, displayed in Figure 1, as evidence of the central importance grammar is given in our classrooms and the remarkable amount of time devoted to teaching and practicing this type of content in relation to the others.

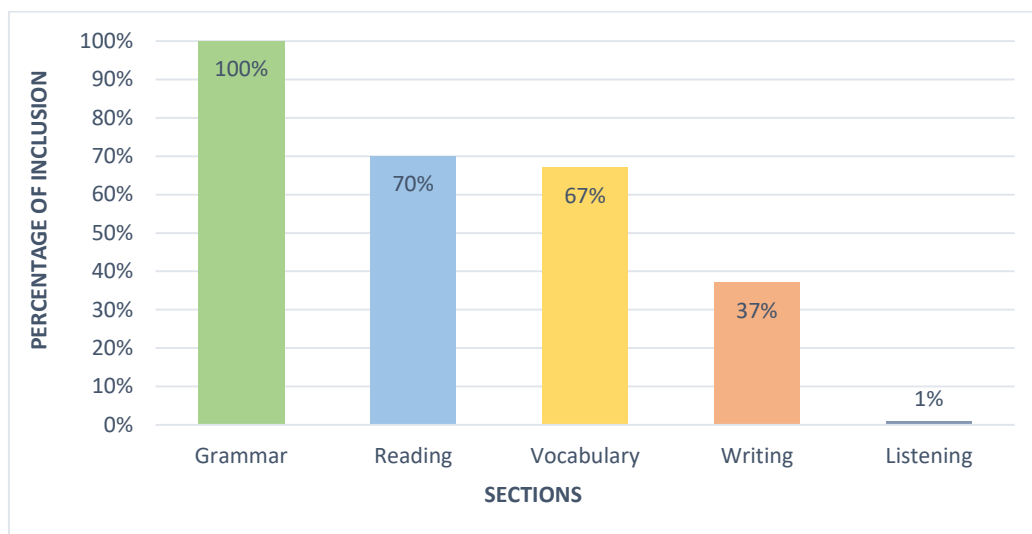


Figure 1. Inclusion of sections in tests.

In turn, the prototypical sequence of sections may indicate that we need to revise the development of topics across various lesson moments and test sections. Finally, the inconsistency in the use of section headings reminds us of the potential benefits of directing students’ attention to the different moments of lessons as well as the different sections of tests.

In relation to the context in tasks, there is a clear tendency towards isolated items, with most of them presented at sentence-level. This is shown in Figure 2.

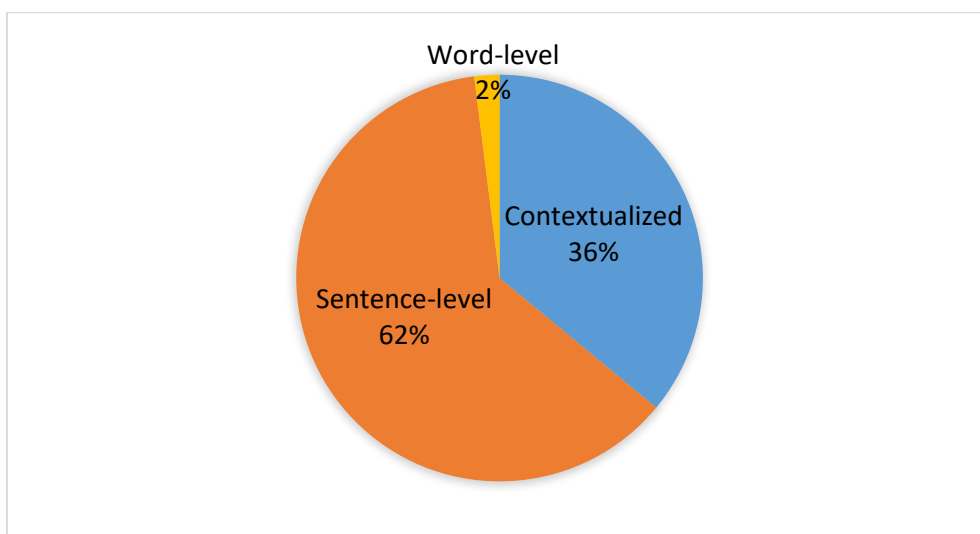


Figure 2. Tasks according to context.

Within contextualized tasks, there has to be further analysis in order to measure whether the degree of contextual clues is enough for students to make connections between the forms and the discursive needs that arise from the situations of use presented (Villicco, 2013). Clear examples of this problem are tasks where the form to be used does not depend on meaning but is established in the instruction, such as “complete with the present simple.” It is worth noticing that instructions like this rely heavily on students’ metalinguistic knowledge. Besides, this design of tasks may indicate that the context provided is not enough for students to know which tense/aspect to use, so the instruction has to provide a clue. This clue, it is argued here, should be better found in the input itself, as contextual clues or in the meaning of the passage, rather than in the instruction. Thus, instead of stating the tense/aspect in the instruction (testing just form), we can leave it for students to decide which tense/aspect is accurate and appropriate according to the meaning and contextual clues of the passage (testing form, meaning and use). It becomes clearer, then, that by modifying the design of the tasks, we can test not only form but also meaning and/or use. This use of context should also be applied to our teaching practices, thus facilitating students’ associations of form, meaning and use.

Regarding the types of responses, the tasks found in the grammar sections do not show a recognizable pattern, since they require selected, limited, or moderate production randomly. However, what they do show is a tendency towards one dimension over the others: the majority of the tasks elicit explicit knowledge of forms, and, to a much lesser extent, of meaning and use.

Our objective here is not to argue in favor of one type of response and against the others: all types are necessary. In this sense, it may be advisable to follow a simple-to-complex sequence: selection and limited production tasks are necessary at initial stages of instruction, when presenting new grammar points, and then there must be a gradual move towards extended production, in which form, meaning and use are smoothly combined. This may be particularly true in low-proficiency courses, such as A1 and A2 levels.

Similarly, and in relation to the dimensions, we could safely say, based on what is known so far, that it could be convenient to design tasks that focus on form (and perhaps meaning) at the beginning of the instruction. Then tasks can progressively move towards the elicitation of the association form-meaning-use (this sequence is formulated on the basis of the concept of learning challenges, proposed by Larsen-Freeman, 2003, 2009). All this can only be achieved, as has been seen, with contextualized tasks, i.e., items presented in extended passages or text

Although it has not been analyzed in this study, the measurement of grammatical knowledge as a component of language skills should not be left out in our classrooms, since it seems to be highly informative. This linguistic performance, as in reading, listening, speaking or writing, plays a crucial role in providing evidence of implicit grammatical knowledge (or lack of it). In fact, modern approaches to measuring grammar test it together with the productive skills (Spinner, 2021), and to a lesser extent with listening and reading. There are some drawbacks in the testing of implicit knowledge of grammar, though. As Larsen-Freeman (2009) puts it “[a] consequence of such decisions, however, is that it is difficult to separate out what in the ability to read or write the texts is due to the lack of knowledge concerning grammatical structures and what might be due to other factors” (p.533). Therefore, relying solely on an integrative approach to measuring implicit knowledge of grammar may not be the best idea in classroom situations, where tests should inform pedagogical decisions. In this line, Purpura (2004) acknowledges that “by eliminating the explicit assessment of grammar, we have no way of providing formative feedback to students” (p.254). All things considered, it is clear that we should use tasks with all types of responses, which focus on all three dimensions in turn, and combined when possible.

The question here, then, is not whether to include explicit-knowledge tasks with selected or limited production; the question is *when* to include them. As Larsen-Freeman (2009) clearly states, “discrete-point and integrative tests represent different approaches to grammar assessment, each of which have a contribution to make” (p.533). Again, varying the tasks to cover all types and all



dimensions over time should constitute valid techniques to measure explicit and implicit grammatical knowledge.

The findings of this study point to particular ways in which we can modify the teaching and testing of grammar, such as presenting contextualized grammar structures and drawing our students' attention to the shifts of language focus during lessons (which should be mirrored by the use of section headings in our tests). Finally, it should be stated that further research on the sequencing of types of tasks and the dimensions tested by each according to the students' level of proficiency may provide revealing insights into the teaching and testing of grammar.

### References

- Abboud, J. (2021). La sección de lectura en exámenes de Lengua Extranjera Inglés en secundarias de Tucumán. In N. Vera, V. D'Andrea and M. Rojas (Comp.), *Traducción y Terminología: puentes interdisciplinarios del conocimiento* (pp. 115-124). Humanitas. Facultad de Filosofía y Letras, UNT.
- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.
- Bachman, L. and Palmer, D. (1996). *Language Testing in Practice*. Oxford University Press.
- Bachman, L. and Damböck, B. (2018). *Language Assessment for Classroom Teachers*. Oxford University Press.
- Bardovi-Harlig, K. (2005). Tense and Aspect in Context. In T. Miller (Ed.), *Functional Approaches to Written Text: Classroom Applications* (pp.186-200). Office of English Language Programs.
- Celce-Murcia, M. (2005). Describing and Teaching English Grammar with Reference to Written Discourse. In T. Miller (Ed.), *Functional Approaches to Written Text: Classroom Applications* (pp.174-185). Office of English Language Programs.
- Ellis, R. (2009). Implicit and Explicit Learning, Knowledge and Instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp and H. Reinders, *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 3-26). Multilingual Matters.
- Heaton, J. B. (1990). *Classroom Testing*. Longman Inc.
- Larsen-Freeman, D. (2003). *Teaching Language. From Grammar to Gramming*. Thomson-Heinle.
- Larsen-Freeman, D. (2009). Teaching and Testing Grammar. In M. Long and C. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 518-541). Blackwell Publishing Ltd.
- López Barrios, M. and Villanueva de Debat, E. (2011). From Grammar Translation to CLT: A retrospective view of coursebooks produced in Argentina. In D. Fernández, A. Armendáriz, R. Lothringer, L. Pico and L. Anglada (Eds.), *Communicative Language Teaching and Learning Revisited. XXXVI FAAPI Conference Proceedings, 22-23-24 September 2011* (pp. 21-33). FAAPI.
- Purpura, J. E. (2004). *Assessing Grammar*. Cambridge University Press.
- Spinner, P. (2021). Measuring grammar. In P. Winke and T. Brunfaut (Eds.), *The Routledge Handbook of Second Language Acquisition and Language Testing* (pp. 233-242). Routledge.
- University of Cambridge Local Examinations Syndicate (UCLES) (2018). *Cambridge English Entry Level Certificate in ESOL International (Entry 2)*. Cambridge University Press and Assessment. <https://www.cambridgeenglish.org/exams-and-tests/qualifications/>
- Villecco, M. E. (2013). La reflexión metalingüística, metadiscursiva y metacognitiva al servicio de la adquisición de competencias discursivas en Lengua Extranjera. In M. E. Villecco (Ed.), *Adquisición y didáctica de lenguas materna y extranjeras: desafíos de adquisición* (pp. 16-21). Editorial Pimiento.