Transcript: SOW-based grammar testing for NCELP Y7–Principles, Design, Creation

Length of Talk: 29 minutes

Presenter: Rowena Kasprowicz

Slide 1 (0.19):

Hello, my name is Rowena Kasprowicz and I am one of NCELP's Research Specialists. In this presentation, I am going to talk through the principles and processes that were followed to create the grammar section of NCELP's Year 7 Spring term language test.

Slide 2 (0.37):

In this presentation, I will:

1. Introduce the team who have been working on the grammar section of the test

2. Talk through the principles that underpin the decisions we made regarding the shape and content of the grammar test

3. Outline the coverage of the grammar section of the test for each of the languages

4. Look at some example questions from the online test and explain how the test has been created

5. In the final section, I'll provide some more detail about how the scoring has been set up for the grammar test items, particularly for the writing section.

Slide 3 (0.40):

Creation of the grammar test was very much a team effort, so my thanks go to everyone who provided input and help in the creation process. In particular, I want to thank Stephen and Nick for the many hours they spent discussing the nuances of how to test particular grammatical features with

me. As well as the proof-readers and native speakers who recorded audio for us. And lastly to Giulia, who really gave lots of invaluable help with setting up the online version of the test and answering our many questions about the way in which the online software worked.

Slide 4 (2.39):

In terms of the principles underpinning the grammar section of the test, our aim was very much to test both learners' receptive and productive knowledge of the grammar that they had learnt during the first half of Year 7, up to February half term. We wanted to know, a) can they correctly recognise and interpret the grammar features when listening and reading and b) can they accurately produce the grammatical structures when speaking and writing.

There were a number of factors which we had to take into consideration when building the grammar tests.

Firstly, we wanted to ensure that the grammar test items were really isolating the students' grammatical knowledge. So, for example, we needed to design questions that were not reliant on students' being able to correctly recall a particular lexical item. We really wanted to be confident that if a student was unable to answer a question correctly, that was because they didn't have the necessary grammatical knowledge, rather than being due to lack of vocabulary knowledge for example. When we look at some example questions later in the presentation, I will explain how we did this for particular bits of grammar.

Secondly, we wanted to ensure that the test provided coverage of all of the grammar features that had been taught so far in the Scheme of Work (SoW), including both morphology (such as verb inflections) and syntax (such as word order for question formation). Therefore, we had to carefully balance how many items students' would answer for a particular grammar feature against the amount of time that was available overall for the grammar section. Of course, there is inevitable variation between the forms of the test for the three languages and this very much reflects the variation in which grammar features have been taught so far in French, German, and Spanish respectively.

Our third consideration was that we wanted to test students' receptive knowledge, i.e. their understanding of the meaning of the grammar feature, as well as their ability to recognise a particular grammatical form. Secondly, we wanted to also test whether students could accurately recall the grammar.

Finally, we wanted to make sure that we had tested students' knowledge in both the written and oral modalities and when we look at some of the example items later you will see how these different considerations came to light within the design of particular questions.

Slide 5 (2.31):

The table here shows the coverage of grammar features in the German version of the test.

By coverage here, this refers to what the pupils are actually being tested on. In the columns labelled Reading, Listening, Writing, and Speaking, I have indicated the number of items each pupil responds to for that grammar feature so you can see how the different grammatical forms were spread across the different modes and modalities.

The first step in building the grammar test was to identify which grammar features and rules had been taught so far and identify the specific structures that had been introduced for each. For example, in terms of subject-verb agreement for weak verbs in German, the 1st, 2nd, and 3rd, person singular forms have been introduced so far and so the items targeting that particular grammatical structure included instances of all of those particular forms that had been introduced so far.

Ideally, we would have tested everything in all modes and modalities, but that was simply not possible within the time constraint that we had.

Therefore, wherever possible, and as you'll see from the table, we have tried to ensure that each structure has been tested receptively and productively either in written or spoken form.

Within the productive items, particularly in the Speaking part of the test, we were able to combine testing of multiple structures within one question, for example by asking students to produce a short phrase or sentence which would contain a number of different grammatical features. For example, in

the Speaking part of the test, we have a set of items in which students had to create statements and questions with weak verbs, therefore were testing both word order and verb agreement in that set of items.

In the German test specifically we made the decision not to test negation (i.e. kein + noun) or the possessive adjective (mein +noun) in this round of testing and that was because the 2nd and 3rd person possessive adjectives will be introduced later in the spring / summer terms. Therefore, we felt it was more efficient to test all of these structures together in one go at the end of Year 7.

Slide 6 (1.29):

The table here shows the coverage of grammar features in the French version of the grammar test. Again, you can see the number of items included for each grammar structure in the Reading, Listening, Writing and Speaking columns. The first two structures in the table were only tested productively, because here we wanted to know that students recognise that the two present tense forms in English (the simple and present continuous) map onto only one form in French. And similarly, we wanted to check that students recognise that French does not have an equivalent of the English 'do-auxiliary' for questions. Therefore, for these two particular structures it seemed to make most sense to provide students with an appropriate English sentence and a test that they knew how to accurately form it in French.

In addition, French only includes one set of items in the Listening part of the test. That is because many of the structures (for example most of the verb endings) are not audible in French, therefore it made most sense to focus on testing these in Reading, rather than Listening parts of the test.

We've tried to make best use of the time available for the grammar section of the test, in this particular case focussing more on the written rather than the oral modality when we are testing receptive knowledge in French.

Slide 7 (0.26):

Similarly, to the previous two slides, here this slide shows the coverage of grammar features in the Spanish version of the test. You can see the particular structures being tested in the left-hand column and how many items were included for each structure in the four columns on the right.

I'll give you a moment to have a look through the table.

Slide 8 (2.43):

In terms of creating the test, we have designed a specific type of question for each grammatical feature that we were including in the test for each language and in the later parts of the presentation we will have a look at some examples of the questions that were included.

In terms of time limit, we had 15 minutes allocated for completing the Reading, Listening, and Writing items, and 4 minutes for the Speaking part.

Pupils will respond to 50 items across the Reading, Listening and Writing sections, and 13 items in total in the Speaking part of the test.

Inevitably there is some variation between the three languages in terms of the weighting of the different modes and modalities, and this is due to the specific grammar features that were tested in each of the languages. As we saw earlier in French, for example, the Listening part of the test is much smaller than it is in German and in Spanish.

In terms of creating the question item pool, for each question type we had a pool of possible items that the pupils could respond to. In the online version of the test, we asked the survey software to select a subset of these items to show to each student (based on the number of items specified for that particular grammar feature, as listed in the coverage tables for each language, which we looked at on the previous slides).

The question items were created using vocabulary from the SoW. We also reviewed all of the items against the vocabulary test to ensure that there was no clashes.

Each pool of items for each grammar structure contains an equal number of instances of each structure that had been taught. For example, the pool of items for the questions targeting subject-verb agreement contained an equal number of items for each of 1st, 2nd, and 3rd person singular forms in German. This means that each student is equally likely to be tested on 1st person singular as they are on 2nd or 3rd person singular, etc.

Over the next few slides, I will show you some examples of the different question types. I've picked a selection of examples from across the languages. Where I have provided an example for a grammar feature in one

language, and that feature has also been tested in another language, then the same principles were applied.

Slide 9 (2.03):

First, we will have a look at some examples of items that were found in the Reading and Listening parts of the test, so these are testing the students' receptive knowledge. Pretty much all of the questions in these sections are multiple choice.

With these items we really wanted to ensure that we were testing students understanding of the meaning of the grammar feature as well as their recognition of the form. So, in the example that you can see on the slide here, which is testing knowledge of subject-verb agreement, rather than giving the answer options in the target language, we have deliberately given them in English. This is therefore really testing that the pupils understand, in this case for example that the –ons verb ending refers to the first-person plural 'we'. If we had given the answer options in French, then the student may recognise the form, for example recognise that nous always appears with the –ons ending, but we can't be 100% sure that the student definitely knows and understands what the –ons ending means.

It's useful to note here that students' knowledge of pronouns is tested in the writing and speaking parts of the grammar test for French and German, were the students have to recall a phrase or a short sentence. Pronoun knowledge are also tested as part of the vocabulary test.

For the instances in French and German, where the verb ending is the same for two 'persons' (for example here 1st and 3rd person singular regular verb endings are the same in French), these options were combined as one option (e.g. I OR he/she above). So, it is important to acknowledge here that with these questions we are not explicitly testing that pupils know that in this case that the 1st / 3rd person take the same ending.

Slide 10 (0.44):

The next example, here, is an example of a question item targeting knowledge of grammatical gender.

Our aim with this question was to try to isolate students' grammatical knowledge of gender and number agreement. Here, the student has to

recognise the gender and/or number of the underlined word, in order to match it to the alternative form which has the same function. So here, successfully completing this item is not reliant on pupils recalling the gender of a particular lexical items. Rather it is trying to test whether they recognise the gender and number as it is marked on the article.

Slide 11 (0.35):

In this third example, we have a receptive item which is testing learners' knowledge of syntax, in this case word order in questions in German.

Here by removing the punctuation, we are testing whether pupils recognise the function indicated by the order of the subject pronoun and the verb in each sentence. In addition, we have used a variety of subjects for both the statements and questions, in order to avoid a reliance of 'du' (you) to indicate questions.

Slide 12 (0.44):

This next item is an example of how we tested students' productive knowledge of syntax. This question focusses on the order of words in phrases which include an article, adjective and noun in Spanish, and we had the equivalent in the French test too. So here students are given the three words and have to drag them into the correct position under columns one, two and three. By giving the pupils the three words, we are isolating students' ability to accurately produce the syntax itself, rather than also requiring them to recall the correct lexical forms.

Slide 13 (1.06):

Here we have examples from Spanish of written productive items.

In general, written productive items required pupils to supply just one element, rather than producing a whole sentence – this was really to make the scoring of the written items much more manageable.

The first example shows an item which is targeting students' awareness of the fact that Spanish does not require an alternative to the English 'do-auxiliary'.

You'll also notice that we have provided a gloss of the verb infinitive for students. This was done in order to ensure that we were isolating grammatical knowledge of these items. By providing the verb infinitive, we are not relying

on pupils being able to recall the appropriate lexical item in order to complete the question successfully. They have the correct word in the target language, so the question is really testing whether they can produce the appropriate verb inflection.

Slide 14 (0.53):

Here we have another couple of examples of written production items.

In general, in the written items we have tried to isolate one particular grammatical feature which pupils have to supply.

These also illustrate another way in which we have tried to isolate grammatical knowledge, in this case knowledge of gender. Here we have indicated the gender of the noun either directly after the noun itself or via the article. This is to ensure that completion of these items was not reliant on pupils having to recall the gender of specific nouns, rather we wanted to know whether students were able to accurately apply the grammatical gender system in the target language and so therefore we explicitly indicated the gender of the nouns that we had used within these items.

Slide 15 (0.53):

This next item is taken from the German version of the test and illustrates a question that is targeting both syntax, in terms of subject-verb word for questions, as well as morphology, in terms of weak verb inflections.

It is worth noting that these written items that are combining testing multiple grammatical features did have implications for the automated scoring for the online version, and I'll explain in a bit more detail in the next section how we went about scoring these sorts of items.

For future iterations of the tests, we will most likely try to separate out testing production of syntax (e.g. drag and drop) and production of verb agreement, in order to simplify the automated scoring process.

Slide 16 (0.45):

We are now going to look at a couple of items from the speaking part of the test. As mentioned earlier, in the speaking part of the test we have combined testing several grammar features within each item, and we did this by asking

students to produce short phrase or a short sentence for each item within the speaking test.

As with the writing items, we have provided a gloss of the key target language words for example giving the verb in items testing in subject-verb agreement, and this was really in order to avoid missing answers being due to student's struggling to recall the correct vocabulary that they needed for that particular item.

Slide 17 (1.04):

Here we have two more examples from the speaking section which are testing irregular verb agreement, & article / noun agreement.

For irregular verbs, we have not provided a gloss of the verb infinitive. This is because in the SoW, irregular verb forms have been taught as individual lexical items, rather than as a transformation from the infinitive. So, we felt that providing the irregular verb infinitive in these items could actually cause confusion rather than being helpful for the students.

For the items testing article agreement, we provide a gloss of the target noun and the target adjective where appropriate, and we have also indicated the gender, so that we are testing whether pupils can accurately produce the correct article and adjective ending, rather than relying on them being able to recall that particular gender for that noun.

Slide 18 (0.30):

We have seen some examples of questions from the online test so now in this final section, I will talk through how we have designed the scoring rubrics for the different sections of the test that the students completed.

In terms of the Grammar test, for the Reading and Listening items this was fairly simple. It was simply a process of indicating which of the multiple-choice options were correct and which were incorrect.

Slide 19 (3.00):

However, for the written productive items a much more nuanced scoring rubric was needed, as here we are scoring whatever the student has typed into the text box.

So, for the online test, possible answers have been manually coded and a mark scheme with full, partial, and no marks has been worked out for each question type. Over the next few slides I will show you some examples of what the partial marking looks like and how this has been applied.

The general principles we applied when developing the scoring rubric are as follows:

1 mark was awarded for each structure that was tested (e.g. for accurate production of subject-verb agreement students would receive one mark per item answered).

Within this, partial marks accounted for the different elements of the structure being tested. For example, in items testing article agreement, partial marks account for whether the student has used the correct gender, correct number, correct case and whether they have correctly produced a definite or indefinite article.

Similarly, for subject-verb agreement, partial marks account for accurate production of the pronoun and the verb inflection.

We were able to deal with the absence of accents and diacritics systematically. We have tolerated missing accents, where these do not alter the meaning of the word and in these cases have not deducted any marks. However, for words where absence of the accent does alter the meaning of the word, we have deducted 0.5 mark. We will see some examples of how this was applied in the later slides.

Ideally, in the Grammar test because the focus here is on testing grammar knowledge, and not on spelling, we would ideally tolerate spelling errors outside of the target structure (e.g. in verb stem as long as verb ending is correct). However, it was not possible to account for other types of spelling errors (aside from missing accents) – as each possible response from a pupil has to be manually coded (see next few slides!). It's not possible to predict all possible spelling mistakes that any one pupil might make. However, it is useful to remember that we did provide students with a gloss of target words within the question, so hopefully this will have minimized the incidence of spelling errors within the words we were asking pupils to produce.

This first round of testing is of course a very much pilot phase, and so once we start receiving responses from pupils we will be able to begin identifying

common errors that are occurring. These could then be built into future scoring rubrics.

Slide 20 (1.10):

The next few slides we will look at some examples of how the partial marking was implemented in different question types.

The examples on the screen here are from items targeting article agreement. For these questions, we coded firstly the option that was fully correct–the top line that you can see for each question there–, and we then indicated the other possible answers which would receive 0.5 mark.

In these sentences, students get 0.5 mark for giving an article with the correct gender. They would lose 0.5 mark for 'another element' (e.g. case / definiteness) being incorrect. For example, in the top example, giving the feminine indefinite article would receive 0.5 mark, because the question asked for the definite article.

In the bottom example, we can see that for masculine articles in German, there are several options that would receive 0.5 mark, this was due to the fact that we also had to account for use of the correct case.

Slide 21 (0.28):

These two examples that we can see on the slide demonstrate the partial marking applied for missing accents, in this case that change the meaning of the target word, the top example taken from the French test and the bottom from the Spanish test. We can see that in both instances when the student has forgotten to include the accent they would receive only half a mark.

Slide 22 (0.49):

In both French and German, the partial marking system was more complicated for the items which targeted subject-verb agreement. This was because the students had to supply two elements, the pronoun and the verb form. Therefore, the scoring had to account for potential inaccuracies in one or both of these elements.

As you can see from the different options that we have coded, students received 1 mark for providing the correct pronoun and the correct verb form. They would receive 0.5 mark for providing the correct pronoun but the

incorrect verb inflection, and vice versa, 0.5 mark if they provided the correct verb inflection but the incorrect pronoun.

Slide 23 (1.59):

This final slide illustrates how this marking was implemented in German test specifically. There was an added layer of complication for these items in German, because we were also testing word order for statements versus questions.

This meant that in the scoring we needed to account for potential inaccuracies in either the word order or the subject verb agreement elements or potentially in both.

Because these items were testing two grammar features (word order and subject-verb agreement), students could score up to 2 marks on each item.

In the first row of the first column, you can see that students would receive 2 marks if both the word order and the pronoun and verb form were correct. 0.5 mark would be deducted if either the verb form or the pronoun were incorrect.

In the final row of the first column, you can see that 1 mark would be deducted if students had used the incorrect word order (but the pronoun and verb inflection were correct).

In the second column, you can see that students would receive 1 mark if they had used the correct word order, but given an incorrect pronoun and incorrect verb form. This recognises and acknowledges that even if pupils have given incorrect pronoun / verb ending, they have correctly positioned the pronoun and the verb (and also that they are correctly recognising that the 'do +aux' and present continuous in English both just need a pronoun and a verb (only) in German.

The final 6 options in the second column then reflect when 0.5 marks would be given, i.e. if the word order was incorrect and either the pronoun or verb form was also incorrect.

Slide 24 (0.41):

Just to summarise, in this presentation I have talked through:

1. the principles that underpin the decisions we made regarding the shape and content of the grammar test

2. the coverage to the grammar section of the test for each of the languages

3. looked at some example questions from the online test and explained how the test has been created

4. explored how the scoring has been set up for the grammar test items, particularly for the writing section

I hope that this presentation has given a useful overview of the design of the grammar test. Thank you very much for listening.