Transcript: NCELP Phonics Assessments–Principles, design, creation

Length of Talk: 37 minutes

Presenter: Dr. Robert Woore

# Slide 1 (0.36):

Hello everybody it’s Robert Woore again here, in this presentation I am going to give some information about the NCELP phonics assessments that we have created to accompany the Scheme of Work. I am going to run through some of the background to the phonics assessments, the principles underpinning them, and talk about how they are intended to work. I hope you find it useful. There is a voiceover which should start automatically with some of the slides and when you are ready to move on just proceed to the next slide. I hope you enjoy it.

# Slide 2 (0.18):

Creating the assessment in phonics, vocabulary, and grammar was a huge team effort, on this slide you can see some of the people who worked on the phonics assessment. We want to show this slide to say a huge thank you to all of these people.

# Slide 3 (0.26):

This is what we will cover in the presentation and it won’t be very long, we will go quite quickly but we will talk a little bit about the principles underpinning the phonics assessment, so what we are trying to test and how we might go about doing that. Secondly, how we designed the tasks and questions that we included in the test, and thirdly, just a word about scoring, what we mean by accuracy and how the procedures work.

# Slide 4 (1.29):

The phonics assessment aims to test pupil’s knowledge of the Symbol-Sound Correspondences (or SSC hereafter) which they covered to date in Year 7. In other words, it is a syllabus-based achievement test. It is trying to find out to what extent they have actually learnt what we are trying to teach them in phonics up to the point of testing.

There are various other presentations and handouts dealing with the rationale for teaching phonics so we won’t go into that in great depth here but we believe SSC knowledge is important, both when going from print to sound so for example when we are reading aloud, it might not actually be out loud but with one’s inner voice. This has been linked to various aspects of language learning such as reading comprehension and vocabulary acquisition. It is also important when going in the other direction, so for example from sound to print, so for example when spelling words when you are writing, also when listening to help you segment the stream of speech that is flowing past you as you listen. To reflect these two directions, print to sound and sound to print, our test includes both a transcription tasks—so to what extent can people hear the language and write it down?—and also a reading aloud test looking at the extent to which people can convert the written language into appropriate phonological forms.

# Slide 5 (1.01):

So, one of the key theoretical underpinnings of various aspects of the NCELP pedagogy is skill acquisition theory, again we won’t go into this in great detail here because it is covered in other presentations and resources, if you would like to refresh your memory on it there is some information on it, for example, in the [grammar presentation](https://resources.ncelp.org/concern/resources/z029p475r?locale=en) from Residential 1 which is available through the Resource Portal. Essentially, learners are moving from on the left declarative knowledge, knowing about a language feature, then onto knowledge about how to use a language feature (procedural knowledge), and finally automatic fast accurate use of the language feature in the final box on the right. They are moving through these stages through expensive repeated meaningful practice. Obviously, the ultimate goal in terms of the phonics strand is for our pupils to be fluent and accurate decoders.

# Slide 6 (1.17):

A fluent decoder in German, for example, would see this word on the top on the left and would immediately recognise it as ‘Wagen’ and pronounce it as ‘wagen’. Fluency and accuracy are both important in foreign language learning, accuracy is obvious­–the spoken and written forms that pupils learn and produce need to be accurate enough to be recognised to allow communication to take place and fluency is important, because for one thing it reduces the burden on the learner’s working memory and then thereby allows other high level process to be carried out more easily such as when reading making inferences, understanding the overall message of the piece, conceptualizing ideas when you are writing and thinking about what you are going to say next, formulating language. We need fluency to allow us not to get bogged down by focusing all our attention on how to pronounce or how to spell a particular word.

However, it is no good having fluency without accuracy—there is no point automatizing incorrect knowledge. At this stage, accuracy is the most important thing.

# Slide 7 (2.50):

If we just pause here for a second to think a little bit more about the relationship between fluency and accuracy. For a learner who is more or less literate or more or less fluent in English as a first language or as a main language when encountering another language that uses the same alphabet as do French, German and Spanish there is the possibility of interference from English SSC correspondences. For example, when seeing this German word they may read it as perhaps /wagen/ by using English SSC correspondences. What we have argued through our decoding strand of NCELP is that we need to intervene in this automatic L1 based process, allowing learners then to access, think about, the knowledge that they have, the declarative knowledge that hopefully they have as a result of our explicit phonics teaching. Hopefully, they will remember, for example, that W in German does not say /w/ as in English but it says /v/ and that this letter a represents a long vowel so then they can work out that the pronunciation is /vagen/ so here they have come out with a more accurate pronunciation but it has taken time, so it is at the cost of fluency.

Indeed, if we look down here, there is evidence to suggest that a more accurate decoding is associating with longer response times, perhaps because learners are accessing declarative knowledge and using it to think about how should I pronounce this word. Our concern was when we were developing the phonics assessment was speeded tests, so tests when they are performing under time pressure might actually make learners more likely to use or fall back on their L1 English or main language and less likely to be able to demonstrate the knowledge at that they have gained of the foreign language phonics and our main target as we saw on the previous slide is accuracy so, ideally, we wanted an untimed, ‘unspeeded’, test. However, of course, we are operating within practical constraints so we did feel like we had to set a time limit because we wanted all of the tests to fit together in a reasonable amount of time that could be done in the context of classroom teaching. So that was an example of a compromise and we will just say a bit more about that later on.

# Slide 8 (1.11):

Turning now to print to sound decoding, so symbol sound correspondences when moving from print to sound. We have to decide how to test this. We looked at a number of previous formats that have been used, such as Erler and Macaro’s rhyme judgement task, in our Fleur study we used something that we called a Sound-Alike Task but ultimately, we felt that the most valid way of assessing print-to-sound decoding was to use a 'reading aloud test’. It is interesting that the Year 1 Phonics Screening Check that is conducted in primary schools also is an individual reading aloud task using both real words and pseudo words. One of the problem is that we come upon real world constraints and it is obviously more time consuming to administer an individual reading aloud test and to more time consuming to mark but we felt actually that this was the most valid way of assessing what we wanted to assess.

# Slide 9 (2.24):

The Year 1 Phonics Screening Check interestingly, as I just mentioned, uses both real words and pseudo words and you can see some example items of the pseudo words over here “vuss” “quop” “zook” and “chack” and then on the right some real words “flat” “skill” “gift” and “coin”. Why does the test include pseudo words? The rationale for that is because pseudo words by definition are unknown to the learners so learners cannot know these words because they are made-up words. They can’t draw on existing knowledge of how those words are pronounced and so it is purer measure of their knowledge of the SSC, of the sub-lexical correspondences between written symbols and the sounds they represent so we considered also using pseudo words for our phonics reading aloud in French, German and Spanish but we felt that this was problematic in ethical pedagogical terms because after all there are so many real words that learners don’t know and it just seemed wrong to get them to interact with made-up words. In the reading aloud task that I did for my thesis with Year 7 learners in French, learners would sometime ask me: “are these real words” or sometimes they would say “I just don’t think these are real, you’ve made them up”. It was good to be able to say to them: “No, these are real genuine French words.”

There is also some evidence that actually participants in tasks like this do treat pseudo words differently to real words in they know they are pseudo words. Of course, we have to be honest with learners and tell them what they are dealing with so in terms of validity, in terms of trying to find out how learners pronounce words and so what we did was we went for real words which we think are unfamiliar or very likely to be unfamiliar to the learners, words which they have not encountered before. In order to do that, we choose words which are low in frequency and which we know they haven’t covered to date in the NCELP Scheme of Work.

# Slide 10 (0.49):

Just to expand a little bit on this difference between familiar and unfamiliar words, so words which learners know and don’t know, I did some work on this with Year 10 pupils in French. For each participant, I created a set of orthographically set word pairs, where they knew one word in the pair but they didn’t know the other word so down here there are some examples. So, a particular learner might have known the word “pain” but not known the word “grain”. Or they might have known the word “trois” but not “chois”. The words are orthographically matched in terms of their spelling body, so they have identical spelling body and that was the bit that I analyzed in terms of their pronunciation.

# Slide 11 (1.33):

Turning to the findings of that study, the percentage of spelling bodies decoded accurately by the Year 10 participants was significantly higher for the familiar items, for the words that they knew, than for the unfamiliar words, the words they didn’t know. As an example, they would say words like “pain” correctly, the word they knew, but they would go on to pronounce this as “grain”. They would pronounce “chat”, the word they knew, correctly, but then pronounce this (<fat>) as /fat/. It suggested that they were treating the items differently in terms of how they went about pronouncing them. It suggests that they were able to recognise this word as a whole and access its pre-stored pronunciation in their long-term memory but what they don’t seem to have done is broken the word down at a sub-lexical level and seen the connection between the “ain” here and the “ain” here. If they had you would have expected them to read the words in a rhyming way as /pain/ and /grain/ but they didn’t. The same here, they don’t seem to have transferred this knowledge of the silent final consonant from the word “chat” to the identical spelling body here in this word “fat”. This really shows the value of including words that learners don’t know in a test where you are trying to get at the knowledge of SSC.

# Slide 12 (0.27):

Here is a quick summary of what we have said so far. When testing print-to-sound decoding we decided to go for a reading aloud test which would be done individually and we decided to use unfamiliar real French, German and Spanish words in order to test their SSC knowledge, avoiding familiar words to stop them from drawing on their existing knowledge of how those words are pronounced.

# Slide 13 (3.04):

We then had to decide how to test SSC when moving in the other direction so from sound to print and for this we decided upon a transcription task, so in other words can learners write down what they here accurately. One of the challenges we then came up against is that languages vary in their orthographic depth so orthographic depth is the extent to which a writing system has transparent and consistent mappings between written symbols on the one hand and spoken sounds on the other. There is much more information about this in some of the other phonics presentations so for example if you looked on the Resource Portal for the phonics presentation from Residential 1 there is more information on this but just to recap we said that Spanish we use these visual metaphors which we thought might be nice to use with pupils to raise awareness of the nature of the learning challenge that they are up against.

In Spanish we had this image of parallel train tracks because for a given written symbol there tends to be usually just one way of pronouncing it so the written letter ‘a’ tends to be pronounced /a/ but conversely for a given sounds /a/ there tends to be just one way to write it, so the written vowel a. English we said is much more complicated and we had a visual metaphor of a cats’s cradle because, and we suggested this might be a nice exercise to do with pupils actually, if you take the sounds /u/ it can be written in multiple different ways as you can see on the left and if you took one of these written representations like the ‘ough’ obviously that can be pronounced in a number of different ways so English is much less consistent in its mappings both from sound to print and from print to sound so for that reason English is considered an extremely deep alphabetic orthography. French, finally, we said was an interesting case because, and we used the image of a tree at least above the ground, because we said for a given spoken form like the word /ver/, the sound /ver/, it could be written in a number of different ways, as you can see here ‘verre’ ‘vers’ ‘ver’ and in some ways, they are not in fact real words, so pseudo-words, like these three here. So, there are actually multiple ways of writing down this same spoken sound but for a given written representation, so for each of these like the leaves on the tree, there is only one way of pronouncing them which is “ver”. So, French is interesting because it is asymmetrical in its mappings.

# Slide 14 (1.16):

So, the upshot of this was, particularly in French of our three target languages, there are multiple possible spellings for a given sound so if we are thinking about a transcription task the problem is that we end up with a proliferation of possible correct answers. As an example, if our spoken word had been ‘ver’ we would have had to allowed all of these real words and all of these pseudo words as correct answers and because the words are unfamiliar to the learners you can’t say what is the correct way to spell it. All of these would have been permissible spellings of this sound. Luckily, we managed to find a simple solution to this challenge, what we did was tweak the response format so we blanked out individual letters so in this example here if we leave two blanks like this \_ \_, then in fact there is only one correct answer in how to write the spoken form “ver” and none of the others would fit the two gaps. We used this blanking out letters mechanism just to constrain the possible responses.

# Slide 15 (1.44):

In terms of item selection, what we did for each language was we listed all of the target SSCs covered in the Scheme of Work and for each of those we then identified three low-frequency words which exemplify that particular SSC. So, in French and Spanish they are all less-frequent than the 5000 most frequent words in the language using the corpus that we used, the reference source that we used. In German, they are all less frequent than the 4,037 most frequent words and the reason for the different figures there is it is just what the frequency dictionaries that we used went up to. We also ensured that the words had not been covered previously in the Scheme of Work and so we concluded that pupils are very unlikely to know any of these words. Hopefully, we are getting a pure measure of their SSC knowledge. Where possible we also controlled for things like orthographic length, so how many letters the words contain in their written form and also their phonological length so how many syllables. In French and German, they are short monosyllabic words and in Spanish they are mostly bisyllabic. We also tried to avoid words that look like English words and we tried to avoid rude words, or words that look like rude English words and so forth, and thank you to those people who picked up on a couple of items that slipped through the net in the first pass. Hopefully, we have ended up with a set of word which meet all of those criteria.

# Slide 16 (0.55):

On this slide, you can then see an example of our set of items, these are the words we used for French. I’m not going to show you the ones from German and Spanish as well but this gives you the idea. For the silent final consonant ending in an -x we have got these words “toux” “poux” and “houx” and for ‘i’ we have “brime” “cime” and “trime” and so for. You can see that mostly we have made the words very similar because as you will see in subsequent slides what pupils get is a random selection of one of these three items for each SSC and so we wanted them to be of similar difficulty. The way we have done that is by having the same spelling body where possible, though that is not the case in all of these words.

# Slide 17 (1.01):

In designing the test, we came upon all kinds of decisions that we had to take such as how many items? Again, this was a balancing act. On the one hand, it would have been nice to have complete coverage of all the SSCs for each individual pupil then we would have been able to provide diagnostic feedback at the individual pupil level, in other words we would have been able to tell individual pupils which SSC they got right and wrong and which they needed to do further work on. But we also had to think about the so-called testing burden, how much time and effort it would take to do the test and we had to bear in mind that they are not only doing the phonics test but also the vocabulary and grammar and we wanted, as far as possible, for it to be able to fit within a single lesson and we didn’t want the assessment to soak up too much teaching and learning time because after all that is the most important thing.

# Slide 18 (2.47):

We came up with the figure of 15 items per pupil both for the reading aloud task and for the transcription task, that’s in the first iteration of these tests and we weren’t exactly sure whether that is the right number and whether it would need to change, perhaps reduce, if pupils seem to need a bit longer for each item, to think about each item. So, it maybe that if you are listening to this presentation at some later point it may be that there is now a different number of items but our initial go at this was to have 15 items per pupil so in the online test, the one that is done via the platform called Qualtrics there is a random selection from all the SSC which are covered in the Scheme of Work, there is a random selection of 15 of those. 15 for the reading aloud task and then a random selection of 15 for the dictation task and then for each of those SSC, you will remember on the previous slide, there were three words per SSC so one of these is randomly selected for each SSC for each pupil. What this mean is: first of all the class as a whole should cover all of the SSCs between them, but an individual pupil will not; and it also means that because for a given SSC there is one word out of a given random three selected it should mean that two pupils sitting next to each other are doing this test online at the same time they should have a different set of items both to read aloud and to transcribe in the dictation test so we think this has some advantages, the main one in our minds when deciding to do it this way was that it would take less time for individuals to complete and that it would be impossible for pupils to copy off each other because they are all doing different items. We think that there should still be scope for diagnostic feedback to inform planning and teaching at the whole class level so as a whole class it should still be possible to see which SSCs students, on the whole, are doing well with.

# Slide 19 (0.59):

There is a screenshot here of some of the instructions for the speaking part of the test so as you will see down here one way we are suggesting is for students to have this website open [Vocaroo](https://vocaroo.com) which allows them to record themselves, upload their recording to the cloud and generate a url which they can then copy and paste back into the Qualtrics survey. So, when the results from the Qualtrics are downloaded each pupil will have a link that can be clicked on in order to listen to what they have recorded in order to be able to mark it. The idea is that they will be able to use this, not only for the speaking part of the phonics test but also for the vocabulary and grammar test too. There are more instructions on how to do that in the overall instructions for completing these assessments.

# Slide 20 (1.27):

Again here, just some of the instructions from the speaking part of the assessment. Part 1 here being the phonics part, you can see here that we decided on two minutes to complete the reading aloud component of the speaking test overall. This is where there are 15 words to read aloud in those two minutes. As I said that number may change over time in light of people’s feedback when they have had a change to trial the tests but just a reminder that this time limit idea is really a compromise. We can see a rationale for having an untimed test, no time limit for the task to give pupils lots of time to think but at the same time it could be problematic, some pupils might take absolutely ages to complete it, some might do it very quickly, you’d get huge individual variation potentially and also we wanted it to fit roughly into a lesson overall for the assessment

We added some reassuring words down here hopefully just to reassure pupil’s that don’t worry we don’t expect them to know the words and just to pronounce them as they think they should be pronounced, to do their best and to have a go.

# Slide 21

No audio.

# Slide 22 (1.14):

On this slide, again we have a screenshot of some of the instructions for the transcription part of the phonics assessment. Just a few things to point out here, for example at the top here you can see these buttons to click in order to insert letters with diacritics. The rest of these instructions I’ll let you read through, they just cover the things that I think we have already talked about earlier in the presentation. They hear each word twice. There is a dash for missing letters. Again, we have tried to offer some reassurance here to reduce test anxiety, to reassure pupils that we don’t expect them to know the words and to just do their best. Again, there is a time limit which as we have said is a compromise really and the number of items in this iteration of the test is fifteen but that may be subject to change depending on people’s feedback once they have had a go and seen how it works. What we don’t really want is for pupils to feel incredible rushed and not be able to draw on the knowledge that we have taught them through the phonics strand of the NCELP pedagogy.

# Slide 23 (0.13):

An example here of what they would see, so they would click here to play and they would have to type in the two missing letters based on what they hear.

# Slide 24 (1.47):

We realised that logistically it might not always be easier nor indeed possible to get pupils to do the online, the Qualtrics version of these assessments so we have also produced pen and paper versions. On the left, you can see the transcription task (where they hear the words and fill in the missing letters) and on the right the reading aloud part where they see these words and read them aloud. I think I forgot to say this, same for the Qualtrics version the target SSC are underlined and bolded in each case so that pupils can direct their attention to those and not worry too much about the other letters which they are not being assessed on so they are being assessed on the particular target SSC in each case.

For these pen and paper versions we have pre-selected 15 items for each of the two tests so we selected 15 of the SSC and for each of those we have selected one of the three words. Obviously, this will be much easier to administer in the whole class setting but would also have some disadvantages so for the transcription task if you do the online version it will be marked for you so that is one advantage to doing it online. Secondly, of course pupils would never even think of this, but it would be possible in the pen and paper version for pupils to copy each other’s answers because they all get the same items and a further consequence of that is you might not get as a whole class, coverage of the SSCs because they have all got the same 15.

# Slide 25 (3.14):

Finally, a note about the scoring of the reading aloud test and particularly a word about foreign accent and how to deal with it. So, for reading aloud there are two ways to think about accuracy. First of all, there is accuracy of the SSC: so how the graphemes or the written symbols are mapped onto the phonemes or sounds of the foreign language. Secondly, the accuracy of the pronunciation of those L2 foreign language phonemes or sounds. The focus of our phonics teaching and therefore of this assessment is the former, the accuracy of of the SSC knowledge, rather than how perfectly, how accurately, those phonemes are pronounced.

So, if we took this word here as an example. First of all, a learner could try to pronounce that word using English SSC, I’m not exactly sure how they would do it, how they would pronounce the word but maybe something like “view” or “veeow” something like that. Or they might know the French SSC as a result of all this phonics teaching and they know that is is really “vo” and they might pronounce that in a very French-like way which would be wonderful. But they might also apply the French SSC correspondence and convert this grapheme /eau/ into this phoneme /vo/ but they might not be able to pronounce that very accurately in French and they might say it as something a bit more like “voo” but that would depend on your regional accent of course. We know that a foreign accent is hard to shift even for the most dedicated learners even if they really want toad for some people it is a marker of identity and perhaps they don’t really want to shift their foreign accent particularly but even if you do want to we know that is is very hard to sound native-like. You can also be intelligible and comprehensible with a foreign accent but the main point here is that we are testing SSC knowledge. We are testing phonics rather than how native-like the pronunciation is, the phonetics. Of course, there is also the question of even if you did want to test the native-like-ness of the pronunciation whose native-like would you take, there are differences in different varieties of French around the world. So obviously the upshot of all of this is that if they read the word aloud using English SSC then it will be marked wrong and if they pronounce it perfectly in French using French SSCs and with a lovely French pronunciation, obviously they will get the mark because that’s right. But we will also give a mark if they decode it using French SSC but they pronounce it in an Anglicized, or whatever depending on their own accent, if they pronounce it not really in a perfect native-like way so we will also award a mark in this case if the learner said “veu”.

# Slide 26 (0.14):

On this slide, you can see the instructions for anyone who is marking the reading aloud test and you can see on here it summarizes that information which I presented on the previous slide.

# Slide 27 (1.25):

The dictation task if done online using the Qualtrics platform will be scored, will be marked, automatically by the software which is obviously a good thing. The reading aloud task, unfortunately the technology is not yet good enough to do that automatically so that still needs to be marked individually by teachers and obviously this has workload implications. I think we acknowledge that people will want to make their own decision on how to do that and whether it is feasible but one option which we might suggest to reduce the burden of having to listen to them all afterwards and mark them which would take a long time, is to sit with individuals as they complete the speaking part of the test. In other words, if they are doing it online they could have the laptop ready or their computer ready to do the reading aloud and then you circulate around and as you get to them they do it and they just mark it as they go along. Or perhaps they come up to the front one at a time or perhaps nip outside the room, perhaps there is a seat just outside the door. Or whatever way would work in your own context. Hopefully there is a way to make that work without it being too much of a burden in terms of workload.

# Slide 28 (0.46):

So that is the end of the presentation. Here is a reminder of what we have covered. We have looked at the principles of our phonics assessments, what we are trying to test and how we might go about doing that in a valid and practical way. Secondly, we have looked at the two tasks we have used for the phonics assessment, the reading aloud and the transcription and we have talked about how we came up with designing the tasks in that way and also how we selected the items. Finally, we talked about how the scoring would work and also what we actually mean by accurate decoding. I hope that has been useful and I hope that you find the assessments themselves to be valuable and useful in your own classrooms.