



FILM 4.

USER PACK

The aim of this User Pack is to help you get the most out of our Language Assessment film series.

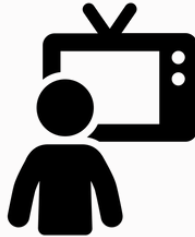
What is included?

- 1. Tips for Watching** - suggestions for how to get the most out of viewing the film.
- 2. Reflection Questions** - questions to help you relate the film content to your context.
- 3. Talk to Text Cards** - cards which provide a text version of the topics discussed in the film.



Film 4:
Assessing Speaking and Listening Skills:
What do we expect from good tests?

[Watch the film here.](#)



Tips for Watching

- Use a device with good sound quality – the audio is an important part of the film.
- Find a quiet location to watch the film – you can also watch the film as a group with colleagues.
- Use the 3 Rs as you watch: Rewind, Reflect, Repeat – pause and re-watch parts of the film to help you consider what has been said or allow more time to read on-screen text.
- Watch the film 2 or 3 times – this will help you to grasp any new ideas presented.
- Take notes while you watch – you may want to talk with colleagues about the content of the film afterwards.



Film 4: Assessing Speaking and Listening Skills: What do we expect from good tests?



Reflection Questions

Take a few minutes to think about what you have learned in the film. If you are viewing the film as a group, you can discuss these questions together.

1. It is important that tests are culturally appropriate. **What types of things should we consider when trying to make test questions culturally appropriate? What kinds of things should we consider when trying to make test stimuli (i.e., the materials we use during the test) culturally appropriate?**
2. The film describes how some children may be disadvantaged by a language test if they speak a different language at home. **Do you feel like this is an issue in your context? How?**
3. Fair tests are important because we use the information from tests to make decisions about children's education. **Have you seen tests or test items that are unfair to children in your context?**

TALK TO TEXT CARDS

Film 4:

Assessing Speaking and Listening Skills: What do we expect from good tests?

1. TESTS MUST BE CULTURALLY APPROPRIATE

In language assessment, we are looking to be able to assess vocabulary, grammar, phonology and discourse. What we are also trying to do is to capture the child's knowledge of the language, but in the process we are also trying to capture well the aspects of that language as it is in use.

It's always very tempting to think you can just take an available language test and translate it into another language. But that doesn't work because the sorts of objects in the environment, the experiences that the children have had, will differ across cultures. You need to select items that are culturally appropriate and then you have a good test.

We need to think about the range of tasks that are going to be developed and then if you are going to use pictorial materials, again, they need to be appropriate for the culture. Also, they need to be simple items which are very easy to name.

You don't want a language test to be hard because something about the picture makes it unnameable. It shouldn't be ambiguous; it shouldn't be confusable - if you give it to several people, they should all give the same name for that picture. So there are a number of different aspects of a test that need to be considered.

Fundamentally, you need to get the test items right, and that's something that requires some skill but can be easily done if you know what is usually in the environment of children of different ages - by which I mean a home environment, school environment, book environment and the culture outside of them.

2. TESTS MUST BE FAIR TO ALL CHILDREN

How can it be checked whether the tests may not be fair for some children?

Evaluation the fairness of a test through the lens of the Rasch model involves evaluating what is referred to as **differential item functioning**. This is a general idea that for certain subgroups of students, some questions on the test might be slightly easier or more difficult than for one group of students as opposed to another. Common groups of students that are evaluated when considering differential item functioning include gender or socioeconomic status of the students, or, in the context of language ability, their home language.

the way of thinking about this is, 'are the questions of equal difficulty for all subgroups of students, like boys and girls, for example'. Sometimes it is the case that certain subsets of questions can be slightly easier or more difficult for boys versus girls. That may have to do with certain factors, like, for example, their familiarity with certain toys that could appear in the question stimulus.

A common point to consider is whether the language of the test is the same as the student's first language in the home. For example, if we have a particular question which uses a commonly used word in the home but not at school, then this could actually manifest as differential item functioning. This is because that particular question becomes easier for the groups of students who speak the language of the test at home versus those who do not.

3. WHY DOES IT MATTER?

It matters in the sense of how we interpret and then use test scores. For example, if we assume that the total score on the test is a reflection of your ability as well as the relative abilities of all the scores on the test, then if that score is not just affected by your language ability, but also by your socio-economic background, your gender, etc. then all of a sudden we are making an interpretation that's confounded by other factors beyond language ability.

4. PUTTING IT ALL TOGETHER

In developing language assessments, researchers first and foremost need to be acquainted with the need for psychometric validity. This often means tests need to be derived from the context and from the culture in which they are meant to be delivered.

It's important to consider when we interpret the scores from those tests, because if we were to make the assumption that a particular score on the test in one language, for example, is equivalent to the same score in another language, then this kind of interpretation can be confounded by bias in the tests.

So we are really looking for a good language assessment that captures that entire variety, from the easy, to the mid-level, to the really difficult. So that, in a sense, is the multifaceted demand of trying to embed the language testing in the language.

