DOCUMENT RESUME

ED 350 819   FL 020 280

AUTHOR   Doye, Peter
TITLE   Authenticity in Foreign Language Testing.
PUB DATE   [91]
NOTE   12p.
PUB TYPE   Reports - Evaluative/Feasibility (142) -- Guides -
                Classroom Use - Teaching Guides (For Teacher) (052)

EDRS PRICE   MF01/PC01 Plus Postage.
DESCRIPTORS   *Language Tests; Pragmatics; *Second Language
                Instruction; *Test Construction; Testing; *Test
                Items; *Test Validity
IDENTIFIERS   *Authentic Assessment

ABSTRACT
                In foreign language testing, as in all testing,
validity is the primary criterion for test quality. However plausible
the concept of validity, in practice it is not always easy to arrive
at congruence between the test situation and the real-life situation
the learner is expected to master. Some language educators make
authenticity a major criterion of test quality. However, complete
congruence of test and real-life situation is impossible, and there
are other considerations than authenticity in testing. A language
test as a social event essentially different from any other social
event in which the learner will need to use the language. The
solution is to find a reasonable balance between authenticity and
abstraction in tests. Pragmatics, with its analyses of speech acts
and their characteristics, can be helpful in finding the right degree
of abstraction for testing. Examples of such test items include a
series of sentences of which portions are illegible and the learner
must supply appropriate words, or a paired or group activity in which
students must elicit information from each other to complete a common
task such as a survey or map completion. (MSE)

Peter Doyé

Authenticity in foreign language testing

1. Validity of foreign language tests

Classical psychometric theory has taught us to evaluate the quality of
educational tests by a number of basic criteria, such as validity,
reliability, economy and utility. Although the characteristics of a good
test can be classified in many different ways, test specialists are in
general agreement that the criteria just named are the ones that any test
producer or user should have in mind when making or applying a test.

They also agree that among the criteria mentioned above validity is the
most important, for unless a test is valid it has no function. The validity
of a test depends on the degree to which it measures what it is supposed
to measure. A good test must   serve   the purpose that it is intended
for, otherwise it is useless. However reliable the results may be,
however objective the scoring may be, if the test does not measure what
the test user wants to know it is irrelevant.

In our context most of the test users are foreign language teachers who
want to know how well their students have learnt the foreign language. For
this purpose they employ tests. My phrase "how well the students have
learnt the foreign language" disguises the complexity of the task. In the
past 20 or 30 years we have all learnt to accept communicative competence
as the overall aim of foreign language instruction. Students are supposed
to learn to understand and use the foreign language for purposes of
communication. This general aim can, of course, be broken down into a
number of competencies in listening, speaking, reading and writing.

In most countries the school curricula for foreign language instruction
are formulated in terms of communicative competencies, and a logical
consequence of this is that also testing is organized according to these
competencies. This approach to testing has been called the "curricular
approach". The foreign language curriculum is taken as the basis for the
construction of foreign language tests. On the assumption that the actual
teaching follows the content prescriptions laid down in the curriculum it
seems plausible also to determine the content of the tests on the basis
of the curriculum. This takes us back to the concept of validity. If the

If the content of a test correspondes to the content prescribed by the curriculum it is said to possess "curricular validity" or "content validity".

2. Authenticity

However plausible the concept of content validity may be, in practice it presents a number of problems. One of these problems is the congruence of the test situation and the real life situation that the learner is supposed to master according to the curriculum. It is on this problem of congruence that I wish to concentrate in my talk. The problem has been described very aptly by Edward Cureton in his article on Validity in Lindquist's well-known book on Educational Measurement:

> If we want to find out how well a person can perform a task, we can put him to work at that task, and observe how well he does it and the quality and quantity of the product he turns out. Whenever a test performance is anything other than a representative performance of the actual task, we must inquire further concerning the degree to which the test operations as performed upon the test materials in the test situation agree with the actual operations as performed upon the actual materials in the situation normal to the task. One way to do this is to make detailed logical and psychological analyses of both the test and the task. From such analyses we may be able to show that many or most of the test operations and materials are identical with or very much like many or most of those of the task, and that the test situation is intrinsically similar to that of the task. On the basis of this demonstration it might be reasonable to conclude that the test is sufficiently relevant to the task for the purpose at issue.[1]

Let us try to apply the ideas expressed in this passage to a very common task that is to be found in any foreign language curriculum: Asking the way in an English speaking environment.

If we want to find out whether students are able to perform this speech act the safest way would be to take them to an English speaking town,

place them in a situation where they actually have to ask the way and
see whether they perform the task successfully and to which degree of
perfection. We all know that this is hardly ever possible, except for
language courses that are being held in an English speaking country.
In the great majority of cases the teaching and learning of English
takes place in a non-English environment. Therefore the second case
mentioned by Cureton comes up when the tester tries to invent a realistic
situation in which the learners have to perform operations congruent with
the ones they would have to perform in situations normal to the task.
Absolute congruence would exist when the tasks in the test situation and
in the corresponding real-life situation would actually be identical.
In this extreme case the test situation and the tasks in it are called
authentic. An authentic test is therefore one that reproduces a real-life
situation in order to examine the student's ability to cope with it.

There are authors who make authenticity one of thr decisive characteristics
of a good test. They derive it from the generally accepted criterion of
validity and regard authenticity as the most important aspect of validity
in foreign-language testing.

To quote just one author who takes this view: Brendan J. Carroll:

> The issue of authenticity must always be an important aspect of any
> discussion on language testing. A full application of the principle
> of authenticity would mean that all the tasks ur.Jertaken should be
> real-life, interactive communicative operations and not the typical
> routine examination responses to the tester's 'stimuli', or part of
> a stimulus-response relationship; that the language of the test should
> be day-to-day discourse, not edited or doctored -in the interests of
> simplification but presented with all its expected irregularities;
> that the contexts of the interchanges are realistic, with the ordinary
> interruptions, background noises and irrelevancies found in the airport
> or lecture-room; and that the rating of a performance, based on its
> effectiveness and adequacy as a communicative response, will rely on
> non-verbal as well as verbal criteria.[2]

Brendan Carroll's whole book can be seen as one great attempt to ensure authenticity in language testing.

3. Limits to authenticity

It is at this point that I begin to have my doubts. However useful the postulation of authenticity as one criterion among others may be, it is certainly also useful to keep in mind that (a) a complete congruence of test situation and real-life situation is impossible and that (b) there are other demands that necessarily influence our search for optimal forms of testing and therefore relativize our attempt to construct authentic tests.

Re (a) Why is a complete congruence of test situation and real-life situation impossible? The answer is simple: because a language test is a social event that has - as one of its characteristics - the intention to examine the competence of language learners. In D. Pickett's ords: "By the virtue of being a test, it is a special and formalised event distanced from real life and structured for a praticular purpose. By definition it cannot be the real life it is probing."[3]

The very fact that the purpose of a test is to find out whether the learner is capable of performing a language task distinguishes it considerably from the corresponding performance of this task outside the test situation. Even if we succeed in manipulating the testees to accept the illocutionary point of a speech act they are supposed to perform, they will, in addition, always have in mind the other illocutionary point that is inherent in a test, namely to prove that they are capable of doing what is demanded of them.

An example of a test that examines the students' competence in asking for a piece of information: Even if by skilful arrangement we manage to lead the students to actually wanting this piece of information, they will always have another purpose of their verbal activity in mind which is: I will show you, teacher, that I am able to ask for information!

Re (b) The other obstacle on the way to perfect authenticity is an
economic one. Through a test we want to get as much information about a
person's communicative competence as possible. The greater the area of
competence we cover by giving a particular test, the better. This requires
a certain amount of abstraction from situational specifics. To use the
example of Asking the Way: What we wish to know is how well the students
can perform the speech act of Asking the Way in a variety of real-life
situations - and the more the better - and not whether they can perform
this act in the particular situation of a particular English city where
they are looking for just one building in a specific street in a certain
quarter of that city. However, we have to embed our task in a realistic
setting that contains all these specifications in order to be plausible
to the students. But this does not mean that we have to include all the
incidentals that might be properties of such a real-life situation. On the
contrary: the more incidentals we include, the more we move away from the
general concept of Asking the Way as most of these incidentals might not
be present in the majority of other situations where "asking the way" is
demanded. Therefore we need not be sorry if we do not succeed in making
a test situation absolutely authentic by providing all the peculiarities,
background noises, hesitations, interruptions, social constraints by
which a real-life communicative situation is characterized. We should
endeavour to employ just the amount of realism that makes it understandable
and plausible, but no more. The fact that we want to know how well the
students master the essentials of our speech act requires abstraction from
incidentals. Pickett gives the example of a simple arithmetic problem:

> If you are asked to find the area of a field 50 metres x 200 metres
> you do not have to get up and walk all over the field with a tape
> measure. You will not be concerned with whether it is bounded by a
> hedge or a fence, whether it is pasture or planted, whether it is
> sunny or wet, or whether it is Monday or Thursday. These incidentals
> are irrelevant to the task of measurement, for which the basic
> information is ready to hand, and we know that the solution will not
> be affected by weather, time, cultivation, perimeter markings or any
> of the other factors which form part of our real-life perception of
> any particular field. The concept of area is an abstraction from all
> possible perceptions and is a constant.[4]

We have to concede that the decision about what are irrelevant incidentals
is easier to make in the case of an arithmetic problem than in a communicative
task, as communicative performance is always embedded in concrete situations
with a number of linguistic as well as non-linguistic elements. But the
arithmetic problem and the communicative task have one thing in common:
Normally, i.e. outside the artificial classroom setting, they occur in real-
life situations that are characterized by a small number of essential features
and a great number of incidentals which differ considerably from one situation
to the next. And if we want to grasp the essential features of a task, we have
to abstract from the incidentals. In this respect abstraction is the counter-
point to authenticity in testing.

What is needed is the right balance between authenticity and abstraction.
We want a fair amount of authenticity but not so much as to obscure the
essential properties of the speech act in question, which by virtue of being
essentials obtain in all its manifestations. In this context, the findings
of modern pragmatics can be of great help, I think. Its analyses of speech
acts have demonstrated that every speech act has its own specific structure
with certain characteristic features. It is on these characteristics that
we have to concentrate if we wish to test the learners' competence in
performing this particular act.

4. Examples

Let us take "Asking for Information" as an example. In his classical book
"Speech Acts. An essay in the philosophy of language" John Searle has
developed a systematic procedure for the description of speech acts, in
which he presents the characteristic features of each act in terms of
four kinds of conditions that are necessary and sufficient for the
successful and non-defective performance of each act. The speech act of
"asking for information" or simply "question" is one of the examples that
Searle uses himself.

The essential characteristic of a question is that it counts as an attempt
to elicit information from a hearer. The two preparatory conditions for
the performance of a question are that the speaker does not know the

answer and that it is not obvious that the hearer will provide the information without being asked. The propositional content of a question depends on what information the speaker needs, of course.

Now, we all know that teaching as well as testing the ability to ask questions is often practised in a way that disregards these conditions. A very common way is to present a number of sentences in which certain parts are underlined and to invite the students to ask for these parts.

Holburne Museum is situated in Pulteney Street.   Example 1
It belongs to the University of Bath.
It is open daily from 11 a.m. to 5 p.m.
Mr. Green works in the Museum library.
He goes there every second morning.
He gets there by bus No. 32.
It takes him right to the main entrance.

This procedure is often used for the simple reason that it is easy to prepare, to administer and to score. But it very obviously violates the essential rules that govern the performance of a question. First of all, the speech act demanded cannot be regarded as an attempt to elicit information. Secondly, the testees do very well know the answer because it is given to them in the statements. It is even underlined, which normally means that the piece of information given is especially important. - a fact that stresses the non-realistic character of the task.

And there is an additional negative feature: the procedure complicates the task for all those learners who find themselves incapable of imagining that they do not possess precisely the information that is given to them and to behave accordingly, i.e. to pretend that they need it.

To conclude: The questions that the students have to ask in this test are no questions at all. The conditions under which they have to perform their speech acts are so basically different from those of real questions that the test cannot be regarded as a means to examine the students' competence in asking questions.

Let us look at the next example which could serve as an alternative to
the previous one: Example 2

Holburne Museum is situated in *[illegible]*
It belongs to the *[illegible]* of *[illegible]*
It is open daily from 11 *[illegible]* to *[illegible]*
Mr Green works at the *[illegible]* *[illegible]*
He goes there every *[illegible]* *[illegible]*.
He gets there by *[illegible]* *[illegible]* *[illegible]*.
It takes him right to the *[illegible]* *[illegible]*.

The difference between the two types of test is minimal on the surface,
but decisive as regards the speech acts that are required to perform the
task. By a very simple design, namely through replacing the underlined
parts of the sentences by words that are illegibly written, the second type
makes a considerable step forward in the direction of an authentic test:
The questions that the learners have to ask are real questions in so far
as the two main conditions of the speech act "QUESTION" as elaborated by
Searle are fulfilled.

First, they can be counted as attempts to elicit information and, second,
the testees do not know the answers yet. What is still missing is an
addressee to whom the questions might be put.   Illegible statements
are quite common, but one would hardly ever try to obtain the lacking
information by a list of written questions. To make this test still more
realistic, one could present the statements not in writing, but in spoken
form with a muffled voice that fails to be clear precisely at those points
where one wishes the students to ask their questions. In this case all the
essential conditions of the speech act "QUESTION" would be fulfilled. But
the test is still far from being authentic.

In a real-life situation one would rarely find such a concentration of
unintelligible utterances and therefore the necessity for a whole series
of questions. Of course we can think of situations in which the necessity
for quite a number of successive questions arises, such as in the situation
of an interview or the situation of a game in which two partners need
certain information from one another in order to complete a common task.

Example 3

The students have to imagine a situation in which they want to find out
about the working conditions of a group of people and have to ask these
people a number of questions in order to make a survey.

S U R V E Y			
Q u e s t i o n	John Payne		
Job			
Number of years in job			
Payment			
Training			
No. of working hours			
Overtime			
Difficult / easy			
Like / dislike			
Holidays			
Atmosphere of work			

Example 4

The students work in pairs. They are given two incomplete maps,
- one for each partner - in which certain geographical information
is missing, and they have to ask each other for the missing
information in order to complete their maps.[4]
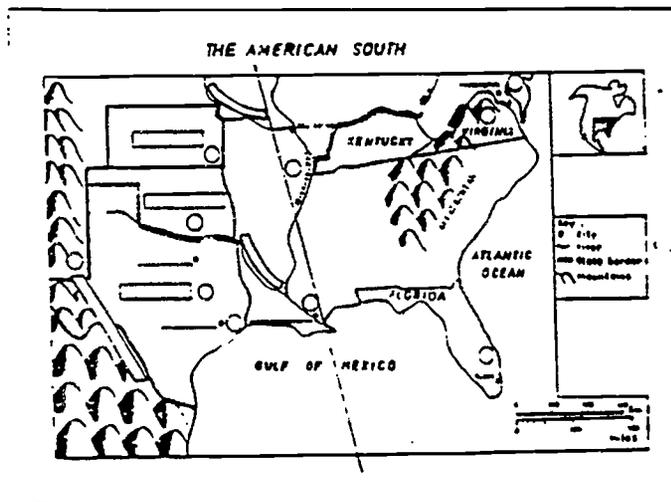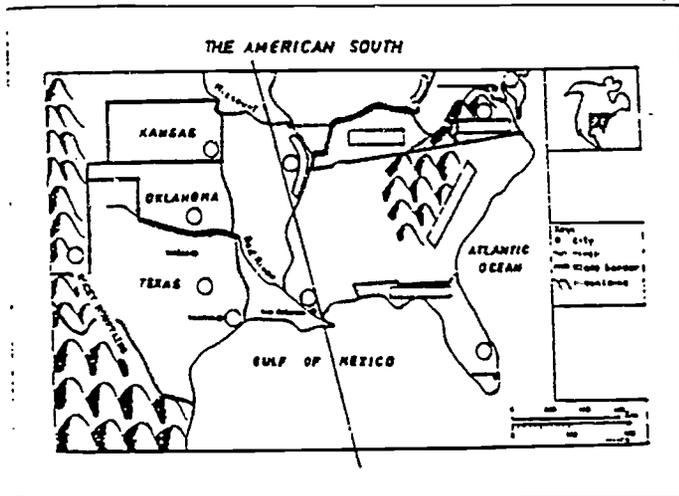
5. The balance between authenticity and abstraction

But to come back to our central problem: How far do we want to go
in our efforts to create authenticity?

In the middle part of my talk, I tried to explain why absolute
authenticity i.e. complete congruence between the test situation
and the so-called real life situation is neither possible nor
desirable.

However much, for validity's sake, we might want to achieve
authenticity in our tests, any attempt to reach it will necessarily
arrive at a point, where it becomes clear that there are limits
to authenticity for the simple reason that a language test - by
its very purpose and structure - is a social event that is
essentially different from any other social event in which language
is used.

Very fortunately, we need not be afraid of protests from our
students. They might be better motivated if we succeed in con-
structing tests that are highly authentic, for then they see the
practical relevance of their tasks.

On the other hand most of them see as we do that a test can never
become absolutely authentic and might find the vain attempts of
their teachers to create fully authentic test situations fairly
ridiculous. Therefore, and for the two main reasons I have pre-
sented we should give up our efforts to achieve the impossible
and be satisfied with finding the right balance between authen-
ticity and abstraction.

1 Cureton, Edward E.: Validity
   In: Lindquist, E.F. (ed.): Educational measurement, American
   Council on Education, Washington, D.C. 1963, p. 622.

2 Carroll, Brendan J.: Testing Communicative Performance
   Pergamon Press, Oxford, 1980. p. 11f.

3 Pickett, D.: Never the Twain ...?
   p. 7.

4 From: Doyé, Peter & Rampillon, Ute: Vertretungsstunden für den Englisch-
   unterricht, Hueber, München 1986, p. 43.