A Study on a Way to Improve the Content Validity of a Classroom "Knowledge" Test in Foreign-Language Courses

Yasusuke Ogawa

As is recognized by most teaehers, periodic assessment of educational progress is essential to effective education and good tests afford very useful assistance to teachers in making those assessments. The teachers are concerned about quality in education — about how much their students have achieved, about how adequate their teaching is, about how successful they are as teachers and how effective the whole business of education is. And their concern about individual students is to do something about them, individually or collectively. Continually they are faced with the necessity of arriving at some decision as to a course of action. They have to decide what to do about an individual or individuals. And they want each one of their decisions to be sound and well—conceived ones. Here it is assumed that "sound decisions arise out of relevant knowledge of the individual or individuals," as Thorndike and Hagen say in their *Measurement and Evaluation in Psychology and Education* (p. 8). In other words, we may say that the more we know about a person, and the more accurately we know it, the more likely we are to arrive at a sound decision about him or a wise plan of action for him.

However, there might be some who oppose to the need of formal classroom tests, saying that a good teacher, working with a class of reasonable size, had no need for tests in order to make sufficiently accurate judgments of student achievement because direct teacher observation is likely to provide a sufficient basis for assessing student achievement. When it comes to the development of physical skills or social behaviors.

they are right. Direct observation provides a much better basis for assessing such skills and behaviors than a written test. Teachers should not ignore their own observations of a student's level of understanding or ability to use knowledge. At this point, we should remember that testing is not really an alternative to teacher observation of student behavior. Testing and teacher observation are not mutually exclusive to each other. "Testing is simply a specialized technique for extending, efficiently recording, and summarizing those observations." [1]

Then, what are the functions of classroom tests? The major function of a classroom test is, as mentioned above, is to measure student achievement and thus to help teachers to diagnose student achievement and also to contribute to the evaluation of the student's educational progress and attainments. Tests can help teachers to give more valid, reliable grades.

Another major function of a classroom test is to motivate and direct student learning. Even when the aims and objectives expressed in the curriculum are sound and well-balanced, the test, when not sound or well-balanced, is likely to direct student learning to undesirable or sometimes harmful goals or results. On the contrary, if the students know in advance they will be tested, if they know the kinds of knowledge and ability the test will require, and if the test does a good job of measuring the achievement of essential course objectives, tken its motivating and guiding influence will be most wholesome. Anticipated tests work as extrinsic motivators of learning efforts, even though they are less desirable or effective than intrinsic motivators. For the grest majority, this kind of motivation provided by tests and other influential factors is indispensable.

Classroom tests have other educational functions. The process of constructing them will cause an instructor to be aware of, and to think carefully about, the goals of instruction in a course. And for the students the process of preparing for a test, of taking it, and of discussing

1 Robert T. Ebel; *Measuring Educational Achievement* (Englewood Cliffs, New Jersey: Prentice—Hall Inc., 1965) p. 4.

it afterward, can be a rewarding learning experience.   Here testing
and teaching need not be thought of as two separate things or mutually
exclusive alternatives.   They are intimately related parts of the total
educational process.

Thus tests can be, and usually are, highly important tools for the
promotion of learning.   We should remember that the educational value
of a test depends on its quality and on the skill of the teacher in using
it.   Good tests, properly made and used, can make valuable contribu-
tion to student's education.   On the other hand, poorly constructed or
poorly administered tests might even do educational disservice or harm
to students.

When we think about the kinds of classroom tests in the foreign—lan-
guage teaching field, we can roughly classify tkem into two groups — that
is, (7) proflciency measurement or skill tests and (2) knowledge measure
ement.   The former kind aims at obtaining information on the student's
mastery of language skills taught in foreign—language courses.   This
takes the form of inventories showing the student's "mastery" and "lack
of mastery" of a number of specific linguistic points in broad skill areas.
On the other hand, knowledge measurement is designed to make an assess-
ment of the student's acquisition of facts or concepts about the foreign
language or related subjects.   In this kind of test, emphasis is on the
intellectual message conveyed rather than on the medium of expression.

It has formally been thought that foreign—language tests mostly con-
sist in, or mean, proficiency measurement or skill testing.   But actual-
ly, foreign—language courses can, and should, have different types of
content and intellectual goals in addition to teaching linguistic skills.
Especially in advanced courses in foreign languages are included the
teaching of the "culture" of the foreign—language country, the study of
"literature" including an appreciation of, the content, historical back-
ground, and stylistic qualities of literary works, and other courses for
language majors deal with the foreign language as a subject matter, such as,

courses in phonetics, linguistics, grammar and similar areas of study, in which the student learns the language just as in other content courses.

Tests designed and used for measuring student attainment in these three broad areas, culture, literature and language as a subject matter may be classified as "knowledge" tests to indicate that they mainly aim at the assessment of the student's acquisition of facts and concepts.

So in this short paper, I would like to discuss ways to make good classroom tests or to imporve the most fundamental or essential quality of classroom "knowledge" tests in foreign language teaching.

The Requirements of a good test — the validity of tests

What is a good test? It is one of the most basic questions for those who make and use classroom tests. All good tests necessarily possess one common, essential quality among other things, that is, *validity.* A good test has to be valid first of all. To be sure, there are other test characteristics which are also of value, but this is an indispensable quality for a test to be good. Without this quality a test would be a poor investment in time and energy.

We cannot too much emphasize the importance of the validity of a test. What is validity? To answer this question we may put it this way. The final proof of the quality of a test must hinge on the answer to a simple and straightforward question: "How well is it doing what it is supposed to do?"[2] Or we may say that "the extent to which a given test fulfills the purpose for which it is designed is the bench mark against which it must ultimately be judged."[3] This is the question of the validity of a test.

As might have been noticed, the validity of a test consists of reliability and relevance. The first part of the above—mentioned question, that is, *"How well* does the test measure?"* concerns the question of reliability. And the second, that is, *"What* precisely does the test measure?"* is that of relevance. In other words, to be valid, a test must be both relevant and reliable at the same time. Relevance and

reliability are necessary conditions for validity.

Reliability is clearly an important component of validity. Another component is relevance. A test which measures with a high degree of accuracy what it is intended to measure is highly valid because it is both highly relevant and highly reliable. But a test could be highly reliable, in theory at least, without being highly valid if lacked a high degree of relevance." [4]

It follows, therefore, that relevance alone or reliability alone does not or cannot make a test a valid one. Either of the two is a necessary condition for validity but can not be a sufficient condition alone. Both of them should go hand in hand in order to make a valid test.

In test construction, validity is the question of what should go into a test or what should be selected to be measured by a test. So, for most classroom tests, relevance is a matter of logical analysis and expert judgment. It usually cannot be judged or measured statistically after the test has been given. What a test actually does measure depends upon the decisions the test maker makes from time to time as he works to build the test. The cumulation of his decisions made at each step in the test—making process determines the relevance of the test. If all the test items are relevant, then the test as a whole is relevant. In other words, if the individual test items require demonstration of mastery of some essential aspect of the course and if they sample or represent proportionally all those essential aspects, then the test as a whole will be relevant.

It naturally follows that good planning is an indispensable stage to produce a well — balanced, good test. So we would like next to consider the steps included in that preparatory stage and what should be

2 John L. D. Clark, *Foreign Language Testing: Theory and Practice* (Philadelphia: The Center for Curriculum Development, Inc., 1972) p. 1.

3 Ibid., p. 1.

4 Ebel, p. 390.

done in each step of the stage.

The Planning of the Achievement Test

The following series of steps will be found quite useful for the develop-
ment of relevant and well-balanced achievement tests.

a) Identify the level or levels of learning to be measured by the test.

b) Define the learning outcome in terms of specific, observable beha-
avio

c) Outline the subject-matter content to be measured by the test.

d) Prepare a table of specifications.

e) Use the table of specifications as a basis for preparing the test.

The most important thing to be considered in test planning is to deter-
mine *what* is to be measured, and to describe it in such precise terms
that test items can be constructed that call forth the desired behavior.

I. Identifying the levels of learning to be measured

This, though often overlooked, is a most important step because it
deals with a fundamental fact, namely, that learning or achievement covers
a range of different levels.   For example, the lowest level of
learning found in schools is typically the accumulation of factual infor-
mation.   Another level is the application or use of factual information.
We differentiate these levels because it is known that different skills
are involved.   If a teacher wishes to help students apply knowledge but
uses only items that test the learning of facts, he has no test data for
assessing the application level of achievement.   Therefore, it is impor-
tant that the teacher recognize the various levels of learning and to
select those levels that are pertinent to the course.

For selecting the levels of learning, several good contributions have
been made to the literature on learning levels.   Perhaps the most use-
ful guide is the *Taxonomy of Educational Objectives* by Benjamin S.
Bloom and others (New York: David Mckay Co., Inc., 1956).   This
is a comprehensive system for classifying objectives within each of three

domains : (1) cognitive domain, (2) affective domain, and (3) psychomotor domain.　The cognitive domsin of the taxonomy is concerned with intellectual outcomes, the affective domain with intereste and attitudes, and the psychomotor domain with motor skills. The classification system for the cognitive domain and affective domain have been developed and published, and the development of the psychomotor area is under way. Since our concern here is with achievement testing, we shall focus attention on the cognitive domain only.

The cogniitive domain of the taxonomy

　　Intellectual outcomes are divided into two major classes : Knowledge and Intellectual Abilities and Skills.　These are further subdivided into six main areas as follows :

　　Knowledge

1. 00 Knowledge (remembering previously learned material)

　　　1. 10 Knowledge of specifics
　　　　　1. 11 Knowledge of terms
　　　　　1. 12 Knowledge of specific facts
　　　1. 20 Knowledge of ways and means of dealing with specifics
　　　　　1. 21 Knowledge of conventions
　　　　　1. 22 Knowledge of trends and sequences
　　　　　1. 23 Knowledge of classification and categories
　　　　　1. 23 Knowledge of criteria
　　　　　1. 25 Knowledge of methodology
　　　1. 30 Knowledge of universals and abstractions in a field .
　　　　　1. 31 Knowledge of principles and generalizations
　　　　　1. 32 Knowledge of theory and structures

　　Intellectual Abilities and Skills

2. 00 Comprehension (grasping the meaning of material)

　　　2. 10 Translation (Converting from one form to another)
　　　2. 20 Interpretation (Explaining or summarizing material)
　　　2. 30 Extrapolation (Extending the meaning beyond the data)

3. 00 Application (Using information in concrete situations)

4. 00 Analysis (Breaking down material into parts)

　　　4. 10 Analysis of elements (Identifying the parts)
　　　4. 20 Analysis of relationships (Identifying the relationships)

4.30 Analysis of organizational principles (Identifying the way the parts are organized)

5.00 Synthesis (Putting parts together into a whole)

5.10 Production of a unique communication
5.20 Production of a plan or proposed set of operations
5.30 Derivation of a set of abstract relations

6.00 Evaluation (judging the value of a thing for a given purpose using definite criteria)

6.10 Judgments in terms of internal evidence
6.20 Judgments in terms of external evidence

The above outline is brief and hence does not supply detailed treatment of each level of learning.   Yet, it points up the fact cognitive behavior covers a wide range and therefore proper assessment of the entire range must involve test items of many different kinds.   The most common error made in achievement testing is that too many items are confined to testing only knowledge, thus neglecting somewhat the testing of intellectual abilities and skills.

As can be seen fn the above taxonomy, the outcomes are arranged in order of increasing complexity.   They begin with the simple recall of factual information, go to the lowest level of understanding (comprehension),   and then proceed through the increasingly complex levels of application, analysis, synthesis, and evaluation,   The subdivisions within each area are also in order of increasing complexity.   This scheme is therefore hierarchical in nature.

The cognitive domain of the taxonomy is especially useful in planning the achievement test.   It focuses on a comprehensive and apparently complete list of mental processes to consider when identifying learning outcomes, and it serves as a guide for stating learning objectives in terms of specific student behaviors.   Although the teacher should not follow

Benjamin S. Bloom and others, *Taxonomy of Educational Objectives: The Classification of Educational Goals* (Handbook I: Cognitive Domain) (New York: David Mckay Company, Inc., 1956)

the taxonomy slavishly or rigidly, it provides a good set of practical suggestions in the first stage of planning.

The instructional objectives for a particular course will depend on the nature of the course, the objectives attained in previous courses, the philosophy of the school, the special needs of the students, and upon other factors bearing on the instructional program. But despite the variation from course to course, most lists of instructional objectives will include learning outcomes in the following areas: (1) knowledge, (2) intellectual abilities and skills, (3) general skills (laboratory, perfor-mance, communication, work—study), (4) attitudes. interest, and appre-ciations. It is in the first two areas, covered by the cognitive domain of the taxonomy, that knowledge achievement testing is most useful. Thus, the first step is to separate from the list of instructional objec-tives those that are testable by paper-and-pencil means. If the instructional objectives have not yet been determined, the cognitive domain of the taxonomy can be used as a frame of reference for doing it.

While the cognitive taxonomy provides a valuable guide for identifying levels of learning, there is no expectation that all of the areas listed will be covered in a particular test or even in a single course. The classification scheme is also neutral concerning the relative importance of the levels listed. Thus, it is the classroom teacher who must decide which levels will guide his teaching and testing, and how much emphasis each one will receive. The taxonomy merely serves as a convenient checklist which prevents relevant areas of learning from being overlooked during the planning for an achievement test.

Ⅱ. Defining the objectives in specific terms

When a satisfactory list of learning levels has been chosen, the next step is to list the specific behaviors which are to be accepted as evidence of adequate achievement. For example, what specific behaviors will show that a student "knows the common terms used in linguistic descrip-tion of a language" or "understands the principles of grammatical analysis

of a language"?　For these two areas, specific behaviors may be
listed as follows ;

(1) Knows common terms used in grammatical analysis of a language
(here we have only a general statement, which must be further specified
as follows ;)

　　1.1 Recgonizes the correct definitions of terms

　　1.2 Identifies the meaning of terms when used in context

　　1.3 Distinguishes between terms on basis of meaning

　　1.4 Selects the most appropriate terms when describing analytical
　　　　procedures.

On the knowledge level, according to Bloom's taxonomy, the above list
may be regarded to be satisfactory for a particular course.　Let us now
identify the specifics pertinent to the course concerning a higher level of
learning.　Suppose we choose the level of synthesis (level ♯5) as the
most desirable target.　Again, we first state a general objective con-
sistent with synthesis and then proceed to break it down into specific
behaviors, as illustrated below.

(5) Understands the principles of grammatical analysis.

　　5.1 Describes each principle in his own words

　　5.2 Gives a specific example of each principle

　　5.3 Explains the relevance of each principle to the major steps in
　　　　grammatical analysis

　　5.4 Predicts the most probable effects of violating each of the prin-
　　　　ciples

　　5.5 Formulates an analysis plan which is in harmony with the princ-
　　　　ciples

As you notice, the terms used to describe the specific behaviors indicate
behaviors which can be demonstrated to an outside observer.　That is,
they are observable behaviors which can be called forth by test
items.　The key terms are listed below, to emphasize what is meant
by defining learning in *specific behavioral terms.*

recognizes	gives an example
identifies	explains
distinguishes between	predicts
selects	formulates
describes	

Terms such as these indicate rather precisely what the student is able
to do to demonstrate his achievement. Such vague and indefinite terms
as "learns," "sees," "realizes," "has acquired," "has developed," and
"is familiar with" should be avoided, since they do not clearly indicate
the terminal behavior to be observed.

In defining the general outcomes, all of the relevant behaviors, of
course, cannot be listed. It is generally satisfactory to include a suf-
ficient number to clarify what the typical student is like who has achieved
the outcomes.

In his *Preparing Instructional Objectives* (San Franeisco: Fearon
Fublishers, Inc., 1962), Mager has suggested that in stating specific
learning outcomes the statement should include three elements: (1) the
specifie behavior, (2) the conditions under which the behavior is to be
demonstrated, and (3) the standard of acceptable performance. Thus,
a statement of a specific learning outcome might appcar as follows:

"When given ten grammatical analysis terms to define in his own
words, the student can correctly define at least eight of them."

Ⅲ Outlining the sudject-matter content

The learning outcomes specify how students are expected to react to
the content of a course. Although it is possible to include both the
student behavior and the specific subject matter, the student is working
toward, in the same statement, it is usually desirable to list them sepa-
rately. This is so because the student can react in many different
ways to the same aspect of content. For example, when we state that
a student can "define a term in his own words," "recall a specific fact,"
or "give an example of a principle," these behaviors can be applied to

almost any aspect of subject matter. Similarly, in studying the taxo-
nomy of educational objectives we may expect students merely to recall
the categories in it, or we could require them to explain the principles
on which it is organized, to summarize its usefulness in test planning,
to classify a given set of learning outcomes with it, or to use it in the
actual construction of a test. Since student behaviors can overlap a
variety of subject matter areas, and vice versa, it is more convenient
to list eaeh separately and to relate them in the table of specifications.

The content of a course may be outlined in detail for teaching pur-
poses, but for test planning only the major categories need to be listed.
The following outline of subject matter topics covered in texts on gram-
matical analysis illustrate sufficint detail for the test plan.

The following topics are borrowed from *English Transformational Grammar*
by R. A. Jacobs and P. S. Rosenbaum.

   A. The study of language
      1. Language as a scientific subject matter
      2. Constituent structure
      3. Deep structures, surface structures, and Transformations
      4. Transformations and elementary transformational processes
      5. Linguistic explanation and ordered rules

   B. Constituents and features
      1. Determining the constituents of a sentence
      2. Noun phrase constituents
      3. Verd phrase constituents
      4. Features, lexical items, and deep structures
      5. Constituent functions

If a test were being planned to cover the total content of a course, it
might be necessary to include only the major headings (A and B) to
prevent the outline from becoming unmanageable.

In using the topics in the course for illustrative purposes, there is
no implication that the content outline should be limited to the material
in a particular book. An achievement test is typically designed to
measure all of the course content, including that covered in class dis-
cussion, outside reading, and any other special assignments. Our example

here is meant to illustrate the approximate amount of details and not
the source of the topies to be included.

IV. Preparing a table of specifications

When the learning outcomes have been defined and the course content
outlined, a table of specifications should be prepared.   This is a table
which relates outcomes to content and indicates the relative weight to
be given to each of the various areas.   The purpose of the table is to
provide assurance that the test will measure a reqresentative sample of
the learning outcomes and the content to be measured.

An example of a table of specifications for a segment of a gram-
matical analysis course is given below.   A more detailed table may be
desirable for some purposes, but this is sufficient for illustration.

Table 1

"Specifications for a test on two topics:
"The study of language," and "Constituents and features."

Objectives (Briefly identified)	Subject Matter Content		Total Number of Items
	The study of language	Constituents and Features	
Grasps Terminology	5	5	10
Grasps Procedures	4	6	10
Knows the Kinds of Transformations and Constituents	4	6	10
Understands Principles	5	5	10
Applications in Analysis	4	6	10
Total Number of Items	22	28	50

The number in each cell of the table indicates the number of test items to be devoted to that area. For example, 10 items in the test will measure knowledge of terms; five of them pertaining to the "study of language" and five of them pertaining to "constituents and features." The number of items assigned to each cell is determined by the weight given to each learning outcome and each subject matter area.

In assigning relative weights to each outcome and each content area, a number of factors will enter into their determination. How important is each area in the total learning experience? How much time was devoted to each area during instruction? Which outcomes have the the greatest retention and transter value? What relative importance do curriculum specialists assign to each area? These and similar criteria must be considered in deciding on the relative importance of each item. In the final analysis, however, the weights assigned in the table should reflect the emphasis given during instruction.

In summary, the preparation of a table of specifications includes the following steps :·

1. Identify the learning outcomes and content areas to be measured by the test.

2. Weight the learning outcomes and content areas in terms of their relative importance.

3. Build the table in accordance with these relative weights by distributing the test items proportionately among the relevant cells of the table.

The resulting two-way table indicates the type of test needed to measure the outcomes and course content in a balanced manner.

V. Using the table of specifications in test production

The table of specifications serves the the test-maker like a blueprint. It specifies the nature of each item in the test. If the table has been carefully prepared, the quality of the test will then depend, largely, on how closely the test-maker can match the specifications,

that is, on how skillfully he can construct test items that call forth
the specific behavior described in the learning outcomes.   While we
can never be certain of a perfect correspondence, this is the key to
effective achievement testing.

In conclusion, it may be said that classroom "knowledge" tests
are essential and effective tools to make our direct observation of
the student's achievement sharper and to make our assessment of the
student's achievement accurate and consequently to make our daily
educational activies effective.   But at the same time, these supposed-
ly effective tools might turn out to be dangerous or harmful ones to
the student's sound and wholesome educational progress, when they are
poorly or improperly prepared and used.

For our tests to be really educationally effective and helpful tools,
they should, first of all, be valid.   The validity, especially the con-
tent validity of tests is the minimum essential among the various
qualities that tests should possess.   This type of validity of a test
necessarily depends upon the test-maker's (1) sound "expert" judgment,
which must eventually pass the test of independent verification, and
(2) techniques and knowledge for channelling his knowledge and judgments
properly into the construction of a classroom test, so that his test
may be a well-balanced one, the one which contains the representative
smapling items of the course content in a well-balanced manner.

In order to make good planning for making a valid test, we should
know in advance what level or levels of learning are to be measured,
what learning outcomes are to be involved in the measurement and also
whichs subject—matter content is to be measured by the test.   And
these learning levels, learning outcomes and subject-matter content should
be proportionately integrated in the finished test.

REFERENCES

Bloom, Benjamin S. and others. *Taxonomy of Educational Odjectives: The Classification of Educational Goals.* Hew York, 1956.

Clark, John L. D. *Foreign Language Testing: Theory & Practice.* Philadelphia, 1972.

Edel, Robert L. *Measuring Educational Achievement.* Englewood Cliffs, New Jersey, 1965.

Harris, David P. *Testing Engli sh as a Second Language.* New York, 1969.

Lado, Robert. *Language Testing.* London, 1961.

Thorndike, Robert L. and Elizabeth Hagen. *Measurement and Evaluation in Psychology and Education.* New York, 1961.