CROSS-CULTURAL TESTING: WHAT TO TEST

J. A. Upshur
University of Michigan

WITH EVER INCREASING numbers of people crossing national, linguistic, and cultural boundaries in order to study and to work, it becomes more crucial to provide instruction and orientation so that they can function effectively in their new locations. A corollary need is to identify those people most in need of instruction or orientation, and to determine what the content of instructional and orientation programs should be.

Considerable work has already been done in devising techniques and tests for the identification of those persons who intend to cross linguistic borders but who haven't sufficient language ability for their new undertakings. Applied linguists and language teachers have done much to determine what the content of foreign language instruction properly should be. Experience, especially that with foreign students in the United States, has shown, however, that measures of language ability alone have limited power to predict who will be able to function effectively in the new linguistic and cultural environment. This is no longer surprising, if indeed it ever was. It has become a cliche to observe that some foreign student is performing poorly because he is "suffering from culture shock."

Such cultural orientation programs as are administered by universities, fellowship sponsoring agencies, professional societies, etc. are handicapped to the extent that participants cannot be "graded" on their lack of cultural understanding before these (usually brief) orientation programs are well under way. They likewise suffer from too little information that specifies which aspects of the new culture are not understood by the participants.

There exists, therefore, a clear need for test instruments and procedures which can supply reliable and valid measures of cultural understanding. As a stimulus to further attempts, as a guide for test construction, and because it is a by-product of an instrument which may soon be in use, Seelye's "Field Notes on Cross-Cultural Testing" is a welcome beginning.[1] And

---

[1]H. Ned Seelye, "Field Notes on Cross-Cultural Testing," *Language Learning*, XVI, 1-2 (1966), 77-85.

precisely because his article is a beginning (as Seelye empha-
sizes[2]), it is an invitation to another step. The purpose of this
reply is twofold: to examine in some detail the direction of
Seelye's beginning in order to ascertain the direction of a next
step, and to anticipate the limitations of tests of cross-cultural
understanding which would be produced by test writers continu-
ing in that direction.

     The purpose of cross-cultural testing which was implied
above should foster the development of tests which are formally
the same as foreign language tests.   That is, the skills, the
abilities and the knowledge to be tested are most relevant to
communication. In language testing one is most concerned with
whether the "linguistic stranger" can speak or write in such a
manner that his intended meanings are understood by the mem-
bers of the foreign language community, and whether he under-
stands their intended meanings when he hears or reads what
they have produced in their language. In cross-cultural testing
one is concerned with whether the "cultural stranger" can be-
have (non-linguistically) in such a manner that his intended
meanings are understood by members of the foreign culture
community, and whether he understands their intended meanings
when he observes their behavior or the products of their be-
havior. Obviously cross-cultural testing is not restricted to
directly observable behavior and behavioral products.   It is
perhaps equally important that the "cultural stranger" should
understand the intended meanings of behavior which is verbally
reported to him, and that the intended meanings of behavior
which he reports are understood. This point is simply that a
school teacher from Cleveland, for example, should not only be
able to understand the intentions of two gentlemen whom she
sees embracing on the platform of the Rome railway station,
but should also understand if the incident is reported to her.

     It should be pointed out that the purpose of Seelye's article
was not to present a test description or a program for test
construction, but to suggest some of the problems encountered
during the preparation of cross-cultural tests.   He succeeded
remarkably well. After reading the paper no one could remain
unaware of the amount of work involved in preparing such a
test. Considerable effort is required in deciding what the content
of the test should be, in writing items to reflect the decisions

---

[2]". . .any conclusions insinuated in this paper should be regarded as tentative; here
we are simply presenting field notes which might be helpful to those engaged in cross-cul-
tural testing." *ibid*, p. 77.

about content, in locating samples for pretesting items, in administering pretests, in analyzing pretest results, and in rewriting items and pretesting again.   One can also appreciate the decisions which must be made about statistical criteria for acceptable items; the decision is necessary because no item will be answered one way by everyone who knows the target culture and another way by everyone who is naive.   One might wish that Seelye had written more about the intriguing problem of "cultural dialect" or "cultural style."   Some of his examples of rejected items illustrate sex differences and differences according to the respondents' social class.   Seelye also mentions age and urban or rural residence as factors which might affect responses on cross-cultural test items, but he does not address directly the central problem which these examples and factors imply—the problem of determining a "core" or "standard culture" shared by all members of a cultural community.   It is certainly arguable that men and women, children and adults live in "separate worlds," that the peasant can never understand the ways of high society, and that a man from the country is always a "hick" when he comes to town.   But the acceptance of such assumptions will not relieve one of discovering the "cultural dialect" in which rural, female, middle-class children (for example) communicate and interact with one another.

One might also wish that Seelye had been more explicit in telling why he wanted a cross-cultural test in the first place, in stating the nature of the information he hoped to derive from his test.   A test which is wanted for ranking people who have lived in a new cultural community according to how much their preconceptions have changed will very likely differ from a test which is wanted for ranking people according to their abilities to use and interpret the patterns of the foreign culture[3] accurately and unambiguously.   Seelye states only that he wants a test which will "determine approximately the level of cross-cultural awareness attained after residence in [a new cultural community]."[4]   Despite this statement Seelye excludes certain items which would indicate cross-cultural awareness: those items which may accurately describe target culture patterns but which natives of that target culture do not recognize.   He makes the aim of his test even more obscure by posing the rhetorical question,   "Would [Americans] survive exposure to

---

[3]The term, cultural pattern, is used here in the same sense as by Robert Lado in *Language Testing* (London: Longmans, 1961), pp. 275-289.

[4]Seelye, p. 77.

[the Guatemalan culture]?"[5]   Because his testing aims are un-
clear, it is not possible to evaluate Seelye's item selection, test
construction and validation procedures. There is no other sound
basis for such evaluation.   One may, however, assume the
broadest possible aims for cross-cultural testing and note what
procedures they would imply.   Then it will become possible to
determine which purposes may be satisfied by Seelye's pro-
cedures (insofar as they have been reported), and what different
or additional procedures would have to be followed by someone
who wished to satisfy other aims.

Four aims of testing are generally recognized:

1. The test user wishes to determine how an individual
   would perform at present in a given universe of situa-
   tions of which the test situation constitutes a sample.
2. The test user wishes to predict an individual's future
   performance (on the test or some external variable).
3. The test user wishes to estimate an individual's present
   status on some variable external to the test.
4. The test user wishes to infer the degree to which the
   individual possesses some trait or quality presumed to
   be reflected in the test performance.[6]

With respect to any one instrument these four aims are not
mutually exclusive.   That is, users may put the same test to
each of these uses, and may have varying degrees of success
according to their different purposes.

These relatively abstract statements of testing aims can
be translated into aims for cross-cultural testing, and implied
prescriptions for test construction can be deduced:

5. The test user wishes to determine how an individual
   would behave and what he would understand in a new
   culture by noting his understanding and behavior in a
   sample of situations from that culture. The test maker
   would collect a representative sample of situations to
   present to the individual so that the individual's under-
   standing and behavior could be observed.

---

[5]*ibid*, p. 78. Although this question was posed in a somewhat different context, it
indicates that Seelye did have in mind a need for cross-cultural understanding which is
consistent with (if not actually identical to) the need outlined in the first paragraphs above.
Hence, one could conclude that he should accept the same testing purpose as that stated
earlier in this paper.

[6]"Technical Recommendations for Psychological Tests and Diagnostic Techniques,"
Supplement to the *Psychological Bulletin*, LI, 2 (1954), p. 13.

6. The test user wishes to estimate from a test score how well an individual will be able in the future to understand and behave appropriately in a target culture community.   The test maker will assemble test items (of any sort) which he can demonstrate to predict such future understanding and behavior.

7. The test user wishes to estimate from a test score how well an individual is able at the time of testing to understand and behave appropriately in a target culture community. The test maker will assemble items which he can demonstrate to indicate such current under-standing and behavior.

8. The test user believes that there is some trait or quality called "cultural awareness" (Any other name will do.) which underlies an individual's ability to com-municate and interact in a foreign culture community, and he wishes to estimate the amount of this trait an individual possesses.   The task of the test maker is especially complex; he must not only determine what the trait called "cultural awareness" is, but must also demonstrate its relevance to effective communication and interaction.[7]

In three cases (5, 6, and 7 above), the test user is concerned with the ability of an individual to function effectively in a large universe of cultural situations.   In the fourth case (8 above), the user is *additionally* interested in the trait which underlies, or is necessary to, that ability.

Neither Lado nor Seelye presents a theory of a trait or process underlying the ability to function in a foreign culture. They do not suggest, for example, that "rigidity" is a human trait which underlies this and other abilities, i.e., that the rigid personality cannot adapt to new cultural ways; and they do not suggest then developing tests of rigidity, which would include items with no apparent relevance to communicating through the

---

[7]An example in a different domain may be useful here. Teachers are often interested in estimating the intelligence of their students because they believe that intelligence is one trait underlying academic achievement. They do not, however, think that intelligence is the same thing as academic achievement (or even of learning ability). They suppose instead that intelligence underlies many behavioral abilities, and that other traits (e.g., motivation, ego-strength, social adjustment, etc.) are significant to academic achievement. The psy-chologists who construct intelligence tests are constantly striving to refine their concept of intelligence, and at the same time they are obliged to demonstrate the relevance of in-telligence for behavior--such as academic achievement.

use of foreign culture patterns.[8]   Both Lado and Seelye seem
to assume that, in the absence of such a theory of underlying
traits or processes, the best place to look for test items is in
the universe of situations one is ultimately (although not exclu-
sively nor necessarily even predominantly) concerned with.   This
appears to be a most reasonable assumption to make.

Before describing the universe of situations—item sources—
from which a cross-cultural test might sample, it is convenient
to make one simplifying assumption; and it becomes necessary
then to make at least one more.   Without these assumptions it
is necessary to include all cultural patterns in the universe of
items.   The first simplifying assumption is consistent with the
view taken by Lado and Seelye, *viz.,* in some instances the
"cultural stranger" will understand and behave correctly in the
foreign culture community by virtue of similarities and identi-
ties between his own and the target cultures.   But if this as-
sumption about generalizations of cultural patterns is to reduce
the universe of items (and, therefore, to be truly simplifying),
it becomes necessary also to assume that the "cultural stranger"
will not *learn* to do anything wrong or to misunderstand any-
thing in the target culture.[9]   The meaning of these assumptions
can perhaps be best clarified by a hypothetical example which
illustrates something which they would preclude.   In both cul-
ture A and culture B (the target culture) it is customary first
to present gentlemen to ladies when making introductions.   In
culture A gentlemen follow ladies through doors, and in culture
B they preceed them.   The first simplifying assumption says
that the pattern of introduction in the target culture need not
be tested because it is identical in form to that of culture A
but that the pattern of "who-goes-first" should be.   The second
simplifying assumption says that even if someone from culture
A sees men preceeding ladies in culture B, he will not gener-
alize this to a concept of "men-and-women-are-treated-differ-
ently" which would lead him to alter his customary pattern of
introductions—he could not change his native pattern and pre-
sent ladies to gentlemen first.[10]

---

[8]Such an apparently irrelevant item might be to require the examinees to recite their
alphabet backwards in, say, five seconds. The item could be scored according to the number
of letters which were recited in the correct reverse order within the time allowed.

[9]Arguments against making this assumption with respect to foreign language testing
may be found in: J.A. Upshur, "Language Proficiency Testing and the Contrastive Analysis
Dilemma," *Language Learning,* XII, 2 (1962), 123-127

[10]This assumption could be weaker than is stated here.   It might be of the form:
cultural mislearning is so infrequent and the results so inconsequential that it is an unimpor-
tant consideration in test construction.

Because a cross-cultural test should provide information on an individual's ability both to understand the foreign cultural patterns and to behave so that his intended meanings are understood, a description of the universe of test items must include: (1) the set of observed behaviors (or behavioral products) to be understood, and (2) the set of behavioral patterns to be appropriately performed. In order to apply the simplifying assumptions it is necessary also that the description show which observed cultural patterns will be understood and which will be appropriately performed.   The form of a description which satisfies these conditions (identifying the understanding and performance domains, and distinguishing between what must be and what need not be learned) can be shown.

The universe of test items is divided into two categories, the understanding and performance domains, as in Figure I. The same patterns which must be understood are those which are used to communicate meaning.   This is a necessary division because it cannot be taken for granted that an individual's understanding and performance of cultural behavior will be identical—just as it cannot be taken for granted that an individual will have the same difficulties in comprehension and production of foreign language behavior.

Universe of
Test Items

Observed patterns
to be understood
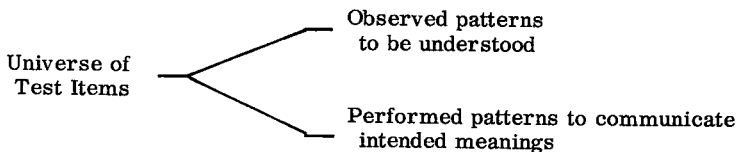
Performed patterns to communicate
intended meanings

Figure I.   *Universe of Items for Cross-Cultural Testing*

In order to apply the two simplifying assumptions, the formal description of the item universe must be extended. Figure II presents the subcategorization of patterns which an individual will observe in the foreign culture and which he must understand.   Figure III presents the subcategorization of the patterns an individual must perform appropriately in order to be understood.

"Familiar" patterns include not only those patterns which are so similar to his own that the cultural stranger would conclude that they are "just like we do it at home," but also those which are similar enough that most cultural strangers (from a single culture) would identify them with patterns of their own
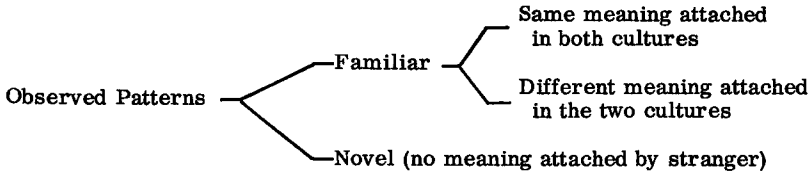
```
                                             ┌── Same meaning attached
                                             │     in both cultures
                              ┌─Familiar ──┤
Observed Patterns ──┤                        └── Different meaning attached
                              │                    in the two cultures
                              └─Novel (no meaning attached by stranger)
```

Figure II. *Subcategorization of Observed Cultural Behavior*

culture. "Novel" patterns are those which are different or am-
biguous enough that cultural strangers cannot attach any mean-
ing or would not agree upon which patterns of their own culture
the observed patterns correspond to. There is probably no *a*
priori way of telling whether any given behavior will seem
"familiar" or "novel." It is more likely the case that this will
have to be an empirical determination.


```
                                      ┌──── Same meaning in
                                      │       both cultures
Performed Patterns ──┤
                                      └──── Different meaning in
                                              the two cultures
```
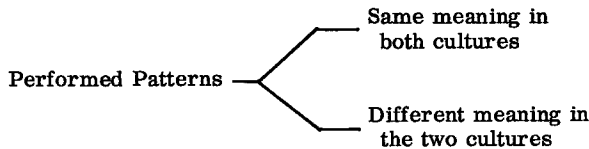
Figure III. *Subcategorization of Performed Behavior Patterns*

Performed patterns with the same meaning in both cultures
are those which would be appropriate in the foreign culture,
i.e., the intended meanings of the cultural stranger would be
communicated. Performed patterns with different meanings
consist of patterns common to both the individual's native cul-
ture and the target culture, but for which each culture has a
different meaning, and consists also of patterns from the native
culture which are not interpretable by members of the foreign
culture community. It is likely true that, similar to the case
of novel observed patterns, there is no *a priori* way of deter-
mining what native patterns will prove uninterpretable to mem-
bers of the target culture.

    The two simplifying assumptions can now be applied to
the description of the universe of cross-culture items. Ob-
served patterns which are familiar to the cultural stranger and
which have the same meanings in both cultures do not have to
be tested; patterns performed by the cultural stranger which

have the same meaning in both cultures do not have to be tested.[11]

Seelye has pointed out that Guatemalans are not always able to give accurate reports of cultural patterns. They might not be able to say what a particular pattern means or to say what pattern would be exhibited in a given situation. Presumably Seelye would agree that the Americans, for whom his test was written, share this inability; he would agree with Lado's general statement that "we are rather helpless to interpret ourselves accurately and to describe what we do."[12] This fact must be taken into account when preparing a cross-cultural test based upon people's reports. Seelye has taken this into account, and the results of the particular way in which he has chosen to handle the problem will be pointed out later.

The description of the universe of cross-cultural items can be extended to account for the inability of Guatemalans (G) and Americans (A) to give accurate reports of cultural meaning and behavior. Figure IV shows the additional subcategorization of target culture patterns which accounts for the fact that Guatemalans and Americans are sometimes able and sometimes unable to give accurate reports of the meanings of behavior. Figure V illustrates the ability or inability of Guatemalans and Americans to give accurate reports of what patterns are appropriate to communicate given intended meanings. According to this complete formal description there are eighteen classes of items in the cross-cultural item universe. For convenience of reference they are labeled: (A), (B), (C), . . ., (Q), (R). The simplifying assumptions provide that eight of these classes need not be tested: (A), (B), (C), (D), (K), (L), (M), and (N). Items from the other ten classes should be sampled if one takes the broadest purposes for cross-cultural testing. It should be noted at this point, however, that these classes are not of equal size, and that in actual test construction one should, *ceteris paribus*, select items to test from each class proportionally to class size.

In the construction of his test for Americans in Guatemala Seelye began by trying to discover examples of cultural contrasts.[13] In that way he attempted to exclude from his test

---

[11]It should not be misinterpreted that this formal description accounts for differences in "form" and "meaning" but ignores differences in "distribution." The same overt behavior segment might be an item in more than one subcategory depending upon its larger context or upon the meaning intended by the performer in the context in which he finds himself.

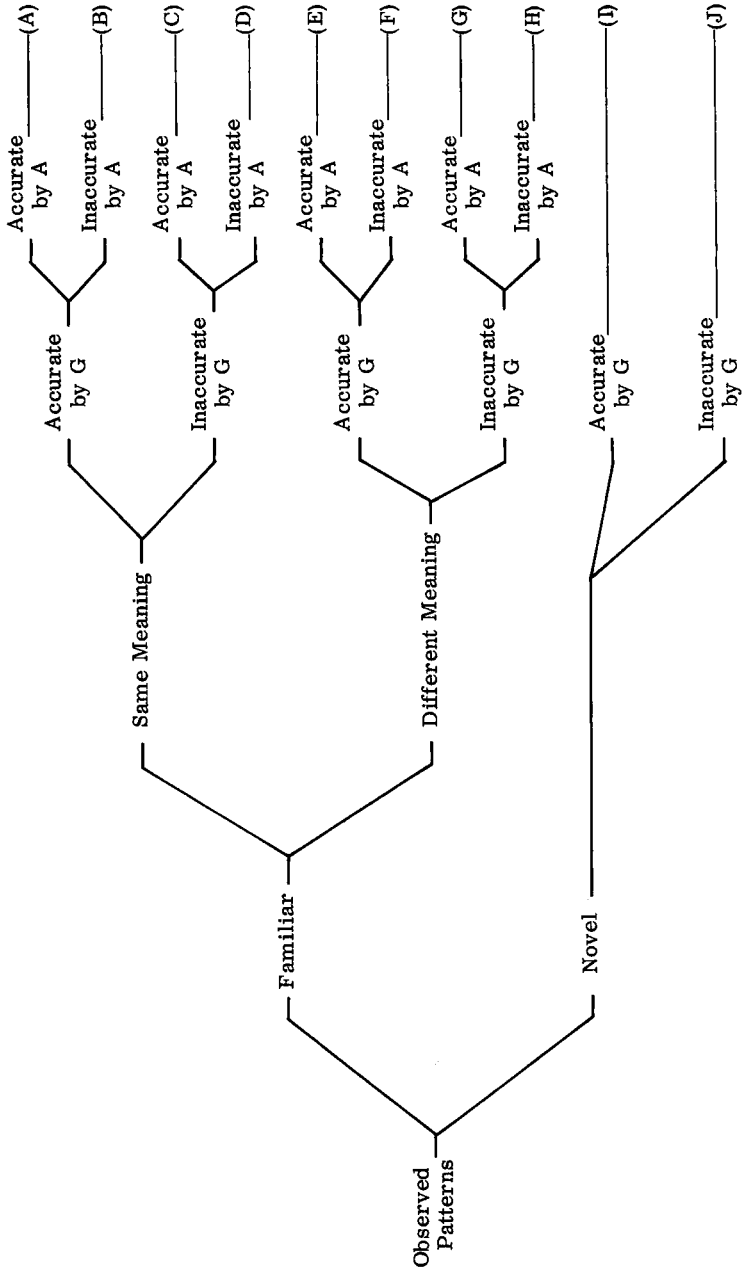[12]Lado, p. 278.

[13]Seelye, p. 78.

Figure IV.   *Description of Observed Pattern Domain for Guatemalan Culture Test Items*

Accurate ———(K)
by A

Inaccurate ———(L)
by A

Accurate ———(M)
by A

Inaccurate ———(N)
by A

Accurate ———(O)
by A

Inaccurate ———(P)
by A

Accurate ———(Q)
by A

Inaccurate ———(R)
by A

Accurate
by G

Inaccurate
by G

Accurate
by G

Inaccurate
by G

Same Meaning
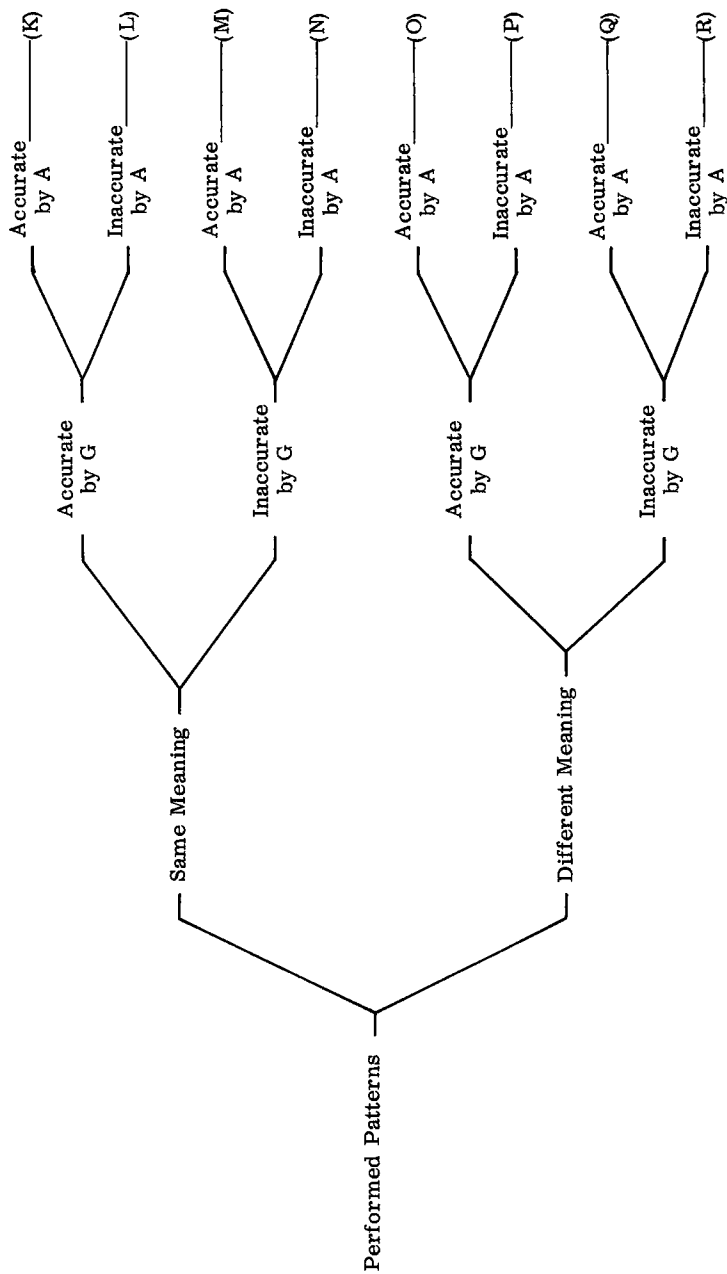
Different Meaning

Performed Patterns

Figure V.   *Description of Performed Pattern Domain for Guatemalan Culture Test*

items which would be properly included in classes (A) - (D) and
(K) - (N). The criterion which he adopts for a cultural con-
trast is, however, a Guatemalan pattern "which contrasts ... to
the pattern which Americans *recognize as their own.*"[14]   This
criterion causes the invalid inclusion of the items of classes
(B) and (L). It further causes the invalid exclusion of any items
in the classes (F) and (P) for which the American "ideal" re-
port coincides with the Guatemalan accurate report.   Seelye
further restricted the items which he would test by excluding
those classes in which Guatemalans do not give accurate reports
of meanings or appropriate behavior: the additional classes
(G), (H), (J), (Q), and (R).[15]   His resulting test includes all
items in two of the eight classes which should have been ex-
cluded; it includes all of the items in only three of the ten
classes which should have been included; it includes a portion
of the items in two of the seven other classes which should
have been included.   The content of the test as compared with
the theoretically ideal cross-culture test is shown in Table I.

It is now apparent that a cross-culture test constructed
along the same lines as Seelye's has limited usefulness. Less
than half of the desired item classes are sampled,[16] and al-
most half as many unwanted classes of items are sampled for
inclusion in the test.[17]   The test might in fact be quite useful

---

[14]Seelye, p. 77, emphasis added.

[15]The only reason he states for making this restriction is that it would be "ironic"
for Americans to be able to report accurately items of Guatemalan culture which Guatemalans
do not (p. 77). Whether or not irony is a sufficiently good reason for making such an exclu-
sion, it is difficult to find anything ironic in such a state of affairs. It doesn't defy credu-
lity or propriety to imagine that an American can learn the meaning of a Guatemalan pattern
rather than its Guatemalan stereotype, and can behave in accordance with what he knows.
Actually the Guatemalan knows the "real meaning" despite his stereotyped inventory
response; otherwise he would not behave as he does behave, and it would be nonsense to
talk about an "ideal" as different from a "real" meaning. The problem can be extrapolated
to the field of foreign language testing, and one would have to conclude, for example, that
if native speakers cannot accurately *describe* some fact of their language, it should not
appear in a foreign language test. In an *ad hoc* "experiment," which is more illustrative
than conclusive, four native speakers of English were asked how they could tell the dif-
ference between *can* and *can't* in speech.   All answered that the presence of /t/ makes the
difference. This was an obviously incorrect report, yet all can distinguish between, "I can
take it," and, "I can't take it."   It is also important for learners of English to be able to
hear the difference. And it is not unthinkable that an English teacher would tell his stu-
dents what to listen for. If Seelye's argument leads to any significant conclusion, it is
probably that a different type of item is sometimes needed.

[16]Actually half (5) of the classes are sampled, but only a part of two of these five
are sampled.

[17]In fact, there will also be included some items which properly belong in (M) and
(N), classes which should not be included. As pointed out above, it is not always possible

Table I. *Makeup of American-Guatemalan Cross-cultural Test*

Classes of Items to be Included	Classes of Items to be Excluded	Classes Actually Included	Classes Included in Part
E		E	
F			F
G			
H			
I		I	
J			
O		O	
P			P
Q			
R			
	A		
	B	B	
	C		
	D		
	K		
	L	L	
	M		
	N		

for purposes 6 and 7 above.   This cannot, however, be taken for granted; it must be demonstrated by comparing test scores with measures of the effectiveness of functioning in the foreign culture.   And although such a test might be useful for these purposes, it is highly likely that a less restricted set of items would yield a more valid, and therefore more useful, test.

Seelye's paper described a number of problems encountered in cross-cultural testing, and as a "case study" it has pointed towards other problems for which solutions should be found. One is the problem of testing methods; techniques which remove limitations upon the classes of cross-cultural items which can be tested must be developed.   Seelye's paper illustrated what the limitations of the testing method which he chose may be.   A second problem is that of determining (preferably by empirical means) whether the second simplifying assumption made by Seelye and Lado, and accepted in this paper, is indeed a valid one.   If it should turn out not to be, the universe of cross-cultural items to be tested is radically altered.   If,

---

(footnote 17 continued)

to predict when patterns of one culture will be generalized to patterns of another. Pretesting would not demonstrate that (M) and (N) items are *not in fact items* from (O) and (P).

however, it proves valid, another problem arises—that of deter-
mining, for pairs of cultures, those patterns which are viewed
by members of both cultures as the same. Another problem
must be solved before it is possible to learn how well a cross-
cultural test measures the ability of cultural strangers to func-
tion effectively in a foreign culture environment. This problem
is to develop measures of non-linguistic communication and
interaction. The final problem alluded to is that of developing
a theory, which can be used by test writers, of the processes
or traits underlying effective communication and interaction in
a foreign language culture.

Seelye's paper raised a number of questions, and answered
many of them. This paper has only raised more questions.
Perhaps the next writer will supply some more answers.