Challenges and Solutions for Testing Less Commonly Taught Languages

ECOLT 2007

Nature of Language

❖ Requirements driven by real world scenarios:

   ➢ Support to US troops

   ➢ Diplomatic efforts

   ➢ Law enforcement

Resource Issues

❖ Defining the requirement

❖ Funding

❖ Available staff

❖ Breadth of requirement

New Challenges

❖ Code Switching

❖ Diglossia

❖ Heritage languages

❖ Non-national languages

❖ Literacy issues

Test developer qualifications

❖ Number of speakers of the language

❖ Familiarity with proficiency testing

❖ Experience in language pedagogy and testing

Testing Receptive Skills

Validation populations

For LCTLs, often we cannot find enough people in the target test population (native speakers of English learning the foreign language in a formal instructional setting) to conduct a large-scale validation. To make up the numbers, we add 4 possible groups of people:

- ➢ Heritage speakers
- ➢ Native speakers
- ➢ "Street learners" (for example, soldiers who have spent time in Iraq and interacted with locals)
- ➢ Learners of related languages

Validation populations

Each population is different from the target population in important ways.

➢ Heritage speakers often lack cultural understanding, and many are illiterate

➢ Native speakers may have trouble understanding English questions and answers/producing answers in English

Validation populations

❖ To mitigate difficulties, be aware of different types of examinees and know whom you're testing: be able to analyze responses group by group.

❖ Alternatives to traditional validation: Angoff method to establish provisional calibrations, then confirm with ongoing administration.

Uneducated developers

If target-language test developers have limited English and/or lack formal education in the language (as may be the case for languages that have no educational system), the testing professionals managing the project must do more:

- ➢ Ask constant and detailed questions

- ➢ Work from translations

- ➢ Go from word-for-word glosses of TL text

Level of language used

Many LCTLs that are not national languages are used only for routine domestic purposes, with a language of wider communication used for higher-level discourse (Arabic dialects and MSA, Philippine languages and English/Spanish)
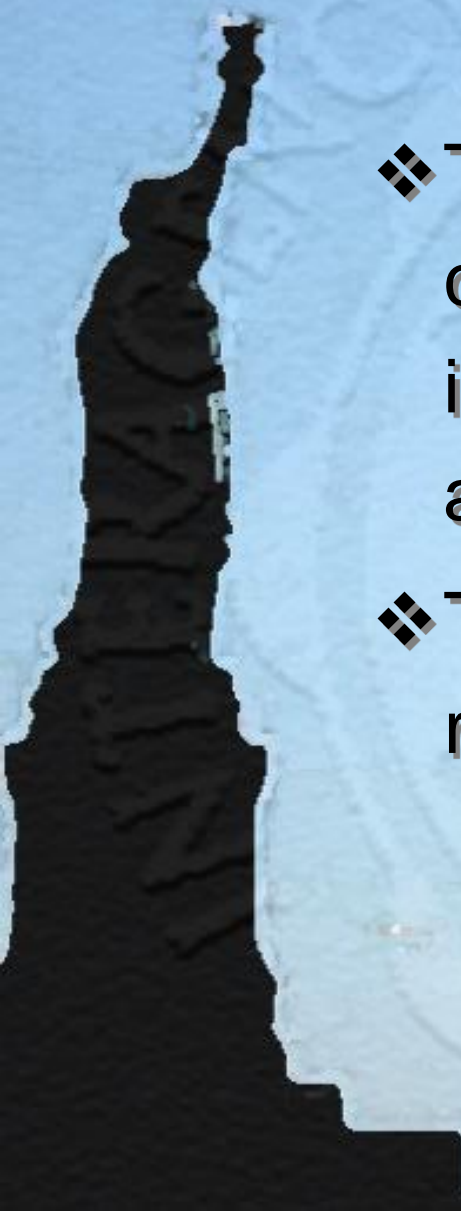
➤ Examine what levels actually need to be tested.

➤ If a testing program usually goes up to level 3, but users of the LCTL typically do not use the language beyond level 2, use a testing format that allows flexibility in which levels are tested.

➤ Don't take level for granted.

Dialect variation

Unlike commonly-tested languages that have a literary standard, test developers cannot assume that any text in "language X" is an appropriate sample of the language.

> Do clients need (for example) Peshawari Pashto, Afghani Pashto, or both?

> Are the target-language test developers aware of dialect differences?

> Do they know the target dialect?

Dialect variation

❖ Test plans need to be very explicit about client needs; specifications and public information documents need to be clear about what dialect is being tested.

❖ TL test developers need to understand the right dialect and know the issues.

Language change

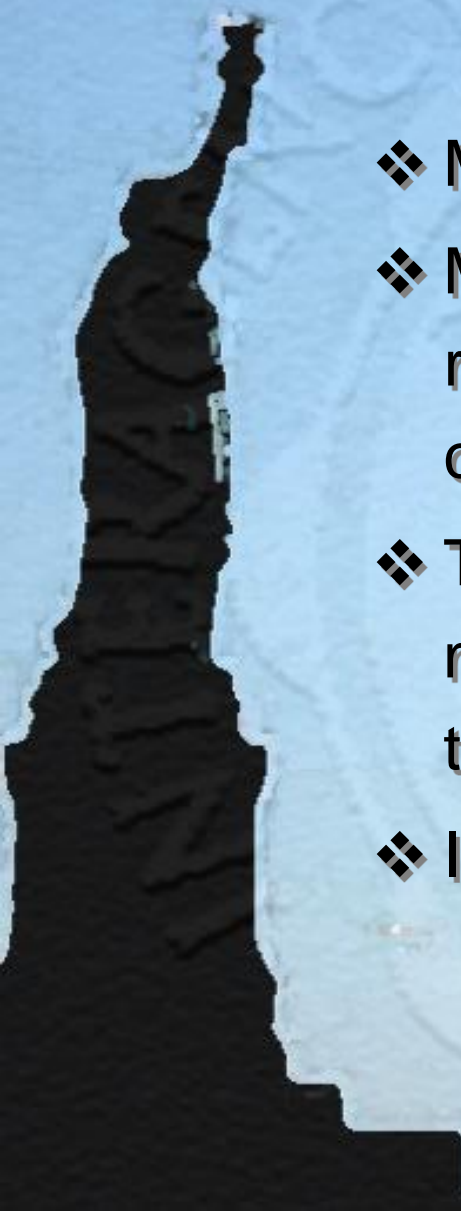Since LCTLs often change rapidly, tests cannot just be developed and left out there.

➢ TL test developers need to be aware of language change issues and willing to accept the fact of language change.

➢ Review tests frequently

➢ Have the capacity to build new items quickly

➢ Use a test format that allows replacement of individual items without re-calibrating the whole pool

Script issues

Some LCTLs, crossing national boundaries, use various script systems (Serbian/Croatian, Kurdish). Do clients require:

➢ Knowledge of at least one script, but not necessarily all?

➢ Knowledge of all scripts?

➢ Can passages in one script be reasonably transcribed into another?

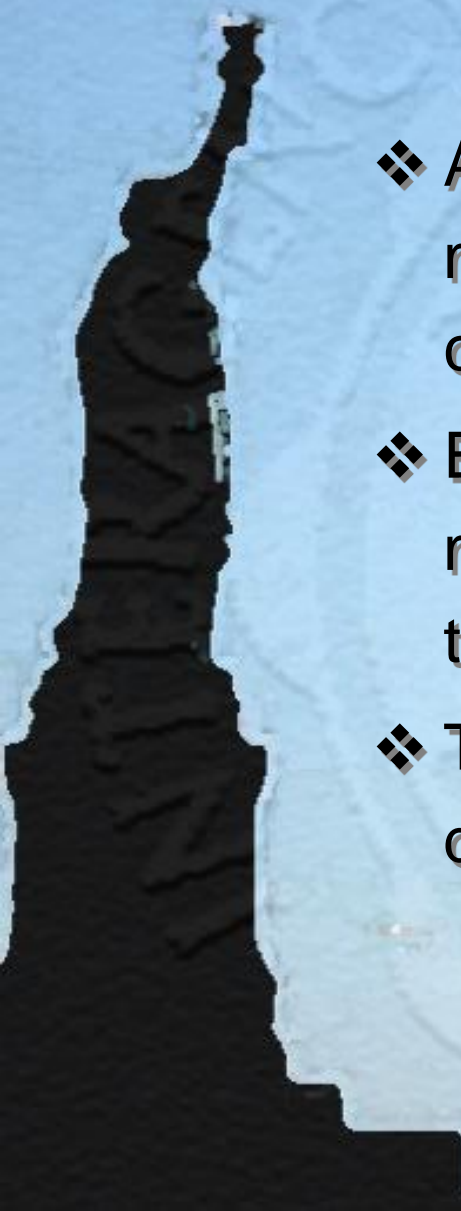➢ Does the script use a font that is not readily available?

Finding materials

❖ Materials in LCTLs are often scarce or unreliable.

❖ Media may be from a diaspora population not representative of the language as it is used in-country.

❖ TL test developers in the US may have spent so many years here that they are not in touch with the language as it is used today.

❖ Internet media may not exist.

Finding materials

❖ Having TL test developers purpose-write passages may help in some cases, but care should be taken that the language feels authentic.

❖ Be aware of authenticity/change issues.

❖ Try to use a variety of diaspora sources, if diaspora sources are the only ones available (US and European).

Diglossia

❖ At low levels, in order to test the specific dialect, it may be necessary to test details and cultural content not usually tested at these levels.

❖ Example: Arabic, if test population knows MSA, most passages in dialect also include MSA. Need to test the MSA parts.

❖ Topics must be especially varied to maximize dialect usage patterns.

Constructed Response Tests

Given problems finding large enough populations for thorough item analysis and calibration, it may be preferable to use constructed-response tests, which are somewhat more direct and flexible (protocols can be adjusted to accommodate novel examinee responses).

Testing Speaking Through Oral Proficiency Interviews

First Languages Tested in the USG

- Arabic
- Cantonese
- Farsi
- French
- German
- Hebrew
- Italian
- Japanese
- Korean
- Mandarin
- Russian
- Spanish
- Vietnamese
- Others…

Fairly Common Languages

❖Armenian, Eastern and Western

❖Dari

❖Hindi

❖Indonesian

❖Pashto

❖Kurdish, Sorani and Kurmanji Serbo-Croatian (now Serbian, Croatian, Bosnian)

❖Uzbek

❖Urdu

❖Many others

Surge Languages

- ❖Baluchi
- ❖Chechen
- ❖Chamorro
- ❖Ga
- ❖Malayalam
- ❖Sindhi
- ❖Tausug
- ❖Twi
- ❖Waray-Waray
- ❖Many others

Oral Proficiency Interview

❖ A.k.a. Speaking Proficiency Test (SPT)

❖ Proficiency, not achievement

❖ 15 - 45 minutes long

❖ 2 testers

❖ Face-to-face or via telephone

❖ 3 sections: warm-up, core, wind-down
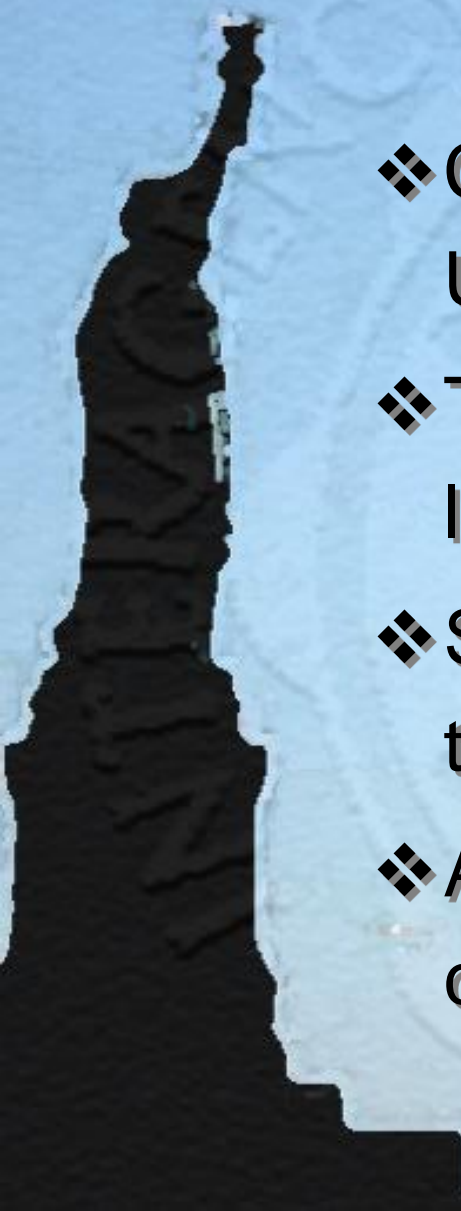
❖ Task-based elicitation of a ratable sample

USG Speaking Testing

❖ Number of speaking tests per year administered by DLI, FBI, and FSI combined?
  ➢ Over 12,000

❖ Number of spoken languages that can be tested by these 3 agencies?
  ➢ Over 100

❖ Number of combined speaking testers that need to be trained, renormed, and checked for quality control?
  ➢ Over 850

Surge Language Issues

❖ Tester Recruitment

❖ Tester Qualifications

❖ Tester Training

❖ Test Administration

❖ Test Scoring

Tester Recruitment Issues

❖ Choice of languages is normally based on USG or national security needs.

❖ Time to find the necessary resources is limited.

❖ Speakers of test language may be difficult to find.

❖ Agencies differ in what type of people they can hire.

Tester Recruitment Solutions

❖ Tester recruiters can double check the urgency of a situation.

❖ Testing organizations can check:
  ➢ testing resources at other agencies;
  ➢ personnel records of employees;
  ➢ language communities in the country;
  ➢ professors or language professionals.

❖ Agencies can require a low-level clearance and not reveal any sensitive information.

Tester Qualification Issues

- ❖ No language teaching or testing experience
- ❖ Background of unrelated skills or professions
- ❖ With or without any academic degrees
- ❖ Low levels of English
- ❖ Language has changed since the tester lived in country
- ❖ Target language is rusty
- ❖ Language of the tester varies from the examinee because of social differences
- ❖ Native or heritage speaker testers who have lower speaking skills than their examinees
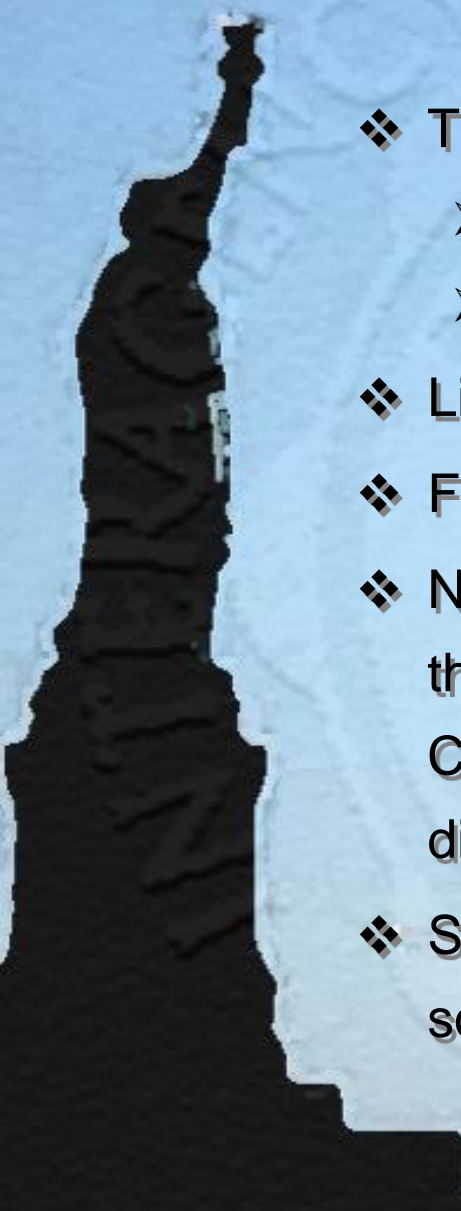
Tester Qualification Solutions

❖ Personalized training, quality control, and more training

❖ Educate trainers in the language

❖ Interpreter/co-trainer in a similar language
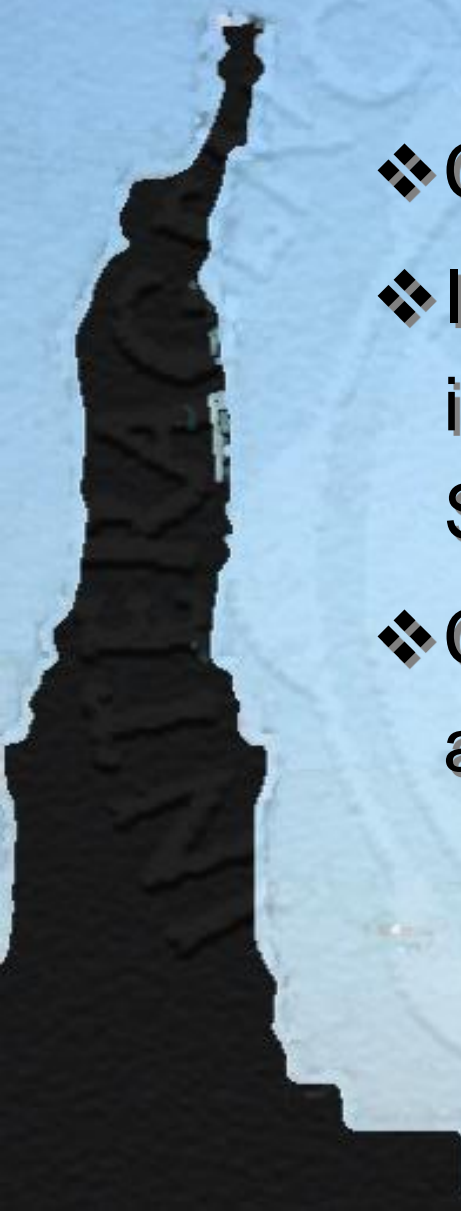
❖ Require language testing for all testers

First Testers in a Language

❖ Experience in the language
❖ Information about time spent in country and how the tester used the language there
➢ Used it in professional contexts (e.g. business transactions, lectures)
➢ Used it daily with family, friends, colleagues, etc.
❖ Previous tests taken in other skills in the language
❖ Authority in the field
❖ Certifications
❖ Awards
❖ Publications

Tester Training Issues

❖ Tester qualification issues become tester training issues

➢ Little conscious understanding of how the test language works

➢ Limited English proficiency

❖ Limited time to conduct training

❖ Few opportunities to conduct practice tests

❖ Necessity to create standardized proficiency tests across the languages, with emphasis on the Middle Eastern, Central and Southeastern Asian languages and their dialects

❖ Specifics of interpreting the ILR Skill Level Descriptions for some LCTL languages

Tester Training Solutions

❖ Continued training after the first test

❖ Intensive training in understanding and interpreting the ILR descriptions for Speaking

❖ Continued collaboration across USG agencies

Test Administration Issues

❖ Sociolinguistic/cultural issues in the target language
  ➢ "taboo" topics
  ➢ gender bias
  ➢ women testing men
  ➢ age and seniority
  ➢ cultural appropriateness of speaking tasks and role-plays

❖ Language interference
  ➢ use of English
  ➢ insertion of words from other dialects or languages of the regions

Test Administration Solutions

❖ Individual refresher training before actual testing sessions

➢ Language-specific strategies and language interference issues discussed

➢ 'Guides' provided before actual tests

❖ To avoid non-Target Language (TL):

➢ Reminder to use TL

➢ Feign lack of non-TL understanding

➢ Ask what that word/phrase would be in TL

➢ Both testers and examinees use circumlocution to overcome communication issues

Testing Speaking: Adjustments to Standardized Procedures

❖ 'Guides' participate in or monitor test administration

❖ Testing aids provided
  ➢ cheat-sheets
  ➢ visual training aids
  ➢ preludes

❖ Silent communication between testers and their co-testers or 'guides'

❖ One brief pause taken to regroup

❖ Guides also assist with scoring

Testing "on the fly"

❖ Written instructions sent to the tester in advance

❖ 20-minute briefing provided before the test

❖ Communication system established for instructions during the test

❖ "Cheat-sheet" used by examiner to direct with the tester

Test Scoring Issues

❖ Inter-language contamination during the interviews

❖ Determining the maximum ILR score possible in a language

❖ Debriefing the tester

Test Scoring Solutions

❖ Careful and detailed explanation of what went on during the exam

❖ Using only highly trained examiners

❖ Careful records of issues in the language to be consulted in the future

❖ Third party reviews

Looking Forward

❖ Collaboration with testing organizations

 ➢ Within the USG

 ➢ Across the USA

 ➢ Around the globe

❖ Proactive training of testers in surge languages when possible

ILR Testing Committee, ECOLT Presentation

Christina Hoffman, Foreign Service Institute,

Beth Mackey, Department of Defense,

Rachel Lunde Brooks, Federal Bureau of Investigation

Mika Hoffman, Defense Language Institute Foreign Language Center

Anna Hardy, Monika Ihlenfeld and Suzanna Gaevsky DLIFLC

Alan Legowik, Defense Intelligence Agency

Paul Tucker, Language Learning Services

Meg Malone, Center for Applied Linguistics

Ben Thomas, Office of the Director of National Intelligence

With special thanks to Ray Clifford, BYU, who served as our Moderator.