Reading and Listening Proficiency Test Development for LCTLs
Arabic, Hebrew, Persian and Turkish as a Case Study

Interagency Language Roundtable
Washington, D.C.
9 April 2010
NMELRC Leadership

**Kirk Belnap**, Director, BYU  
**Maggie Nassif**, Administrative Director, BYU

<table>
<thead>
<tr>
<th>Mahmoud Al-Batal</th>
<th>Erika Gilson</th>
<th>Shmuel Bolozky</th>
</tr>
</thead>
<tbody>
<tr>
<td>UT - Austin</td>
<td>Princeton</td>
<td>Massachusetts - Amherst</td>
</tr>
<tr>
<td>Associate Director</td>
<td>Associate Director</td>
<td>Associate Director</td>
</tr>
<tr>
<td>Professional Development</td>
<td>Language Assessment</td>
<td>Pathways to Proficiency</td>
</tr>
</tbody>
</table>

**Kirk Belnap**

- Ruth Adler Ben-Yehuda, Brown  
- Kristen Brustad, UT Austin  
- Muhammad Eissa, Chicago  
- Suzan Oezel, Indiana  
- Vardit Ringvald, Brandeis  
- Martha Schulte-Nafeh, UT Austin  
- Dwight Stephens, Duke

**Erika Gilson**

- Roger Allen, Penn  
- Mahdi Alos, USMA  
- Micheline Chalhoub-Deville, UNC Greensboro  
- Nihan Ketrez, Yale  
- Salah-Dine Hammoud, USAFA  
- Roberta Micallef, Boston  
- Vardit Ringvald, Brandeis  
- Martha Schulte-Nafeh, UT Austin  
- Kamran Talattof, Arizona

**Shmuel Bolozky**

- Benjamin Hary, Emory  
- Ahmad Karimi-Hakkak, Maryland  
- Sylvia W. Onder, Georgetown  
- Vardit Ringvald, Brandeis  
- Muhammad Eissa, Chicago  
- Kamran Talattof, Arizona
Mandate for Title VI Language Resource Centers

“Improve the Nation’s Capacity to Teach and Learn Foreign Languages Effectively”
LRC Priorities
(according to Title VI Legislation)

• Research
• Materials Development/Dissemination
• Performance Testing
• Teacher Training
• Assess LCTL Needs, Develop Action Plans
• K-12
• Advanced Summer Intensive Programs
NMELRC Mission

• reach more students
• increase quality of learning opportunities for all students
“Assess LCTL Needs, Develop Action Plans”

• surveys of students, teachers, administrators
• site visits, telephone interviews
• study of hiring/staffing practices, implications
• collect outcomes data from language programs and funding agencies
How are we doing in terms of outcomes?
**Critical Languages Scholarship Program**

<table>
<thead>
<tr>
<th>Language</th>
<th>Level</th>
<th>2008 applied</th>
<th>2008 awarded</th>
<th>2009 applied</th>
<th>2009 awarded</th>
<th>2010 applied</th>
<th>2010 awarded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic</td>
<td>Beginning</td>
<td>1345</td>
<td>73</td>
<td>1495</td>
<td>73</td>
<td>694*</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>Intermed.</td>
<td>759</td>
<td>65</td>
<td>872</td>
<td>65</td>
<td>561</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>234</td>
<td>65</td>
<td>388</td>
<td>55</td>
<td>264</td>
<td>60</td>
</tr>
<tr>
<td>Persian</td>
<td>Intermed.</td>
<td>60</td>
<td>8</td>
<td>77</td>
<td>8</td>
<td>87</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>17</td>
<td>7</td>
<td>29</td>
<td>7</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Turkish</td>
<td>Beginning</td>
<td>293</td>
<td>17</td>
<td>279</td>
<td>25</td>
<td>409</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Intermed.</td>
<td>73</td>
<td>19</td>
<td>102</td>
<td>17</td>
<td>51</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>28</td>
<td>14</td>
<td>32</td>
<td>10</td>
<td>21</td>
<td>15</td>
</tr>
</tbody>
</table>

*Starting in 2010, "beginning" requires applicants to have taken one year of college-level Arabic or the equivalent prior to the beginning of the summer institute.*
Case Study

CASA
(The Center for Arabic Study Abroad)
## CASA Applications and Fellowships

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Applications</th>
<th>Summer-Only Fellowships</th>
<th>Full-Year Fellowships</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000-2001</td>
<td>42</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>2001-2002</td>
<td>54</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>2002-2003</td>
<td>50</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>2003-2004</td>
<td>81</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>2004-2005</td>
<td>99</td>
<td>6</td>
<td>28</td>
</tr>
<tr>
<td>2005-2006</td>
<td>118</td>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td>2006-2007</td>
<td>129</td>
<td>4</td>
<td>36</td>
</tr>
<tr>
<td>2007-2008</td>
<td>127</td>
<td>6</td>
<td>35</td>
</tr>
<tr>
<td>2008-2009</td>
<td>123</td>
<td>6</td>
<td>38</td>
</tr>
<tr>
<td>2009-2010</td>
<td>162</td>
<td>5</td>
<td>42</td>
</tr>
<tr>
<td>2010-2011</td>
<td>192</td>
<td>2</td>
<td>42</td>
</tr>
</tbody>
</table>
This advertisement offers new computers

- **A** for rent with a low security deposit.
- **B** for rent for as little as 43 LE per month.
- **C** for sale with a 2 year warranty plan.
- **D** for sale with a long payment plan.
What is the cow saying?

- A  The cow wants to get his CDs.
- B  The cow wants to sell his CDs.
- C  The cow wants to record his CDs.
- D  The cow wants to pay for the CDs.
This passage says that Tom Roland wanted:

- A to correspond with the queen.
- B to send his balloon to England.
- C his letter to go to another country.
- D to travel in a hot-air balloon.
<table>
<thead>
<tr>
<th>Started</th>
<th>Time</th>
<th>Name</th>
<th>Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 Apr 10</td>
<td>1h 8m</td>
<td>Arab</td>
<td>0</td>
<td>53</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 Apr 10</td>
<td>0h 56m</td>
<td>Arab</td>
<td>0</td>
<td>59</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 Apr 10</td>
<td>0h 49m</td>
<td>Arab</td>
<td>0</td>
<td>47</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 Apr 10</td>
<td>1h 10m</td>
<td>Arab</td>
<td>0</td>
<td>47</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 Apr 10</td>
<td>0h 49m</td>
<td>Arab</td>
<td>0</td>
<td>76</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 Apr 10</td>
<td>0h 40m</td>
<td>Arab</td>
<td>0</td>
<td>71</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 Apr 10</td>
<td>0h 53m</td>
<td>Arab</td>
<td>0</td>
<td>65</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 Apr 10</td>
<td>0h 57m</td>
<td>Arab</td>
<td>0</td>
<td>76</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 Apr 10</td>
<td>0h 57m</td>
<td>Arab</td>
<td>0</td>
<td>41</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 Apr 10</td>
<td>0h 48m</td>
<td>Arab</td>
<td>0</td>
<td>71</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Four Challenges Encountered When Testing Proficiency in the Receptive Skills

Ray Clifford
Four Challenges

1. Defining the proficiency construct to be tested.
2. Developing curriculum-independent proficiency tests.
3. Applying proficiency-based scoring criteria.
4. Reporting proficiency results.
Four Challenges

1. Defining the proficiency construct to be tested.
   a. Language ability is a multidimensional trait.
   b. There are many definitions of “Reading” (and “Listening”) comprehension.

2. Developing curriculum-independent proficiency tests.

3. Applying proficiency-based scoring criteria.

4. Reporting proficiency results.
What is reading?

• How many variables are involved in defining reading?
  – How many are related to the author and the text s/he produced?
  – How many are related to the reader’s skills and purposes?

• Assuming five variables for the author and 5 variables for the reader, how many types of reading are there?
How many types of reading are there?

<table>
<thead>
<tr>
<th>Rating Level</th>
<th>Author</th>
<th>Reader</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Purpose</td>
<td>Topical Domains</td>
</tr>
<tr>
<td>3 / Superior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 / Advanced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 / Intermediate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 / Novice</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- For 10 factors, each with 4 levels there are 40 cells in which a rating may be assigned.
- With one rating per factor, how many different rating profiles are possible?
For 10 factors, how many rating profiles are possible?

<table>
<thead>
<tr>
<th>Rating Level</th>
<th>Author</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Reader</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Purpose</td>
<td>Topical Domains</td>
<td>Genre</td>
<td>Text Type</td>
<td>Accuracy</td>
<td>Purpose</td>
<td>Topical Domains</td>
<td>Type of Reading</td>
<td>Reading Strategy</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Superior</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10 factors with 4 levels produces \(4^{10}\) …
or 1,048,576 possible rating profiles.
How might this unwieldy complexity be made more manageable?

• We can reduce the scoring complexity by aligning the rating factors!

• For instance, it would make sense to align the author “topical domains” with the author “purposes” typically associated with those topics.
For 9 factors, how many rating profiles are possible?

<table>
<thead>
<tr>
<th>Rating Level</th>
<th>Author</th>
<th>Reader</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Purpose</td>
<td>Topical Domains</td>
</tr>
<tr>
<td>Superior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novice</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9 factors with 4 levels produces $4^9 \ldots$

or **262,144** possible rating profiles.
How might this unwieldy complexity be made more manageable?

- We can reduce the scoring complexity by further aligning the rating factors!
- For instance, it would make sense to align the author “genre” with the author “purposes” generally associated with those topics.
- (In fact, we could reapply the “alignment” principle across all factors.)
For 8 factors, how many rating profiles are possible?

<table>
<thead>
<tr>
<th>Rating Level</th>
<th>Author</th>
<th>Reader</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Purpose</td>
<td>Topical Domains</td>
</tr>
<tr>
<td>Superior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novice</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8 factors with 4 levels produces $4^8$ …

or 65,536 possible rating profiles.
For 7 factors, how many rating profiles are possible?

<table>
<thead>
<tr>
<th>Rating Level</th>
<th>Author</th>
<th>Reader</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Purpose</td>
<td>Topical Domains</td>
</tr>
<tr>
<td>Superior</td>
<td></td>
<td>Green</td>
</tr>
<tr>
<td>Advanced</td>
<td></td>
<td>Blue</td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td>Yellow</td>
</tr>
<tr>
<td>Novice</td>
<td></td>
<td>Gray</td>
</tr>
</tbody>
</table>

7 factors with 4 levels produces $4^7$ …

or 16,384 possible rating profiles.
For 6 factors, how many rating profiles are possible?

<table>
<thead>
<tr>
<th>Rating Level</th>
<th>Author</th>
<th>Reader</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Purpose</td>
<td>Topical Domains</td>
</tr>
<tr>
<td>Superior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novice</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6 factors with 4 levels produces $4^6$ …
or 4,096 possible rating profiles.
For 5 factors, how many rating profiles are possible?

<table>
<thead>
<tr>
<th>Rating Level</th>
<th>Purpose</th>
<th>Topical Domains</th>
<th>Genre</th>
<th>Text Type</th>
<th>Accuracy</th>
<th>Purpose</th>
<th>Topical Domains</th>
<th>Type of Reading</th>
<th>Reading Strategy</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5 factors with 4 levels produces $4^5$ …

or 1,024 possible rating profiles.
For 4 factors, how many rating profiles are possible?

<table>
<thead>
<tr>
<th>Rating Level</th>
<th>Author</th>
<th>Reader</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Purpose</td>
<td>Topical Domains</td>
</tr>
<tr>
<td>Superior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novice</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4 factors with 4 levels produces $4^4 = 256$ possible rating profiles.
For 3 factors, how many rating profiles are possible?

3 factors with 4 levels produces $4^3$ … or 64 possible rating profiles.
For 2 factors, how many rating profiles are possible?

2 factors with 4 levels produces $4^2$ … or 16 possible rating profiles.
How Many Rating Profiles are Possible?

<table>
<thead>
<tr>
<th>Rating Level</th>
<th>Author</th>
<th>Reader</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior</td>
<td>Purpose</td>
<td>Purpose</td>
</tr>
<tr>
<td></td>
<td>Topical Domains</td>
<td>Topical Domains</td>
</tr>
<tr>
<td></td>
<td>Genre</td>
<td>Type of Reading</td>
</tr>
<tr>
<td></td>
<td>Text Type</td>
<td>Reading Strategy</td>
</tr>
<tr>
<td></td>
<td>Accuracy</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Advanced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novice</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 factor with 4 levels produces $4^1 = 4$ possible rating profiles.
Benefits of Aligning Reading Factors

• Complexity is reduced.
• Each level becomes a separate “Task, Condition, and Accuracy” ability criterion.
• With aligned criteria:
  – Raters can look for sustained ability.
  – Assign ratings using a floor and ceiling approach.
• This hierarchy of levels defines the construct of what it means to be able to “read”.
What is “reading”?

• For achievement and performance testing, there are a million possible definitions of reading.

• For proficiency testing, reading requires:
  – Automatic comprehension rather than laborious decoding.
  – “Reading to learn” rather than “learning to read”.
  – Understanding of a text for the purpose for which it was written.
  – General, transferable ability beyond one’s own areas of specialization.
Our Definition of Reading

• Proficient reading: The active, automatic process of using one’s internalized language and culture expectancy system to comprehend an authentic text for the purpose for which it was written.

• Author purpose and reading purpose.
  – Orientation – Get necessary information
  – Instruction – Learn
  – Evaluation – Evaluate and synthesize
Four Challenges

1. Defining the proficiency construct to be tested.

2. Developing curriculum-independent proficiency tests.
   a. Communication tasks to be performed.
   b. Contexts and typical topical domains.
   c. Communication accuracy.
   d. Associated text types and genres.

3. Applying proficiency-based scoring criteria.

4. Reporting proficiency results.
The type of learning expected:
There are 3 Types of Learning

A. Direct Application
B. Near Transfer
C. Far Transfer
The 1st Type of Learning

• With direct application learning, students...
  – Memorize and practice specific responses.
  – Focus is on the content of a specific course, textbook, or curriculum.
  – Learn only what is taught.
The 2nd Type of Learning

• With near transfer learning, students…
  – Go beyond rote responses to rehearsed and semi-rehearsed responses.
  – Focus on a predetermined set of tasks or settings.
  – Apply what they learn within a range of familiar, predictable settings.
The 3rd Type of Learning

• With learning for far transfer, students…
  – Develop the ability to transfer what is learned from one context to another.
  – Acquire the knowledge and skills needed to respond spontaneously to new, unknown, or unpredictable situations.
  – Learn how to continue learning and to become independent learners.
The type of test used:

3 Types of Tests

A. Achievement
B. Performance
C. Proficiency
The 1st Type of Test

- **Achievement tests measure:**
  - Practiced, memorized responses.
  - What was taught.
  - The content of a specific textbook or curriculum.
The 2nd Type of Test

• **Performance tests measure:**
  
  – Rehearsed and semi-rehearsed responses.
  
  – Ability to respond in constrained, familiar, and predictable settings.
  
  – Whether learning transfers to similar situations.
The 3rd Type of Test

• **Proficiency tests measure:**
  – Whether skills are transferable to new tasks.
  – Spontaneous, unrehearsed abilities.
  – General ability to accomplish tasks across a wide variety of real-world settings.
Aligning Learning and Testing

- **Direct Application <=> Achievement**
  - Memorized responses using the content of a specific textbook or curriculum.

- **Near Transfer <=> Performance**
  - Rehearsed ability to communicate in specific, familiar settings.

- **Far Transfer <=> Proficiency**
  - Unrehearsed general ability to accomplish real-world communication tasks across a wide range of topics and settings.
### General Proficiency Requires a Transfer of Learning

A By-Level Proficiency Summary with General Text Characteristics and Learning Types

[Text Type] + Green = Far Transfer, Blue = Near Transfer, Red = Direct Application

<table>
<thead>
<tr>
<th>ILR LEVEL</th>
<th>FUNCTION/TASKS</th>
<th>CONTEXT/TOPICS</th>
<th>ACCURACY</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>All expected of an educated NS [Books]</td>
<td>All subjects</td>
<td>Accepted as a well-educated NS</td>
</tr>
<tr>
<td>4</td>
<td>Tailor language, counsel, motivate, persuade, negotiate [Chapters]</td>
<td>Wide range of professional needs</td>
<td>Extensive, precise, and appropriate</td>
</tr>
<tr>
<td>3</td>
<td>Support opinions, hypothesize, explain, deal with unfamiliar topics [Multiple pages]</td>
<td>Practical, abstract, special interests</td>
<td>Errors never interfere with communication &amp; rarely disturb</td>
</tr>
<tr>
<td>2</td>
<td>Narrate, describe, give directions [Multiple paragraphs]</td>
<td>Concrete, real-world, factual</td>
<td>Intelligible even if not used to dealing with non-NS</td>
</tr>
<tr>
<td>1</td>
<td>Q &amp; A, create with the language [Multiple sentences]</td>
<td>Everyday survival</td>
<td>Intelligible with effort or practice</td>
</tr>
<tr>
<td>0</td>
<td>Memorized [Words and Phrases]</td>
<td>Random</td>
<td>Unintelligible</td>
</tr>
</tbody>
</table>
The Major ACTFL Levels

ILR

3

SUPERIOR
Can support opinion, hypothesize, discuss topics concretely and abstractly, and handle a linguistically unfamiliar situation

2

ADVANCED
Can narrate and describe in all major time frames and handle a situation with a complication

1

INTERMEDIATE
Can create with language, ask and answer simple questions on familiar topics, and handle a simple situation or transaction

0

NOVICE
Can communicate minimally with formulaic and rote utterance, lists and phrases

NATIONAL MIDDLE EAST LANGUAGE RESOURCE CENTER
The Phenomenon of Shrinking Educational Expectations

• Students don’t want to waste their time studying what is not going to “needed.”

• For students (and often teachers, parents, and administrators); the tests used and not a course’s stated learning objectives define what is “needed.” Therefore,
  – Limited-scope tests reduce the breadth of learning.
  – Simple tests reduce the level of learning.
Shrinking Educational Expectations
(When only Achievement Tests are Used)

1. High academic goals are set and learner outcomes are defined.
2. Developers include examples of the most important goals in a textbook.
3. Teachers present as much of the textbook as time allows.
4. Students are tested on some of the items taught.
Expanding Educational Expectations (When Proficiency Tests are Added)

1. Set instructional goals and define expected learner outcomes.

Real-world Instructional Domains: cognitive understanding, psychomotor skills, and affective insights.

2a. Course developers sample from the real-world domain areas to create a textbook.

2b. Test developers use an independent sample of the real-world domain areas to create proficiency tests that are not based on the textbook.

3. Teachers adapt text materials to learners’ abilities, diagnose learning difficulties, adjust activities and add supplemental materials to help students apply new knowledge and skills in achievement and constrained performance areas, and then in real-world proficiency settings.

4. Students practice, expand, and then demonstrate their unrehearsed, extemporaneous proficiency across a broad range of real-world settings that are not in the textbook.
Adding Proficiency Expectations and Proficiency Testing Requires:

- Improved assessment literacy for everyone: Teachers, Administrators, Students, and Parents.
- Ongoing communication among stakeholders.
- A tolerance for formative assessment that allows programs to “fail forward.”
- Clearly stated Expected Learner Outcomes.
- Assessment practices that match the ELOs.
Four Challenges

1. Defining the proficiency construct to be tested.

2. Developing curriculum-independent proficiency tests.

3. Applying proficiency-based scoring criteria.
   a. Ratings are Criterion-Referenced rather than Norm-Referenced.
   b. Ratings assigned are non-compensatory.

4. Reporting proficiency results.
## ILR Proficiency Level Summary

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>FUNCTION/TASKS</th>
<th>CONTEXT/TOPICS</th>
<th>ACCURACY</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>All expected of an educated NS</td>
<td>All subjects</td>
<td>Accepted as an educated NS</td>
</tr>
<tr>
<td>4</td>
<td>Tailor language, counsel, motivate, persuade, negotiate</td>
<td>Wide range of professional needs</td>
<td>Extensive, precise, and appropriate</td>
</tr>
<tr>
<td>3</td>
<td>Support opinions, hypothesize, explain, deal with unfamiliar topics</td>
<td>Practical, abstract, special interests</td>
<td>Errors never interfere with communication &amp; rarely disturb</td>
</tr>
<tr>
<td>2</td>
<td>Narrate, describe, give directions</td>
<td>Concrete, real-world, factual</td>
<td>Intelligible even if not used to dealing with non-NS</td>
</tr>
<tr>
<td>1</td>
<td>Q &amp; A, create with the language</td>
<td>Everyday survival</td>
<td>Intelligible with effort or practice</td>
</tr>
<tr>
<td>0</td>
<td>Memorized</td>
<td>Random</td>
<td>Unintelligible</td>
</tr>
</tbody>
</table>
## Oral Proficiency Interview

© ACTFL, Inc., 1999 All rights reserved.

<table>
<thead>
<tr>
<th>Proficiency Level*</th>
<th>Global Tasks and Functions</th>
<th>Context / Content</th>
<th>Accuracy</th>
<th>Text Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior</td>
<td>Discuss topics extensively, support opinions and hypothesize. Deal with a linguistically unfamiliar situation.</td>
<td>Most formal and informal settings / Wide range of general interest topics and some special fields of interest and expertise.</td>
<td>No pattern of errors in basic structures. Errors virtually never interfere with communication or distract the native speaker from the message.</td>
<td>Extended discourse</td>
</tr>
<tr>
<td>Advanced</td>
<td>Narrate and describe in major time frames and deal effectively with an unanticipated complication.</td>
<td>Most informal and some formal settings / Topics of personal and general interest.</td>
<td>Understood without difficulty by speakers unaccustomed to dealing with non-native speakers.</td>
<td>Paragraphs</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Create with language, initiate, maintain, and bring to a close simple conversations by asking and responding to simple questions.</td>
<td>Some informal settings and a limited number of transactional situations / Predictable, familiar topics related to daily activities.</td>
<td>Understood, with some repetition, by speakers accustomed to dealing with non-native speakers.</td>
<td>Discrete sentences</td>
</tr>
<tr>
<td>Novice</td>
<td>Communicate minimally with formulaic and rote utterances, lists and phrases.</td>
<td>Most common informal settings / Most common aspects of daily life.</td>
<td>May be difficult to understand, even for speakers accustomed to dealing with non-native speakers.</td>
<td>Individual words and phrases</td>
</tr>
</tbody>
</table>

[*A rating at any major level is arrived at by the sustained performance of the functions of the level, within the contexts and content areas for that level, with the degree of accuracy described for the level, and in the text type for the level. The performance must be sustained across ALL of the criteria for the level in order to be rated at that level.]*
Applying C-R Rating Criteria Requires the Use of “Fuzzy Logic”

1. Success is defined as a “standard” with task, condition, and accuracy criteria.
2. Base-level ratings are non-compensatory.
3. The hierarchy of level-specific standards forms a “Guttmanesque” scale.
4. Ratings are based on unrehearsed ability to accomplish communication tasks in “real world” settings, rather than on rehearsed abilities in curriculum-dependent situations.
What is Fuzzy Logic?

• A form of multi-valued logic derived from fuzzy set theory. (Latfi Zadeh, 1965).

• The logic used in many real-world, “soft-computing” applications.
  – Robotics.
  – Artificial intelligence.
  – Neural networks.
  – Natural Language Processing.

• Fuzzy logic is the way humans think.
What is Fuzzy (Human) Logic?

• In contrast with binary logic (where sets use 0/1 or “either/or” logic), fuzzy logic variables may have a range of values representing their degree of membership in a given set or class.
  – Binary logic: Is the greenish color blue or yellow?
  – Fuzzy logic: How much blue and how much yellow does the greenish color contain?
What is Fuzzy (Human) Logic?

• In contrast with “probability logic” (where one estimates the likelihood that something is true), fuzzy logic variables represent the degree of membership an item has in a given set or class.
  – Probability: What is the chance that a greenish color is really blue? (Or really yellow?)
  – Fuzzy logic: How much blue and how much yellow does the greenish color contain?
Proficiency Ratings and Fuzzy Logic

• Raters DON’T use binary or probability logic to determine whether an individual meets the standard for a given level.

• Raters DO assess how many of the level-specific proficiency criteria are met.
  – No evidence.
  – Some random evidence.
  – Emerging evidence
  – Evidence of developing skills.
  – Evidence of sustained ability.
With this Profile, What is the Speaker’s Proficiency Level?

- Level 2
- Level 3

Sustained
Developing
Emerging
Random
With this Profile, What is the Speaker’s Proficiency Level?

- Level 2: Sustained
- Level 3: Developing

Emerging
Random
With this Profile, What is the Speaker’s Proficiency Level?

- Sustained
- Developing
- Emerging
- Random

Level 2

Level 3
With this Profile, What is the Speaker’s Proficiency Level?

Level 2
Sustained
Developing
Emerging
Random

Level 3
With this Profile, What is the Speaker’s Proficiency Level?

Level 2
Sustained
Developing
Emerging
Random

Level 3
With this Profile, What is the Speaker’s Proficiency Level?

- Sustained
- Developing
- Emerging
- Random

= Level 2+

Level 2

Level 3
With this Profile, What is the Speaker’s Proficiency Level?

Level 2
- Sustained
- Developing
- Emerging
- Random

Level 3
Human Logic, Ratings, and Total Scores

• 3 Job candidates (Alice, Bob, and Carol) have just taken a reading proficiency test covering Levels 1, 2, and 3.
• All three have different reading abilities.
• All three received the same total score of 195. (65% of a possible 300 points).
• How is that possible?
Criterion-Referenced Scoring Provides the Answer

• Test each proficiency level separately.

• Check for sustained ability at each level.
  – Sustained (consistent evidence) $\approx 67\%$ to $100\%$
  – Developing (a lot, not enough) $\approx 51\%$ to $66\%$
  – Emerging (some evidence) $\approx 26\%$ to $50\%$
  – Random (no evidence) $\approx 0\%$ to $25\%$

• Apply the non-compensatory, “floor” and “ceiling” criteria used in rating speaking skills.
What is Alice’s Reading Proficiency Level?

Total Score: 195 (65%)

- Sustained: 85%
- Developing: 70%
- Emerging: 40%
- Random: 0%

Level 1 | Level 2 | Level 3
What is Alice’s Reading Proficiency Level?

Total Score
195 (65%)

Level 1: 85%
Level 2: 70%
Level 3: 40%

Threshold 2 (Advanced Low)
What is Bob’s Reading Proficiency Level?

Total Score: 195 (65%)

- Sustained: 90%
- Developing: 85%
- Emerging: 20%
- Random
What is Bob’s Reading Proficiency Level?

Total Score
195 (65%)

Level 1: Sustained - 90%
Level 2: Developing - 85%
Level 3: Emerging

Random: 20%

Solid 2 (Advanced Mid)
What is Carol’s Reading Proficiency Level?

Total Score: 195 (65%)

- Sustained: 90%
- Developing: 60%
- Emerging: 45%
- Random:
What is Carol’s Reading Proficiency Level?

Total Score
195 (65%)

Sustained
90%

Developing
60%

Emerging
45%

Random
1+ (Intermediate High)
Four Challenges

1. Defining the proficiency construct to be tested.

2. Developing curriculum-independent proficiency tests.

3. Applying proficiency-based scoring criteria.

4. Reporting proficiency results.
   a. Level of sustained ability.
   b. Status within the “non-sustained” level.
What is Alice’s Reported Reading Proficiency Level?

Total Score: 195 (65%)

- Sustained: 85%
- Developing: 70%
- Emerging: 40%
- Random

Threshold 2 (Advanced Low) with Emerging abilities at Level 3 (Superior)
What is Bob’s Reported Reading Proficiency Level?

Total Score 195 (65%)

Level 1
Sustained 90%
Developing 85%
Emerging 20%
Random

Solid 2 (Advanced Mid) with Random abilities at Level 3 (Superior)
What is Carol’s Reported Reading Proficiency Level?

Total Score
195 (65%)

Level 1
90%

Level 2
60%

Level 3
45%

Sustained
Developing
Emerging
Random

1+ (Intermediate High)
With Developing abilities at Level 2 (Advanced)
Summary
We are Addressing the “Four Challenges” By

1. Defining the proficiency construct to be tested.
2. Developing curriculum-independent, Criterion-Referenced proficiency tests.
3. Applying non-compensatory, proficiency-based scoring criteria.
4. Reporting results for both the sustained level and status at the next higher level.
ACTFL

Building Capacity for Testing
Reading and Listening
Did You Know?

• NATO Benchmark Advisory Tests
  – 4 skills
• Item development and review (ILR)
• Pro-chievement Tests
• AAPPL
  – Interpretive Reading and Listening assessment of academic language
• Arabic Consensus Project
ACTFL Proficiency Guidelines
2010 Revisions

• Reading and Listening
  – First major revision since 1986 publication
  – Addition of sub levels at Advanced
    • Advanced Low
    • Advanced Mid
    • Advanced High

• Paperless distribution via a dedicated web site
• Level-specific descriptions enhanced by sample texts
Welcome to the ACTFL 2010 Proficiency Guidelines project.

The ACTFL Proficiency Guidelines, first published in 1986, and revised again in 1999 and 2001, are undergoing extensive revision. The Proficiency Guidelines 2010 will better describe the current understanding of the major stages of second language growth and will be brought to life in ways not possible before the advent of the Internet. This website is the means by which ACTFL would like to communicate drafts of the revisions to the profession and receive its feedback. It is also the means by which the final version will be made available to the profession. The ACTFL Guidelines will contain not only new text but will be supported by multimedia examples of proficiency and text in various languages along with assessments and a self-study course in text typology. Please explore the site, learn from it, make recommendations. We look forward to hearing from you!

Click to send feedback on the web site
ACTFL 2010 Proficiency Guidelines

At the Advanced Level, readers can understand the main idea and supporting details of longer authentic texts. Readers are able to compensate for limitations in their vocabulary by using contextual clues. Comprehension is supported by greater knowledge of the conventions of the language and at times may even surpass the Advanced Level if the reader is familiar with the subject matter of a text. These readers are beginning to read for learning.

Advanced texts are authentic and consist of connected discourse up to several paragraphs. Their structure is generally clear and predictable. For the most part, the prose is uncomplicated and the subject matter is familiar. There may or may not be evidence of the author's voice.

Advanced Level readers demonstrate a growing independence and an ability to read subject matter that is new to them. They have sufficient control of language to understand sequencing, time frames and chronology.

See below for the three sublevels of Advanced:

**Advanced High**

Able to follow many of the essential points of written discourse at the Superior level in areas of special interest or knowledge. Able to understand parts of texts which are conceptually abstract and linguistically complex, and/or texts which treat unfamiliar topics and situations, as well as some texts which involve aspects of target-language culture. Able to comprehend the facts to begin to make appropriate inferences. An emerging awareness of the aesthetic properties of language and of its literary styles permits comprehension of a wider variety of texts, including literary texts. Misunderstandings may occur.

**Advanced Mid**

Able to read somewhat longer prose of several paragraphs in length, particularly if presented with a clear underlying structure. The prose is predominantly in familiar target language sentence patterns. Reader understands the main ideas and many supporting facts. Comprehension derives not only from situational knowledge and subject matter knowledge but also from increasing control of the language.

**Advanced Low**

Able to read paragraph length prose with a clear underlying structure. The prose is in familiar sentence patterns. The reader understands the main ideas, and some supporting facts. Comprehension derives primarily from situational knowledge and subject matter knowledge.
Sample Advanced Texts

- Obituaries
- Wedding announcements
- News articles
- Business letters
- Instructions
- Reviews
When New Hampshire lawmakers decided to bill negligent hikers for their rescues, they figured they would solve some budget problems and teach hapless tourists a lesson. Then a 17-year-old Eagle Scout got lost on Mount Washington and ended up with a bill for $25,000.

Now New Hampshire officials, facing possible litigation, are defending a law that hasn’t solved their budget troubles and puts the state at odds with national search-and-rescue groups that say billing hikers is dangerous.

“It certainly has put us on the hot spot,” says Lt. Todd Bogardus, head of search and rescue for the state Fish and Game Department.

Few other states bill those who are rescued — and most of the other laws target skiers who stray off marked slopes.

Jack Schwartz, 8, has had seasonal allergy symptoms in the past, but his parents never needed to give him medication.

Some people with allergies should consider immunotherapy in the form of allergy shots.

This year was different. On the way to his baseball league’s opening celebration in Stow, Massachusetts, the second-grader started crying so hard his mother was afraid he’d have a panic attack after pollen blew into his eyes through open car windows.

“He’s eyes were just burning; they were really red,” says his mother, Julie Glovin. Luckily, the family ran into a friend whose son also had seasonal allergies, who recommended antihistamine eye drops for Jack. “Within a minute, he was much better,” Glovin says. Now he’s taking Zyrtec for kids, and so far is doing well.

As spring is the time for getting outside to enjoy blooming flowers and baseball pitches, it’s also a time of misery for allergy sufferers.

This year, some parts of the country are...
**ACTFL 2010 Proficiency Guidelines**

At the Superior Level, readers are able to comprehend texts from many genres dealing with a wide range of subjects, both familiar and unfamiliar. Comprehension is no longer limited to the reader's familiarity with subject matter but also comes from a command of the language that is supported by a broad vocabulary, an understanding of more complex structures and experience with the target culture. Readers at the Superior Level read to learn; they read within the normal range of speed; they can draw inferences from textual and extralinguistic clues.

Superior texts use precise, often technical vocabulary and complex grammatical structures. These texts discuss abstract topics using abstract linguistic formulations. Superior texts often are written for professional audiences, can be argumentative and are frequently linked to cultural norms.

The Superior Level reader reads in many ways like a native. At this level the reader is able to get meaning from lengthy texts of a professional or academic or literary nature, but not fully understand texts that are richly embedded with cultural references and assumptions.
Sample of Superior Texts

News Analyses
- Sample 1
- Sample 2

Op/Ed Pieces
- Sample 1
- Sample 2

Professional Journal Articles
- Sample 1
- Sample 2

Technical Report
- Sample 1

Straightforward Literary Texts
- Sample 1
ACTFL 2010 Proficiency Guidelines

The Power of Posterity By DAVID BROOKS Published: July 27, 2009

Every day, I check a blog called Marginal Revolution, which is famous for its erudite authors, Tyler Cowen and Alex Tabarrok, and its intelligent contributors. Last week, one of those contributors asked a question that is fantastical but thought-provoking: What would happen if a freak solar event

If you take an individualistic view of the world, not much would happen immediately. There are millions of people today who do not reproduce, and they lead happy, fulfilling and productive lives.

Even after the event, material conditions would be exactly the same. People would still have an incentive to go to work, pay off their bills and educate the children who were already with us. For 20 years, there would still be workers flowing into the labor force. Immigrants from the other side of the earth could eventually surge into the areas losing population. If anything, the mass-sterilization might reduce the environmental strain on the planet. People might focus on living for the moment, valuing the here and now.

But, of course, we don’t lead individualistic lives. Material conditions do not drive history. People live in a compact between the dead, the living and the unborn, and the value of the thought experiment is that it reminds us of the power posterity holds over our lives.

If, say, the Western Hemisphere were sterilized, there would soon be a cataclysmic spiritual crisis. Both Judaism and Christianity are promise-centered faiths. They are based on narratives that lead from Genesis through progressive revelation to a glorious culmination.

Believers’ lives have significance because they and their kind are part of this glorious unfolding. Their faith is suffused with expectation and hope. If they were to learn that they were simply a dead end, they would feel that God had forsaken them, that life was without meaning and purpose.

The secular world would be shattered, too. Anything worth doing is the work of generations — ending racism, promoting freedom or building a nation. America’s founders, for example, felt the eyes of their descendants upon them. Alexander Hamilton felt that he was helping to create a great empire. Noah Webster composed his dictionary anticipating that America would someday have 300 million inhabitants, even though at the time it only had 6 million.

These people undertook their grand projects because they were building for their descendants. They were motivated — as ambitious leaders, writers...
ACTFL 2010 Proficiency Guidelines

Advanced

At the Advanced Level, listeners can understand the main ideas and most supporting details of connected discourse. Listening extends to a variety of topics beyond immediate survival needs. Listeners are able to compensate for limitations in their vocabulary and structural knowledge by using real world knowledge. In fact, listeners may understand oral texts above the Advanced Level if they possess significant familiarity with the topic or context. These listeners are beginning to listen to learn.

Advanced Level listeners can process authentic connected speech equivalent in length to a multiparagraph written text. This speech is uncomplicated and generally has a clear and predictable structure.

Advanced Level listeners demonstrate a growing independence and an ability to listen to subject matter that is not very familiar to them. They have sufficient knowledge of language structure to understand basic timeframe references.

Advanced High

Able to follow many of the essential points of oral discourse at the Superior level in areas of special interest or knowledge. Able to understand parts of speech which is conceptually abstract and linguistically complex, and/or texts which treat unfamiliar topics and situations, as well as some speech which involves aspects of target-language culture. Able to comprehend the facts presented in oral discourse and to begin to make appropriate inferences. Nevertheless some misunderstandings may occur.

Advanced Mid

Able to listen to and understand speech of several paragraphs in length, particularly if presented with a clear underlying structure. The speech is predominantly in familiar target language sentence patterns. Listener understands the main ideas and many supporting facts. Comprehension derives not only from situational knowledge and subject matter knowledge, but also from increasing knowledge of the language.

Advanced Low

Able to understand paragraph-length speech with a clear underlying structure. The language is in familiar sentence patterns. The listener understands the main ideas, and some supporting facts. Comprehension derives primarily from situational knowledge and subject matter knowledge.

Advanced Texts

Oprah Winfrey
Policing Report
News Report
Virtual museum tour
Local news report
Superior

At the Superior Level, listeners are able to understand speech in a standard dialect on a wide range of familiar and unfamiliar subjects. They can follow linguistically complex extended discourse such as that found in academic and professional settings, in lectures, speeches and reports. Comprehension is no longer limited to the listener's familiarity with subject matter but also comes from a command of the language that is supported by a broad vocabulary, an understanding of more complex structures and experience with the target culture. They can understand not only what is said, but sometimes what is left unsaid, such as inferences or extralinguistic clues.

Speech at the Superior Level uses precise, often technical vocabulary and complex grammatical structures. This speech deals with abstract topics that are conveyed to professional audiences, can be argumentative and is frequently linked to cultural norms. Listeners at Superior Level speech show some appreciation of idioms, colloquialisms, and register.

The Superior Level listener listens in many ways like a native listener. At this level the listener is able to understand lengthy, formal speech but not necessarily fully understand richly embedded cultural references or assumptions. Listeners at the Superior Level listen to learn.

Superior Texts:
NPR commentary
Ben Bernanke before Congress
Reading of Robert Frost poetry
Point/Counterpoint

Bigthink (Augusten Burroughs on Oprah)
http://www.youtube.com/watch?v= NhLcngE2RCQ&feature=channel

This I Believe (Luis Urueta)
http://www.npr.org/templates/player/mediaPlayer.html?action=1&t=1&list=false&id=103362391&n=103397414
SNL
Weekend Update (Michael Phelps bust)
http://www.youtube.com/watch?v=IHpgZc9FkK8&feature=related

Ricky Gervais
http://www.youtube.com/watch?v=2ITK1ymams&feature=channel

Sir Ian explains Richard III
http://www.youtube.com/watch?v=3cWJshB4yZ8&feature=related

President of Stanford introduces Steve Jobs:
Reading and Listening Test Delivery

- PATH tests developed by NMERC to be delivered from the ACTFL/LTI Test Delivery Platform
- Internet delivered
- Can be delivered individually or bundled in any combination
  - With OPIs, OPIcs and WPTs
  - No minimum or maximum number
- ACTFL ratings assigned based on Revised Proficiency Guidelines
- Machine scored
  - Official ACTFL certificates issued
Next steps

• ACTFL Professional Development
  – Familiarization with Revised Guidelines
  – Workshops in Text Selection and Rating
  – Implications of the Guidelines for Instruction
• Potential for ACE Credit Recommendation
• Adaptive Reading and Listening Tests
  – Self assessment up front
  – Branching
• Development underway in CTL’s
  – English, French, German, Latin, Spanish
• Critical languages are leading the way