Empirical Validation of Reading Proficiency Scales

Results 2012
Ray Clifford & Dana Bourgerie
Prior Research Results: Reading

• Students were able to capture some meaning from all levels of texts, so there is no sequence of text difficulty as implied by the ACTFL [and by ILR] scales. [Allen et al, 1988]

• Text types do not form a hierarchy of increasing difficulty. [Lee and Musimeci, 1998:]
Prior Results: Reading

- There is no empirical evidence to validate the *a priori* [ILR and ACTFL] definitions of levels. [Alderson, 2000.]

- The ACTFL and ILR scales are frequently described as criterion-referenced, however there is a circularity in the descriptors between texts and readers ability. [Brown, 2002]
A Real-World Example

• The following post card was delivered to me a few years ago.
• Are all of the text types used equally difficult?
A Postcard Offer

PHONE-A-GRAM  URGENT-U-REPLY

CONGRATULATIONS:
You will definitely receive one of the free prizes listed below:

1. PORTABLE CASSETTE PLAYER
2. COLOR TV
3. 35MM CAMERA
4. GIFT CERTIFICATE
5. FISHING ROD AND REEL
6. $500 CASH
7. 6 PIECE STEAK KNIFE SET
8. BLENDER
9. INSTANT LOAD CAMERA
10. 10 SPEED BICYCLE

★ VALID THIS WEEK ONLY!! ★

To Claim Your Gift
Call: [ASK FOR UNITED EXT.]

OFFER EXPIRES

Call Immediaely!

Urgent That You Call At Once!
This special offer is made available to you by your local Independent Distributor for United Laboratories of America, Inc. — Now celebrating over two decades of reliable service.

★ GIFT AT NO COST OR OBLIGATION ★

Although the purpose of this offer is to acquaint you with our film processing service, you are not required to make a purchase to receive your gift. If married, both husband and wife must be present. We ask only that you present your Phone-A-Gram in person so that we can erase the magic ink and determine your prize.

SEE REVERSE SIDE FOR ADDITIONAL INFORMATION
A Fine Print Definition of “NO COST”

United Laboratories Of America, Inc.
P. O. Box 4499 • Station A
Dallas, Texas 75208

PHONE-A-GRAM  URGENT-U-REPLY

Note: The address was printed here.
A Fine Print Definition of “NO COST”

Should a legal adjudication be made that valid consideration was transferred or expended by offeree pursuant to the specific terms of this written offer, then in such event restitutory indemnification shall be accomplished hereunder upon written request.
What Researchers Haven’t Understood

• Text types are not defined by what a reader can do, but by the task or purpose the author is trying to accomplish.

• Language proficiency (and what the proficiency scales describe) is a multi-dimensional trait.

• The best way to measure the development of multidimensional traits is to define progress using steps or a hierarchy of levels within the multidimensional continuum.
C-R Tests of Reading Proficiency

1. Define the construct to be tested.
2. Use the proficiency scales to describe Task, Conditions, and Accuracy (TCA) criteria for each level to be tested.
3. Train the items writers to develop C-R, TCA-aligned items that target a specific level.
4. Test whether the C-R, TCA-aligned item sets cluster in a difficulty hierarchy.
1. The Construct to be Tested

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

<table>
<thead>
<tr>
<th>Author purpose</th>
<th>Reading purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orient</td>
<td>Get necessary information</td>
</tr>
<tr>
<td>Inform</td>
<td>Learn</td>
</tr>
<tr>
<td>Evaluate</td>
<td>Evaluate and synthesize</td>
</tr>
</tbody>
</table>
2. Level-specific, TCA-aligned Criteria

• When an educated writer is writing for an educated audience...
  – The author’s purpose normally determines the genre or text mode that is used.
  – The text mode normally determines the text type that is used for the intended audience.

• When testing for proficient reading, ...
  – The reader task matches the writer’s purpose.
  – The question asked or test item used must match the reader’s task.
The Effect of Test Item Choices

http://www.gocomics.com/fminus/2012/08/02

STUDIES HAVE SHOWN THAT STUDENTS PERFORM MUCH BETTER ON SINGLE-CHOICE TESTS.
3. Trained Item Writers

- Trained the items writers on the construct.
- Trained item writers on the proficiency scale.
- Trained item writers on external text modes, internal text types, and their relationship to author purpose.
- Trained item writers to align the reader’s task with the author’s purpose.
- Retrained the item writers.
- Mentored the item writers.
4. Confirming that the C-R, TCA-aligned items do cluster in a difficulty hierarchy.

• Checked for alignment before piloting items using a Modified-Modified-Angoff procedure.
  – Estimated at-level difficulty.
  – Estimated difficulty for learners at 1 level lower.
  – Estimated difficulty for learners at 1 level higher.
  – Estimated item discrimination across levels.

• Confirmed alignment and discrimination with WINSTEPS IRT analyses.
Results, Reading Simplified Chinese
(Parallel results obtained in other languages.)

**ITEM LEVEL RELIABILITY—Simulated Data Fully Crossed**

| TABLE 3.1 Chinese Simplified Deleted Items—Simul ZOU037WS.TXT Oct 10 16:23 2012 |
| INPUT: 213 Person 50 Item MEASURED: 213 Person 50 Item 2 CATS WINSTEPS 3.70.0.3 |

<table>
<thead>
<tr>
<th>SUMMARY OF 213 MEASURED Person</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>MEAN</td>
</tr>
<tr>
<td>S.D.</td>
</tr>
<tr>
<td>MAX.</td>
</tr>
<tr>
<td>MIN.</td>
</tr>
<tr>
<td>REAL RMSE</td>
</tr>
<tr>
<td>MODEL RMSE</td>
</tr>
<tr>
<td>S.E. OF Person MEAN = .11</td>
</tr>
</tbody>
</table>

Person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .94
Reading Simplified Chinese
Initial Means and Confidence Intervals
With this documented alignment,

• We can produce Reading tests that:
  – Are criterion-referenced.
  – Can be computer adaptive.
  – Include the advantages of testlet-based design.
  – Use “floor and ceiling” scoring.

• Generate reports that show:
  – The reader’s sustained level of proficiency.
  – Progress the reader has made toward the next higher proficiency level.
Thank you.

Questions?