Assessing Learning Outcomes in STARTALK Programs

Presented by: Megan C. Masters and Margaret E. Malone

STARTALK Spring Meeting
Saturday May 5, 2012
Atlanta, GA
Objectives

- In this session, we will:
  - Provide an overview of research completed to date, including a description of the research project, participant characteristics and assessment instruments
  - Discuss the impact of training on increasing learners’ self-assessment accuracy
  - Share a sneak peek at an enhanced, interactive self-assessment training intervention designed for **ALL** summer 2012 STARTALK learners
  - Detail study participation for summer 2012
STARTALK Assessment

• Problem – How can STARTALK appropriately assess student learning in summer programs in multiple languages without discouraging students or losing the focus on proficiency?
STARTALK Assessment

• Solution
  – Prioritize student self-assessment and promote goal setting using LinguaFolio (LFO).
STARTALK Assessment

• Research Questions

1. Will self-assessment data provide useful (valid and reliable) data?
2. Can students really self-assess?
Methodology

• Compare an objective test of language ability with subjective indicators of learning outcomes by students (LFO) and teachers

• Data collected during summer 2010 and summer 2011
Overview of the Study

• Who has participated?
• What were the characteristics of the programs?
• What tools did we use to collect information?
Participants

STARTALK Programs
N = 29

STARTALK Learners
N = 730

STARTALK Instructors
N = 65
Program Characteristics

• 60 + contact hours (~3 weeks or longer)
• High school
• Arabic, Chinese, and Hindi
• Mostly Novice-Intermediate
• Geographic Variation
• Willingness to participate
# Assessment Instrument Overview

Each instrument administered at end of STARTALK Program

<table>
<thead>
<tr>
<th>Tool</th>
<th>Name</th>
<th>Developer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student self-assessment</td>
<td>LinguaFolio Online - 56 Mandatory Can-Do Statements</td>
<td>National Council of State Supervisors for Languages (original)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Center for Applied Second Language Studies at Univ. of Oregon (online)</td>
</tr>
<tr>
<td>Teacher assessment</td>
<td>Teacher Retrospective Observation of Student Progress - 56 Mandatory Can-Do Statements</td>
<td>Center for Applied Linguistics (CAL)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dan Reed, PhD of Michigan State University</td>
</tr>
<tr>
<td>Objective test</td>
<td>Computerized Assessment of Proficiency (CAP) - Four skills - Multiple Choice (L,R) - Constructed Response (S,W)</td>
<td>Center for Applied Second Language Studies at Univ. of Oregon (CASLS)</td>
</tr>
</tbody>
</table>
Example of LinguaFolio Online (LFO)

![LinguaFolio Online Example](image)

<table>
<thead>
<tr>
<th>Description</th>
<th>I Can Do With Help</th>
<th>I Can Do</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>I can greet people in a polite way using single words and memorized phrases.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can say hello to a person I do not know.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can say hello to an adult.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can say hello to my teacher.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can say hello to someone my age or younger.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can introduce myself to someone.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I can.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

This is a tool for tracking language proficiency.
# Example of TROSP

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Rating (circle one)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>The student can...</em></td>
<td>Not Covered</td>
</tr>
<tr>
<td>label some familiar things.</td>
<td>NC</td>
</tr>
<tr>
<td>write about him or herself using learned phrases and memorized expressions.</td>
<td>NC</td>
</tr>
<tr>
<td>write lists that help in day-to-day life.</td>
<td>NC</td>
</tr>
<tr>
<td>describe aspects of daily life using phrases and simple sentences.</td>
<td>NC</td>
</tr>
<tr>
<td>write short notes using phrases and simple sentences.</td>
<td>NC</td>
</tr>
<tr>
<td>express opinion on familiar topics using a series of sentences with some details.</td>
<td>NC</td>
</tr>
<tr>
<td>write questions to obtain and clarify information.</td>
<td>NC</td>
</tr>
<tr>
<td>write personal communications on familiar topics using connected sentences with supporting details.</td>
<td>NC</td>
</tr>
<tr>
<td>write a description or explanation of a familiar topic using connected sentences with supporting details.</td>
<td>NC</td>
</tr>
</tbody>
</table>
Example CAP Test - Reading

Situation
You decide to drop in on your Arabic instructor when you see this note on his office door.

Question 1/1
What should you do?
- knock on the door
- come back after lunch
- come back next week
- do not disturb
Example CAP Test - Listening

Item: CS-L2-018-A

**Situation**
You are leaving home when you hear your Chinese neighbors talking.

**Question 1/1**
Where is the woman going?
- to a store
- to a gym
- to a restaurant
- to a theater

**Indicates the correct answer**

**Dialogue**

Woman: Had your meal yet?
Man: Just did. Where are you going?
Woman: (I'm) Going to a store to get something.
Man: Bye then!
Woman: Bye.
Writing Task A
You have a new pen pal from New Delhi and you need to write the first email to him/her to introduce yourself. In addition to your name, tell him/her at least two things about yourself, and ask at least two questions about him/her. (The questions should cover different topics than the statements.) You may want to include:

- interests & hobbies
- school
- city
- physical description
- family members
- pets
- etc

Be sure to include an appropriate greeting and sign-off. Write as many words as you can.
Example CAP Test-Speaking

Speaking Task

You are living in New Delhi and have just been to your first big party. Your Indian friends want to know what parties are like where you are from. Describe a party you went to in the past including some of the following:

- where it was
- what you were celebrating
- who was there
- what you did
- the food and drink served
- the music or other entertainment
Data Analyses (2010)

Two main aspects of data analyses were of interest:

1. How do students’ self-assessment ratings compare with teachers’ ratings of students on the same 56 CanDo statements?

2. How do student and teacher ratings compare with the objective measure?
Results: Cross Program Variation 2010

Each assessment instrument provided sufficiently reliable data, but the validity varied across institutions.

Figure 1. Mean outcomes across programs (2010)
In other words…

Student self-assessment

Teacher assessment of students
Does Self-Assessment Training Help?

Summer 2011:

Online Self-Assessment Training Opportunities Provided for:
- STARTALK Teachers
- STARTALK Learners
Three aspects of data analyses were of interest:

1. How do students’ self-assessment ratings compare with teachers’ ratings of students on the same 56 CanDo statements?

2. How do student and teacher ratings compare with the objective measure?

3. Did self-assessment training help increase the accuracy of ratings?
Results: Cross Program Variation 2011

Each assessment instrument provided sufficiently reliable data, but the validity varied across institutions.

Figure 2. Mean outcomes across programs (2011)
In other words...
Effect of Self-Assessment Training on Learners

- Training **did not** significantly increase accuracy of learners’ self-assessments ($t = -.246$, $p = 0.808$)

**Figure 3.** Slopes as Outcome Analysis of LFO
Effect of Self-Assessment Training on Teachers

- Training **significantly increased accuracy of teachers’ assessments** of learners (t = 2.618, p = 0.015)

![Figure 4. Slopes as Outcome Analysis of TROSP](image-url)
Overall Effect of Self-Assessment Training

Online Self-Assessment Training Opportunities Provided for:
- STARTALK Teachers
- STARTALK Learners

Diagram:
- Program Directors
- Teachers
- Students
Summary of Findings

• The effects of LFO training seemed to reach teachers, but not students
  – Why?
  – LFOtraining.com

• Students less confident about their abilities than teachers
  – Why?
  – What can we do?
Enhanced Training: Summer 2012

• Additional systematic training is necessary to increase learner self-assessment accuracy
  - Interactive training video has been designed SPECIFICALLY FOR STARTALK STUDENTS
  - Learners will be asked to view authentic video-based interactions between interlocutors at various levels of proficiency
  - Learners will be asked to rate actors’ performances according to the LFO “Can Do” scale
    - I Can Do
    - I Can Do with Help
    - This is a Goal
A Sneak Peek

• Training example

• All teacher and student-related training resources can be found in one place at:

  www.LFOtraining.com

• Complete self-assessment training module will be ready by June 1, 2012
Interested in Participating?

- Participation Involves
  1. Computerized testing of all students
     - 1 hour / student total
     - Coordinated and set-up by CASL/CAL researcher
  2. Teacher completion of TROSP (56 items)
     - Likely to be completed while students are completing computerized test
  3. Student completion of LFO mandatory can-do statements (56 items)
     - To be completed by students at end of STARTALK program
Interested in Participating?

• Our goal is to collect data from as many different STARTALK programs as possible
• We’d LOVE for you to participate!
• Checklist:
  ✓ Arabic, Chinese or Hindi Program
  ✓ At least 60 contact hours of instruction
  ✓ Novice to Intermediate level students
  ✓ Access to computers for all students
Thank you to our Sponsors!

Associate Directorate of Education and Training (ADET)

Office of the Director of National Intelligence (ODNI)
Thank you to our Team!

- Steven J. Ross (Principal Investigator, CASL)
- Martyn K. Clark (CASL)
- Anne Donovan (CAL)
- Kassandra Gynther (CASL)

To volunteer to participate, or to obtain more information, please contact:

mmalone@cal.org

mmasters@casl.umd.edu

Questions?
Q&A

• Why do we care about self-assessment accuracy?
• Why might it be so hard for learners to self-assess their language skills?
• What assessment practices are implemented in your STARTALK classroom?
Additional slides

• Triangulation of outcomes (correlations)
• Rasch analyses
Triangulation of Outcomes

Correlations among measures are mid to low-range

Summer 2010

<table>
<thead>
<tr>
<th></th>
<th>TRO</th>
<th>CAP</th>
<th>LFO</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRO</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAP</td>
<td>.529</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>LFO</td>
<td>.252</td>
<td>.424</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Summer 2011

<table>
<thead>
<tr>
<th></th>
<th>TRO</th>
<th>CAP</th>
<th>LFO</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRO</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAP</td>
<td>.602*</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>LFO</td>
<td>.183</td>
<td>.338</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Rescaling of assessment instruments

- Rasch rating scale model/Rasch mixed-scale model
  - Each person ability and task difficulty estimate converted to a logit metric
  - Scaling of person ability and task difficulty allows for visual examination of relative overlap
  - Scale set at zero to allow histograms to be mirror images
Methodology

• Outcomes on each instrument re-scaled with a Rasch item-response theory model
  – Allowed for learner outcomes to be generated on a single comparable scale

• Multi-level modeling used to assess the influence of program-level variables on individual difference variables at learner level
  – Training (students/teachers)
  – Heritage learners
Learners less confident in their ability to perform tasks
Teacher-based Assessment

Teachers more confident in their students’ ability to perform tasks
Exploring Program Variation

Residual = Predicted – Observed Score

Normal distribution of Residuals

Over-estimation

Under-estimation

Studentized Residual

LANG
- Arabic
- Chinese
- Hindi

PROG

17.5  20.0  22.5  25.0  27.5
Exploring Program Variation

Residual analysis of CAP performance as predicted by LFO and TROSP

Summer 2010

Summer 2011