Semi-direct Tests: Tools for Self Assessment and Feedback on the Path to High Level Proficiency Levels?

Flagship: Results 2012

Margaret E. Malone
Background: Direct and semi-direct tests

Research: Development and feasibility of computer-based tests

Products: Self-access rater training and student tutorials

Potential washback for Flagship
There is a dearth of assessment materials in less commonly taught languages.

(Jackson & Malone, 2009; O’Connell & Norwood, 2007)

A knowledge of assessment fundamentals, or assessment literacy, is necessary to properly select, develop, and use assessments.

(Inbar-Lourie, 2008; Taylor, 2009)
Part 1: Semi-direct oral proficiency testing

- Background on oral proficiency testing
- Development of the COPI
The Oral Proficiency Interview: A face-to-face interview conducted between a trained tester and examinee in the target language

FSI/ILR/ACTFL. Originally developed for government use and revised for academia

Advantages: Reliability, consistency between government and academia, positive washback on classroom teaching and assessment for trained testers

Disadvantages: Asymmetrical power structure of interviewer and examinee, restriction in role of interviewer, time and resources
1970s: The Recorded Oral Proficiency Exam (ROPE)

1980s: Simulated Oral Proficiency Interview (SOPI), Video Oral Communication Instrument (VOCI)

1990s-2000s: Computerized Oral Proficiency Instrument (COPI), Oral Proficiency Interview (Computer)

2010 and beyond: Student tutorials and return of self-assessment.
The OPI and COPI

**COPI**
- Relies on computer and guide, with text and illustrations, to elicit language
  - Directions in English or target language, prompt in target language
  - The interlocutor can change to reflect language and culture
- Recorded for rating and review

**OPI**
- Relies on highly trained interviewer to elicit language
  - Interview in target language (except role play instructions)
  - Warm-up, Level Checks, Probes, Wind-down
- Recorded for future confirmation of rating
CAL researchers wanted to use computer technology to overcome some limitations of SOPI method, esp. to allow more *examinee autonomy*.

Feasibility study 1997-2000 (Spanish, Arabic, Chinese)

Task refinement study (*CAST Project*) 2002-2005
Elements of examinee control

- Self-assessment determines the proficiency level of first task administered
- Topic of tasks
- Level of task difficulty (alternating tasks)
- Thinking and response time
- Language of task instructions (Advanced and Superior tasks only)
You will now assess your own Modern Standard Arabic speaking abilities. The results of this self-assessment will help you select the level of your first COPI task.
Now let’s go on to the next task. Compared to the last task, how challenging would you like the next one to be? Click on the box below that describes your preference.

- The same
- More challenging
- Less challenging
In which language would you like the task description to be? Click on the box below that describes your preference.

- English
- Arabic
Feasibility study: Examinee planning and response time

- Use varied by global proficiency level of examinees as well as by level of COPI task
- Least planning time for Novice-level tasks, then Intermediate, then Advanced, then Superior
- Examinees rated at Superior and Advanced-High levels used less planning time on average than those at Advanced and Intermediate sublevels
- At tasks at the same level, examinees with higher proficiency ratings provided longer responses than those with lower proficiency ratings
- Response time used on the COPI for Superior and Advanced tasks was much longer than that provided on the SOPI
Since early 1980s, CAL has developed years of experience in designing tasks on semi-direct tests of oral proficiency to optimize elicitation of speech samples ratable according to the criteria of the ACTFL Speaking Proficiency Guidelines (mostly through cognitive labs and examinee surveys).

**CAST Project (2002-2005)**
- Examined characteristics of OPI, SOPI and COPI tasks
- Identified characteristics of tasks successful in eliciting speech samples ratable according to the ACTFL Speaking Proficiency Guidelines
- Determined (rhetorical) functions necessary to be elicited at each major ACTFL proficiency level
COPI model: Facilitating research

- Research on student performance, rater behavior and individual tasks facilitated by
  - Standardized tasks
  - Digitization and storage of examinee responses
  - Tracking of examinee behavior
  - Electronic storage of ratings and rater comments on each task
  - Tracking of rater behavior within rating interface

- Task standardization may also facilitate development of automatic oral proficiency scoring on constrained tasks
Part 2: Oral proficiency resources for teachers and students

- Self-access rater training materials
- Student tutorial
Two Arabic assessment resources

- **Multimedia Rater Training Program (MRTP)**
  A self-paced computerized training program on rating oral proficiency assessments based on the *ACTFL Proficiency Guidelines – Speaking (Revised 1999)*

- **Student Self-Assessment Tutorial**
  A self-paced online tutorial for students to explore the basics of oral proficiency assessment and to assess their own oral proficiency skills
Purpose of the resources

- To give users control and agency in assessing oral proficiency
  - Instructors can effectively incorporate oral proficiency assessment in instruction.
  - Students can assess their own skills and set learning goals.

- To encourage positive washback, or the beneficial influence of assessment on instruction and learning
To ensure feasibility of material delivery and adherence to project timeline, development process is collaborative and iterative.
To ensure usefulness of materials and to promote positive washback, development process incorporates stakeholder feedback at every step.

- **Development Stage 1**: Exploratory focus groups
- **Development Stage 2**: Resource review
- **Development Stage 3**: Validity and reliability check
Stakeholder feedback: Instructors

- Instructor feedback gathered through
  - Semi-structured focus groups
  - Independent review of draft tutorial
  - Conference presentations
# Exploratory focus groups: format

<table>
<thead>
<tr>
<th>N</th>
<th>Program</th>
<th>Arabic Background</th>
<th>Heritage Learners</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Large public university</td>
<td>Native speakers</td>
<td>Some in student population</td>
</tr>
<tr>
<td>4</td>
<td>Suburban public high school district</td>
<td>Native speakers</td>
<td>Many in student population</td>
</tr>
</tbody>
</table>

- 2 moderators: 1 facilitator, 1 note taker
- Duration: about one hour
- Audio-recorded and later transcribed
- Participants compensated for their time
“We ignore to some extent the speaking part, which I want to see more emphasis on.” (HS)

“With limitations of time and resources, I don’t think it’s easy for us to meet our goals and objectives through testing.” (university)

“Lacking the good materials…is one of the serious problems that we are facing.” (HS)

“Training really situates the instructor in a position where he or she knows what to expect from the students.” (university)

“What should we do? Is it ok if the students speak fluent dialect, is this enough or no?” (HS)
Tutorial review: Format

- 2 native speaker instructors at a large public university
- Sent link to live tutorial with instructions for review of each page and each module
- Responses returned electronically
- Participants compensated for their time
Tutorial review: Questions

- New information you learned; new information your students might learn
- Information you already know; information your students might already know
- Information to be added or deleted
- Areas of confusion
- Reactions to images and audio samples
- Technical difficulties
- Additional questions and comments
Feedback from instructors

- Results from focus groups:
  - Desire for additional assessment training and materials
  - Issues surrounding teaching and assessing dialect and heritage learners

- Results from tutorial review:
  - Value of tutorial for students
  - Goals of students
Stakeholder feedback: Students

- Student feedback gathered through
  - Semi-structured focus groups
  - Independent review of draft tutorial
  - Group interviews
### Exploratory focus groups: Format

- University student group: 1 facilitator, 1 note taker
- High school student group: 1 facilitator took notes
- Duration: about one hour
- Audio-recorded and later transcribed
- Participants compensated for their time

<table>
<thead>
<tr>
<th>N</th>
<th>Program</th>
<th>Arabic Background</th>
<th>Heritage Learners</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Large public university</td>
<td>Upper- and lower-level learners</td>
<td>N=2</td>
</tr>
<tr>
<td>6</td>
<td>Urban public high school</td>
<td>Lower-level learners</td>
<td>N=0</td>
</tr>
</tbody>
</table>
“I don’t really know if what I’m getting in class is really an accurate reflection of what I can and cannot do.” (university)

“I like the speaking tests better. I think I learn more when I have to speak it.” (HS)

“I’d just like to emphasize again how helpful it is to be assessed more on speaking than we are currently.” (university)

“I did self assessments by talking to people…I try to talk to them and when they talk back and if I understand then I feel happy.” (HS)
Tutorial review: Format

- Sent link to live tutorial with instructions for review of each page and each module
- Some responses returned electronically
- Group interview duration: about 1 hour, 45 minutes
- Audio-recorded and later transcribed
- Participants compensated for their time

<table>
<thead>
<tr>
<th>N</th>
<th>Program</th>
<th>Arabic Background</th>
<th>Heritage Learners</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Large private university</td>
<td>Upper- and lower-level learners</td>
<td>None</td>
</tr>
<tr>
<td>10</td>
<td>Large public university</td>
<td>All in second-year intensive class</td>
<td>Some</td>
</tr>
</tbody>
</table>
Tutorial review: Questions

- New information learned
- Information to be added or deleted
- Areas of confusion
- Reactions to images and audio samples
- Technical difficulties
- Additional questions and comments
“Maybe just a little more information, not about like the history of ACTFL, but who uses ACTFL …and maybe what are the other standardized tests that are used.”

“I think the anecdotal information would be great here, you know, just somebody saying, you know, ‘I’m an intermediate and I feel like I can survive in Cairo decently and make friends.’”

“I really want to improve my oral proficiency, I want more specific tips. I want more things that I can do. Or more specific references I can read about.”

“What I realized was that I shouldn’t just stick with the book, but I really need to get out there and like start talking you know Arabic with native speakers.”

“And I liked the listening parts and I liked the charts a lot. It's just there was -- it was very wordy.”
Feedback from students

Results from focus groups:
- Frustration with lack of oral skills practice and assessment
- Desire to learn more about oral proficiency

Results from tutorial review:
- Request for streamlined tutorial content
- Appreciation of audio samples
Applications of these assessment resources

- **Rater training**
  - Provides self-access training on oral proficiency rating
  - Has potential for washback to the classroom for task and activity development
  - Allows instructors to find patterns across students and tasks (e.g., all students in 101 receive a lower score on task X than Task Z. Why?)

- **Student tutorial**
  - Allows students to learn about the proficiency levels
  - Provides background on expectations for proficiency levels for Flagship students
  - Supplies student testimonials on the importance of assessment and oral skills practice
Please visit this free tutorial at http://www.cal.org/aop
Thank you!

- Questions?
- Comments?

mmalone@cal.org