Effects of video vs. audio modes on ratings of learner recordings

Betsy Lavolette, Michigan State University
June 14, 2012
CALICO
Motivation

- Does it make a difference if teachers use audio or video to assess speaking skills?
- How can teachers make an informed choice of assessment tools?
What is “speaking”?  

• No mention of gestures or facial expressions in descriptions of speaking in  
  – ACTFL guidelines (ACTFL, 2012)  
  – Common European Framework (CoE, 2001)  
  – TOEFL speaking rubrics (ETS, 2008)  

• Why not?
Background

- Studies of how language learners use video information in listening tests (Wagner, 2007, 2008)
- Test takers score
  - HIGHER (Sueyoshi & Hardison, 2005; Wagner, 2010)
  - LOWER (Suvorov, 2009)
  - SAME (Londe, 2009)
on listening tests when provided with video vs. audio texts
- No research has examined how raters rate video vs. audio recordings made by learners
Research questions

1. a. Do students practice more when making either audio or video recordings? b. Why?

2. a. Do students prefer recording audio or video? b. Why?

3. a. Do teachers prefer rating audio or video? b. Why?

4. Do teachers rate audio and video recordings differently? If so, how?
Participants & materials

- Recordings: 39 ESL students at English language center
- Ratings: 13 language teachers

- Prompts & rubric: TOEFL iBT independent speaking
  - e.g., Choose a teacher you admire and explain why you admire him or her. Please include specific examples and details in your explanation. (ETS, 2006)
  - Holistic rubric (0 to 4)
Student demographics

- N = 39
- Mean age = 20
- 27 male, 12 female
- L1
  - Chinese: 16
  - Arabic: 15
  - Korean: 4
  - Other: Japanese (1), Kurdish (1), not specified (3)
- Mix of proficiencies from intermediate to advanced
Teacher characteristics

- \( N = 13 \) (12 completed exit questionnaire)
- Mean age = 27
- 6 male, 6 female
- L1
  - English: 7
  - Arabic, Japanese, Spanish, German, French, Afrikaans: 1 each
- All graduate students in language assessment course and language teachers
Procedure: Recordings

- ESL students recorded in language lab
  - 2 audio
  - 2 video
- 45 seconds each
- CLEAR Audio & Video Dropboxes
- Opinion/demographic questionnaire
Procedure: Ratings

- Each rater rated all recordings
- Types of recordings:
  - Audio
  - Video
  - Audio track from video
- Raters were aware of purpose of study
- All raters noticed that they rated some recordings twice
1.a. Do students practice more when making either audio or video recordings?

<table>
<thead>
<tr>
<th></th>
<th>Audio</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>n (students who answered both prompts)</td>
<td>35 (90%)</td>
<td>27 (70%)</td>
</tr>
<tr>
<td>Mean tries/prompt</td>
<td>2.2</td>
<td>1.4</td>
</tr>
<tr>
<td>SD tries/prompt</td>
<td>2.6</td>
<td>1.0</td>
</tr>
<tr>
<td># of students who tried 2X or more</td>
<td>17 (49%)</td>
<td>10 (37%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Audio</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students who indicated more practice in each mode</td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>
1.b. Why do students practice more when making audio recordings?

- “Because I want to make sure that you can understand me very well.”
- “I need to make senses without movements help”
1.b. Why do students practice more when making video recordings?

- “because it was amazing!!”
- “i have to make myself ready”
- “may be kind of nerves”
- “Because I have to control my movements”
- Because I can see myself and then I can learn from myself. When I watch myself I feel that this is another person who wants to teach me, and I can learn from it.
4. Do teachers rate audio and video recordings differently?

• Did you watch the video clips while you rated them?

<table>
<thead>
<tr>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>9</td>
</tr>
<tr>
<td>No</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>81.8</td>
</tr>
<tr>
<td>No</td>
<td>18.2</td>
</tr>
</tbody>
</table>

• Did you feel that you rated the audio files differently than the video files?

<table>
<thead>
<tr>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>3</td>
</tr>
<tr>
<td>No</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>25</td>
</tr>
<tr>
<td>No</td>
<td>75</td>
</tr>
</tbody>
</table>
4. Do teachers rate audio and video recordings differently?

<table>
<thead>
<tr>
<th>Rater</th>
<th>INTERrater reliability</th>
<th>INTRArater reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>.541</td>
<td>.499</td>
</tr>
<tr>
<td>R4</td>
<td>.590</td>
<td>.613</td>
</tr>
<tr>
<td>R5</td>
<td>.658</td>
<td>.521</td>
</tr>
<tr>
<td>R8</td>
<td>.431</td>
<td>.120 (n.s.)</td>
</tr>
<tr>
<td>R10</td>
<td>.642</td>
<td>.524</td>
</tr>
<tr>
<td>R11</td>
<td>.648</td>
<td>.728</td>
</tr>
<tr>
<td>R14</td>
<td>.638</td>
<td>.676</td>
</tr>
<tr>
<td>R16</td>
<td>.687</td>
<td>.711</td>
</tr>
<tr>
<td>R17</td>
<td>.563</td>
<td>.577</td>
</tr>
<tr>
<td>R18</td>
<td>.613</td>
<td>.418</td>
</tr>
<tr>
<td>R19</td>
<td>.517</td>
<td>.566</td>
</tr>
<tr>
<td>R20</td>
<td>.621</td>
<td>.616</td>
</tr>
</tbody>
</table>

Percent agreement: 47%
4. Do teachers rate audio and video recordings differently?

- 3 repeated-measures ANOVAs:
  - All 13 raters
  - Reliable raters
  - Unreliable raters

- To determine if mean ratings significantly differ based on modality:
  - Audio
  - Audio track from video
  - Video
4. Do teachers rate audio and video recordings differently?

- All 13 raters
- Means:
  - Audio = 2.49
  - Audio from video = 2.31
  - Video = 2.27
- Omnibus test significant
  - V = .17
  - F(2, 33) = 3.46
  - p = .043
- Pairwise comparisons:
  - Approaching significance: Audio > video, p = .063, r = .38 (M)
4. Do teachers rate audio and video recordings differently?

- Reliable raters
- Means:
  - Audio = 2.46
  - Audio from video = 2.21
  - Video = 2.27
- Omnibus test significant
  - $V = .225$
  - $F(2, 33) = 4.79$
  - $p = .015$
- Pairwise comparisons:
  - Audio > audio from video, $p = .049$, $r = .40$ (M)
4. Do teachers rate audio and video recordings differently?

- **Unreliable raters**
- **Means:**
  - Audio = 2.54
  - Audio from video = 2.44
  - Video = 2.25
- **Omnibus test significant**
  - $V = .44$
  - $F(2, 33) = 13.11$
  - $p < .001$
- **Pairwise comparisons:**
  - Audio > video, $p = .010$, $r = 0.48$ (M)
  - Audio from video > video, $p = .001$, $r = .58$ (L)
Conclusions

• Trends toward:
  – Students practice more on audio vs. video
  – Students prefer recording audio over video
  – Teachers prefer rating video over audio

• Raters overall:
  – Audio > video

• Reliable raters
  – Audio > audio from video

• Unreliable raters
  – Audio & audio from video > video
Pedagogical implications

• Don’t let students choose between audio and video
• Audio & video recordings are different assignments
  – Performance may differ
  – Rating process may differ
Thank you to
Dr. Paula Winke
Dr. Dennie Hoopingarner & CLEAR
ELC teachers & students
LLT 808 students

betsy@msu.edu