A meta-analysis of effects of multiple-choice and open-ended formats on reading and listening test performance

Yo In’nami, Toyohashi University of Technology, Japan
innami@hse.tut.ac.jp

Rie Koizumi, Tokiwa University, Japan
koizumir@tokiwa.ac.jp

LTRC 2008

Outline
1. Introduction
2. Literature review
3. Research questions
4. Method
5. Results
6. Implications
7. Suggestions for future research
1. Introduction

- Effects of test format

Taro is taller than Hanako and Jiro.

Question: Who is the tallest?

<table>
<thead>
<tr>
<th>MC</th>
<th>OE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Taro.</td>
<td></td>
</tr>
<tr>
<td>B. Hanako.</td>
<td></td>
</tr>
<tr>
<td>C. Jiro.</td>
<td></td>
</tr>
</tbody>
</table>

* | *A. Taro. | Taro. |
| B. Hanako. |   |
| C. Jiro. |         |
2. Literature review (1)

- Format difficulty order from the literature

<table>
<thead>
<tr>
<th>Domain</th>
<th>Difficulty order</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 reading (e.g., Davey, 1987; Elinor, 1997)</td>
<td>MC &lt; OE, MC = OE</td>
</tr>
<tr>
<td>L2 reading (e.g., Shohamy, 1984; Trujillo, 2006)</td>
<td>MC &lt; OE, MC = OE</td>
</tr>
<tr>
<td>L2 listening (e.g., Teng, 1999)</td>
<td>MC &lt; OE</td>
</tr>
</tbody>
</table>

2. Literature review (2)

**Meta-analysis** (e.g., Cooper & Hedges, 1994):
- summarizes a set of empirical data across studies;
- fully reports selection and inclusion criteria;
- codes and uses information from the study findings and characteristics systematically.
2. Literature review (3)

Examples of meta-analyses:
- Ross (1998)

No meta-analysis of differences of difficulty between MC and OE → Our study

3. Research questions

(RQ1) Which format is easier, MC or OE?
(RQ2) To what extent is it easier?
(RQ3) Are there any moderator variables related to format effects?

- Separate investigations for L1 reading, L2 reading, and L2 listening studies
4. Method (1)

- Collecting previous studies articles, reports, proceedings (published and unpublished) through books, journals, databases

- 56 data sources from 37 studies

- Coding of studies: statistics (mean, $SD$, $r$, $n$) & moderator variables

4. Method (2)

15 moderator variables

(1) Between-subjects or within-subjects designs (counterbalanced or non-counterbalanced)
(2) Random or non-random assignment
(3) Item stem or non-stem equivalency
(4) Access or no-access to the text when answering
(5) Text explicit or implicit questions
(6) No. of options in an MC format
4. Method (3)
(7) L2 proficiency level (middle or high)
(8) Age (primary, secondary, or adult)
(9) Learners’ L1
(10) Learners’ L2

(11) Reliability of an MC format
(12) Reliability of an OE format
(13) Reliability of scoring in an OE format
(14) % of correct answers in an MC format
(15) % of correct answers in an OE format

4. Method (4) Data analysis

• Calculation of an effect size for each study

\[
\frac{(MC \ mean - OE \ mean)}{(pooled \ SD)}
\]

• Calculation of a mean effect size & confidence interval across studies
4. Method (5) Data analysis

- Examination of the effects of moderator variables
  (a) Classify studies according to each variable
  (b) Perform a single variable linear regression

- Effect size interpretation based on Cohen (1988):
  Small: |0.20| to |0.50|
  Medium: |0.50| to |0.80|
  Large: |0.80| or above

5. Results

5.1 Format effects in L1 reading

(RQ1) MC < OE (k = 22)
(RQ2) The degree of difference: small to large (0.65 [0.24, 1.06])
  (A positive effect size with a 95% confidence interval that did not include zero indicated MC < OE.)
(RQ3) Moderator variables:
  - between-subjects designs
  - within-subjects designs
  - random assignment
  - stem equivalency, etc.
5.2 Format effects in L2 reading
(RQ1) Overall, MC = OE (k = 11)
(RQ2) No difference (0.22 [–0.23, 0.66])
(RQ3) Moderator variables:
   between-subjects designs
   random assignment
   stem equivalency
   learners with a high L2 proficiency level
   →Under these conditions, MC < OE

5.3 Format effects in L2 listening
(RQ1) MC < OE (k = 5)
(RQ2) The degree of difference: medium to large (1.11 [0.57, 1.66])
(RQ3) Moderator variables:
   between-subjects designs
   random assignment
   stem equivalency
   access to the text when answering
   (note-taking allowed), etc.
5.4 Key moderator variables
(a) between-subjects designs
(b) random assignment
(c) stem equivalency

Possible reasons
- (a) Exclude carryover effects?
- (b) & (c) Eliminate irrelevant factors?

5.5 Comparison with traditional literature review

<table>
<thead>
<tr>
<th>Domain</th>
<th>Traditional literature review</th>
<th>Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 reading</td>
<td>MC &lt; OE</td>
<td>MC &lt; OE (small to large)</td>
</tr>
<tr>
<td></td>
<td>MC = OE</td>
<td></td>
</tr>
<tr>
<td>L2 reading</td>
<td>MC &lt; OE</td>
<td>MC = OE (MC &lt; OE under certain conditions)</td>
</tr>
<tr>
<td></td>
<td>MC = OE</td>
<td></td>
</tr>
<tr>
<td>L2 listening</td>
<td>MC &lt; OE</td>
<td>MC &lt; OE (medium to large)</td>
</tr>
</tbody>
</table>
6. Implications

- Using an MC format rather than an OE format would make the test easier.

- e.g., in a test with an SD of 10, 6.5 points higher in MC than OE in L1 reading, 11.1 points higher in MC than OE in L2 listening

7. Suggestions for future research

- Larger aggregations of studies in L2 listening
- Other moderator variables to be included (e.g., passage length or learners’ background knowledge)
- Effects of other test formats to be meta-analyzed (e.g., cloze and c-tests)
Thank you for listening!

The full reference of studies can be obtained from the authors.

7. Acknowledgement & references

Supported by Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science
The full reference of studies can be obtained from the authors.


