Replication of structural equation models in language testing and learning

Yo In'nami
(Toyohashi Univ. of Tech., Japan; innami@hse.tut.ac.jp)
Rie Koizumi
(Tokiwa Univ., Japan; koizumir@tokiwa.ac.jp)
AAAL2010, March 8, 2010

1. Replication

1. Replication

• Replication of structural equation models (SEM)
  – Common in social-science books and articles (e.g., Bollen, 1989)
  – Rare in L2 testing and learning

2. Data availability for replication

2. Data availability for replication

• Replication of structural equation models requires the variance/covariance or correlation matrices.
  – Read articles
  – Contact authors

1. Replication

• Increasing interest in the language learning literature (e.g., Gass & Mackey, 2007; Language Teaching Review Panel, 2008)
  – Replication helps empirical and theoretical findings build on one another (e.g., Kline, 2004)
  – Replication encourages critical appraisal of studies, as conducting good replication studies presupposes proper understanding of the original studies (e.g., King, 1995)

1. Replication

• Replication of structural equation models can be conducted for each previous study, but of greater interest would be a synthesis of a replication of structural equation models.
  • Hulland, Chow, and Lam (1996) collected 343 models in marketing using SEM in 186 articles and found that only 112 out of the 343 models (33%) reported sufficient information for reanalysis. They successfully replicated 75 out of the 112 models (67%).
3. This study

- To what extent is information for the replication provided by contacting authors?
- To what extent can the original models be successfully replicated?

4. Method: Article collection

- Studies using SEM in L2 testing and learning were collected in October 2008.
  - Studies in 20 representative journals (e.g., Applied Linguistics, TESOL Q) were retrieved via ERIC, LLBA, & search engines available at journal homepages.
  - Keywords: confirmatory factor analysis, covariance structure (analysis), structural equation modeling…
  - To cross-check the studies electronically identified, we conducted manual searches in these journals.

4. Method: Research Question 1

- We extracted information necessary for replication from each article.
  - A variance/covariance matrix
  - A correlation matrix (with means and SDs)
- When such information was missing, we e-mailed all the (co-)authors and requested the matrix.
  - We explained the purpose of our research.
  - If we received a reply, we asked why the matrix was not included in the article.

4. Method: Research Question 2

- We replicated each model with Amos, by exactly following the procedures reported in the original articles.
- We divided our replication analyses into 3 stages:
  - (a) Preliminary analysis (non-convergence and negative error variances)
  - (b) Model fit indices & (c) parameter estimates (i.e., factor loadings, correlations, & error variances)
  - We calculated the absolute value differences in (b) model fit indices and (c) parameter estimates between each original model and its replicated one.
    - For example, CFI = 0.95 for the original model & 0.94 for our replicated model, then code the absolute value difference of 0.01.

- ↑ Bae & Bachman replicated: CFI = .950, NNFI = .929
4. Method: Research Question 2

- We counted the number of models whose absolute value differences in model fit indices and parameter estimates between the original and replicated models lay out of the mean +/- 2SD range (thus judged to be outliers, or failures of replication).
  - For example, a mean absolute value difference in NFI between the original and replicated models = 0.036 with an SD of 0.052, then the number of models whose value differences lay out of -0.068 and 0.140 (0.036+/-.2*0.052) was counted.

5. RQ1: Availability of information for replication from the authors

- 50 articles used SEM in 7/20 journals.
- 19/50 articles (38%) contained enough information for reanalysis.
- We e-mailed the 33 authors of the remaining 31 articles missing information for reanalysis.
  - 18 authors responded, 4/18 authors provided data.

<table>
<thead>
<tr>
<th>Reply</th>
<th>Information obtained = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>--Variance/covariance matrix = 2</td>
</tr>
<tr>
<td></td>
<td>--Correlation matrix with means and SDs = 1</td>
</tr>
<tr>
<td></td>
<td>--Raw data = 1</td>
</tr>
<tr>
<td></td>
<td>Information not obtained = 14</td>
</tr>
<tr>
<td></td>
<td>--Data too old to be retrieved = 6</td>
</tr>
<tr>
<td></td>
<td>--No further response despite promise = 6</td>
</tr>
<tr>
<td></td>
<td>--Data discarded when moving; decades old, occupied much space in the office = 1</td>
</tr>
<tr>
<td></td>
<td>--Large-size raw data caused a transmission problem = 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No reply</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Information obtained = 4</td>
</tr>
<tr>
<td></td>
<td>--Variance/covariance matrix = 2</td>
</tr>
<tr>
<td></td>
<td>--Correlation matrix with means and SDs = 1</td>
</tr>
<tr>
<td></td>
<td>--Raw data = 1</td>
</tr>
<tr>
<td></td>
<td>Information not obtained = 14</td>
</tr>
<tr>
<td></td>
<td>--Data too old to be retrieved = 6</td>
</tr>
<tr>
<td></td>
<td>--No further response despite promise = 6</td>
</tr>
<tr>
<td></td>
<td>--Data discarded when moving; decades old, occupied much space in the office = 1</td>
</tr>
<tr>
<td></td>
<td>--Large-size raw data caused a transmission problem = 1</td>
</tr>
</tbody>
</table>

5. RQ2: Replicability of the original models

- We reanalyzed the 124 models with sufficient information for reanalysis.
- Preliminary analysis:
  - 19/124 models faced problems with model unidentification (13 models) or not positive definite matrices (6 models). 3 of the unidentified models and 2 of the models with not positive definite matrices had the same problems in the original analysis. Thus, the 10 unidentified models and the 4 models with not positive definite matrices did not have such problems in the original study but did in our replication, suggesting that 11% (14/124) of the models reanalyzed failed to be replicated. In other words, 89% of the models were successfully replicated.

5. RQ2: Replicability of the original models

- For the remaining 105 models without analytical problems, the differences in values of (b) fit indices and (c) parameters were examined.
  - Differences in values of fit indices between the original studies and our reanalysis were, in most cases, very small
    - Near-zero values of median and mode, and the large positive values of skewness and kurtosis
  - The number of models whose absolute value differences in model fit indices indicated outliers was small (0 to 13%). In other words, successful replication rate was high, ranging from 87% to 100%, depending on the fit index used.
    - NFI difference between the original and replicated models was small (median = 0.015, mode = 0.000), distributed near zero (skewness = 2.214), had a strongly peaked distribution (kurtosis = 5.670), with one of the 24 models judged to be an outlier (4%).
5. RQ 2: Replicability of the original models

- Parameter estimates differed little between the original and replicated models (median = 0.040; mode = 0.010), with most estimates distributed around the mean (skewness = 2.284; kurtosis = 5.529). Only two out of the 35 parameters (6%) failed to be replicated, suggesting the successful replication rate of 94%.

6. Summary

- Research Question 1 (To what extent is information for the replication provided by contacting authors?)
  - Information necessary for reanalysis was obtained from only four out of the 33 authors (12%). Such information was not provided in the articles primarily because doing so was not requested by the editor(s)/reviewers.
  - Our success rate for getting the data was lower than but similar to previous studies (24%, Wolins, 1962; 26%, Wickerts et al., 2006; 38%, Craig & Reese, 1973).
  - Many of the journals in L2 testing and learning follow the APA Publication Manual, in which data sharing is encouraged. Nevertheless, our mostly failed attempt to obtain data from the authors suggests in reality a very limited availability of data from the authors.

- Research Question 2 (To what extent can the original models be successfully replicated?)
  - 124/360 models were reanalyzed.
  - 89% of the models were successfully replicated in the preliminary analysis.
  - 87-100% (depending on the fit index used) of the fit indices were successfully replicated.
  - 94% of the parameters were successfully replicated.

7. Three implications

- Report (a) the variance/covariance matrix and/or (b) the correlation matrix with means and SDs for any SEM analysis
  - Enables editors and reviewers to conduct model replication
  - Archive data at journal/authors’ homepages
    - Precludes errors in manually inputting data from articles
    - Avoids paper length constraints
    - Reduces response bias: Renowned researchers contacting the authors of the original studies are more likely to receive responses to inquiries
  - Encourage reanalysis of structural equation models
    - Reanalyze existing structural equation models before undertaking new studies