An Application of Multilevel Modelling to Meta-Analysis, and Comparison with Traditional Approaches

Alison O’Mara & Herb Marsh

Department of Education, University of Oxford
Substantive and methodological synergy

- Enhancing self-concept of children and adolescents through interventions
- Methods of meta-analysis
Meta-analysis

- Systematic synthesis of various studies on a particular research question
- Collect all studies relevant to a topic
- “Content analysis”
- An effect size is calculated for each outcome
- Effect sizes with similar features are grouped together and compared
- This allows identification of moderator variables
Model assumptions in meta-analysis

- Fixed effects
  - All of the variability between effect sizes is due to sampling error alone (Hedges & Vevea, 1998)
  - Effect sizes are independent

- Random effects
  - Variability between effect sizes is due to sampling error plus variability in the population of effects
  - This model assumes that studies are heterogeneous to an extent (Erez et al., 1996), because each study has different contexts, researchers, and even methods.
  - Effect sizes are independent
Multilevel modelling meta-analysis

- Multilevel
  - Meta-analytic data is inherently hierarchical (i.e., effect sizes nested within studies)
  - Variability between effect sizes is due to sampling error plus variability in the population of effects
  - Effect sizes are not necessarily independent

- Allows for multiple effect sizes per study (Goldstein, 1995; Hox, 2002; Bryk & Raudenbush, 1992)

- Provides more precise and less biased estimates of between-study variance than traditional techniques (Van den Noortgate & Onghena, 2003)
Self-concept interventions

- Unclear whether self-concept interventions are effective
- Problems in literature:
  - Methodological considerations
  - Conceptual inconsistencies
    - Focus of this presentation
### Theoretical perspectives

<table>
<thead>
<tr>
<th>UNIDIMENSIONAL</th>
<th>MULTIDIMENSIONAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self evaluations are consistent across different contexts</td>
<td>Domains of self-concept are distinct from each other</td>
</tr>
<tr>
<td>Self-concept is the sum or total perception of the self</td>
<td>E.g., math self-concept, physical appearance self-concept, social self-concept</td>
</tr>
<tr>
<td>Instruments measure global evaluations (“I am a good person”), or sum together evaluations of different aspects to yield ‘total’ self-concept score</td>
<td>Instruments measure specific domains (“I am good at math”)</td>
</tr>
</tbody>
</table>
The problem…

- Evaluating self-concept interventions from unidimensional perspective loses information.
Sampling

- Selection criteria
  - Measure of self-concept/ self-esteem at posttest
  - Mean age of 18 or younger
  - Control group
  - Published

- Total yield of 145 articles from the years 1958 to 2000

- 200 interventions

- 460 effect sizes
**Effect size calculation**

**Standardised Mean Difference**
(Hedges & Olkin, 1985)

\[
ES = \frac{\bar{X}_{G1} - \bar{X}_{G2}}{s_{pooled}}, \text{ where } \\
\sqrt{s_{pooled}^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}
\]

**Hedges correction for small sample size bias**

\[
d_i = ES \left[ 1 - \frac{3}{4N - 9} \right]
\]
Weighting

- In fixed and random effects, the effect sizes are weighted by the inverse of the variance to give more weight to effects based on large sample sizes.

- Variance is calculated as

\[ v_i = \frac{(n1 + n2)}{(n1 \cdot n2)} + \frac{d_i^2}{2(n1 + n2)} \]
Fixed effects meta-analysis

- The analog to the ANOVA homogeneity analysis is appropriate for categorical variables
  - Also referred to as Q-test
  - Follows a chi-square distribution
  - Looks for systematic differences between groups of responses within a variable
- Can also conduct regression analyses (not discussed here)
Random effects meta-analysis

- Follows the same procedures as fixed effects models (i.e., homogeneity analyses and regression), except that it adds a random variance component to the variance.

- The variance component is typically calculated as:

\[ \nu_\theta = \frac{Q - (k - 1)}{\sum w_i - (\sum w_i^2 / \sum w_i)} \]

- The new weighting is by the formula:

\[ w_{i\text{RE}} = 1/(\nu_i + \nu_\theta) \]
To help minimise violations of assumption of independence in fixed and random effects analyses, Cooper’s (1998) shifting unit of analysis was used.

Effect sizes are aggregated based upon the particular moderator variable, such that each study only includes one effect size per outcome on that particular variable.
Multilevel meta-analysis

- Levels
  - Level 3: publication level component
  - Level 2: study/intervention level component
  - Level 1: effect size outcome level component
- Intercept-only model gives overall mean effect size

\[ d_{ijk} = \beta_{000} + \nu_{0k} + u_{0jk} + e_{ijk} \]

- \( \nu_{0k} \) is the random error at level 3,
- \( u_{0jk} \) is the random error at level 2, and
- \( e_{ijk} \) is the random error (residual) at Level 1.
Software

- Fixed and random effects: macros for SPSS (Lipsey & Wilson, 2001) using method of moments
- Multilevel: MLwiN using restricted maximum likelihood estimation (see Hox, 2002)
## Results summary – ‘empty model’

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed effects</th>
<th>Random effects</th>
<th>Multilevel</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta$/intercept (SE)</td>
<td>.31(.02)</td>
<td>.51(.07)</td>
<td>.47(.06)</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>.28,.35</td>
<td>.38,.64</td>
<td>.37,.61</td>
</tr>
<tr>
<td>$p$-value $\chi^2$ test ($df = 144$)</td>
<td>$p &lt; .001$</td>
<td>$p &lt; .001$</td>
<td>$p &lt; 0.001$</td>
</tr>
</tbody>
</table>

Heterogeneous outcomes: need to model moderator & predictor variables
Other ways of showing heterogeneity between studies in MLM

- The intercepts for the different studies (level 3 residuals, $v_{0jk}$) have a variance, $\sigma^2_{v0}$, of .186 (SE = .085)
- ICC = .271.
Construct validation

- Target self-concept domains - self-concept domains with focal relevance to the intervention’s goals
- Target-related - logically related to the intervention’s goals, but are not primary
- Non-target - not expected to be enhanced by the intervention

Example: Reading self-concept intervention
- Target = Reading self-concept
- Target-related = School self-concept
- Non-target = Physical appearance self-concept
## Predictor variable – outcome relevance

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed</th>
<th>Random</th>
<th>Multilevel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>.49</td>
<td>.55</td>
<td>.55</td>
</tr>
<tr>
<td>Target-relevant</td>
<td>.11</td>
<td>.49</td>
<td>.47</td>
</tr>
<tr>
<td>Non-target</td>
<td>.08</td>
<td>.21</td>
<td>.26</td>
</tr>
<tr>
<td>( p )-value ( \chi^2 ) test</td>
<td>( p &lt; .001 )</td>
<td>( p &lt; .001 )</td>
<td>( p &lt; 0.001 )</td>
</tr>
</tbody>
</table>

- **Reference variable**: Self-concept domains with focal relevance to the intervention goals.
- **Non-target**: Logically related to treatment goals, but not targeted.
- **Target-relevant**: No expected to be enhanced by the intervention goals.
Implications

- Demonstrates importance of substantive/methodological synergies
  - Multidimensional constructs require MLM

- Use of multilevel modelling in meta-analysis
  - Results differ from previous meta-analyses using fixed effects model and random effects model
  - Similar to random effects when not too heterogeneous
  - Slight differences likely due to estimation procedures for calculating random error variance components (non-iterative vs. iterative)
  - Less likely to reach significance (larger confidence intervals)
Limitations and future directions

- Fine-tuning multivariate approach using response variables (e.g., Kalaian & Raudenbush, 1996; Goldstein, 1995)
- Multilevel missing data imputation in MLwiN
- Simulation
Analyses not discussed here…

- Other moderator variables
  - E.g., random assignment, control group type
- Follow-up data analysis
- Inter-rater reliability (Cohen’s kappa)
- Publication bias
  - Fail safe $N$
  - Trim and fill procedure (Duval & Tweedie, 2000a, 2000b)
  - Power analysis
Questions

- alison.omara@edstud.ox.ac.uk
- herb.marsh@edstud.ox.ac.uk