Selecting Parsimonious Statistical Models in Evaluation: Using the AIC (Akaike Information Criterion) as a Tool for Model Selection

Shelly Engelman, Ph.D.
Tom McKlin, Ph.D.
The Findings Group, LLC
**Conceptual Overview**

- **Goal** = To identify a robust, parsimonious statistical model in order to:
  - Identify variables that have the strongest impact on outcomes
  - Add theoretical value to other programs/contexts
  - Test and utilize in other evaluation contexts (theory-driven evaluation)

- **Problem** = Regression models are commonly used to evaluate how well variables explain variance in the outcome. R-square is a commonly used statistic to evaluate model fit.
  - However, regression models are misused and are fraught with problems for assessing a model with multiple, complex relationships

- **Solution** = Structural Equation Modeling using AIC (Akaike Information Criterion) fit statistic has many advantages over regression models and R-square
Project Overview

• Project: NIH-funded program to expand the use of inquiry-based laboratory exercises using the Bean Beetle Model system.
  • The program provided faculty with summer workshops to develop inquiry-based laboratory studies using bean beetles.
Project Overview, Continued

• Does inquiry-based learning improve students’ science knowledge and research skills?
  • Significantly different from traditional curricula?

• Mechanism: How does inquiry-based learning improve student learning? What is the theory of change?

• Model of change apply across genders? Races/ethnicities?
Methods

• Biology-Based Learning Intervention (4-year Phase 2 project)
• 17 college/university faculty trained to use intervention in inquiry-based lab classes
• How does the intervention impact students’ science context knowledge and research skills?
• N = 602 students from 8 colleges/universities
  – 327 Females; 207 Males
  – 90 Asians; 86 Blacks; 48 Hispanics; 312 Whites; 66 Multiracial
Measures

Perceptions of Instructional Practices (Flora & Cooper, 2005)

1. **Assessment for Learning** (4-items)
   - “My instructor graded students through methods such as presentations, portfolios, and exhibitions.”

2. **Authentic Activities** (3-items)
   - “You work on projects that are meaningful to you.”

3. **Facilitated Learning** (4-items)
   - “You participated in whole-class discussions where your instructor talked less than the students.”

4. **Complex Tasks** (3-items)
   - “You are asked to apply prior knowledge to new tasks.”

5. **Inquiry Based Learning** (4-items)
   - “You work on tasks that require you to develop your own experimental designs.”

Mediating Variables:

- **Confidence** (modified from Baldwin et al., 1999) (5-items)
  - “I am confident in my abilities to design an experiment that is a valid test of a hypothesis.”

- **Attitudes** (modified from Knezek & Christensen, 1996) (4-items)
  - “I enjoy scientific laboratory activities very much.”

Outcome Variables:

- **Science Knowledge and Research Skills**
  - Lawson’s Classroom Test of Scientific Reasoning
  - Nature of Science (Lederman et al., 2002)

- **Intention to Persist** (modified from Williams, Weibe, Yang, & Miller, 2002)
  - “I can see myself working as a researcher in a science field.”
Theoretical Model

Weiner’s Attribution Model:
Cognitions ➔ Emotions ➔ Behavior

- Assessment for Learning
- Authentic Activities
- Facilitated Learning
- Complex Tasks
- Inquiry-Based Learning

Confidence

Science/Research Knowledge

Attitudes

Intention to persist
MultiModel Assessment: Fix Path(s) to 0
Top 6 Advantages to Using AIC

1. Assesses a complex model with multiple relationships simultaneously
2. Considers multiple models before selecting the best model
3. AIC penalizes attempts at over-fitting a model which leads to a parsimonious model
4. Parsimonious models often generalize to other evaluation contexts
## Results

<table>
<thead>
<tr>
<th>Fit Statistics</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6 (Final Model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>535.01</td>
<td>497.57</td>
<td>402.82</td>
<td>399.13</td>
<td>311.09</td>
<td>288.47</td>
</tr>
</tbody>
</table>

- AIC allows evaluators to consider multiple models before selecting the best model. The model with the lowest AIC is considered most robust and parsimonious.
Final Model: 
Partially Mediated Model

- Authentic Activities
- Complex Tasks
- Confidence
- Science/Research Knowledge
- Intention to persist
Final Model Fit:

**Males**

\[ \chi^2 (df=32) = 36.78, p < .05; \text{RMSEA} = .03, \text{CFI} = .95. \]

**Females**

\[ \chi^2 (df=34) = 49.12, p < .05; \text{RMSEA} = .05, \text{CFI} = .96. \]

**Neuroscience Scholars**

\[ \chi^2 (df=43) = 77.02, p < .05; \text{RMSEA} = .061, \text{CFI} = .95. \]
Useful Tools


2. SPSS AMOS - free download of student version: [http://www.amosdevelopment.com/download/](http://www.amosdevelopment.com/download/)

3. SEM Workshop at Data Analysis Training Institute of Connecticut (DATIC) with David A. Kenny and Betsy McCoach: [http://datic.uconn.edu/](http://datic.uconn.edu/)

4. Instructional Videos: [www.thefindingsgroup.com](http://www.thefindingsgroup.com)
Thank you!
Top 4 **Disadvantages** to Using regression and R-square to Evaluate a Model...

1. Complex models with more than one relationship are dealt with in a **piecemeal nature**
2. **Lack of parsimony** because adding additional variables to a model artificially increases the R-square
3. **Difficult to replicate** because of the lack of parsimony
4. **Does not generalize** to other evaluation contexts
Results

<table>
<thead>
<tr>
<th>Fit Statistics</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6 (Final Model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>535.01</td>
<td>497.57</td>
<td>402.82</td>
<td>399.13</td>
<td>311.09</td>
<td>288.47</td>
</tr>
</tbody>
</table>

- AIC allows evaluators to consider multiple models before selecting the best model. The model with the lowest AIC is considered most robust and parsimonious.
- Model 6 “fit” the data for both male and female students.
- Our conclusion is that students’ perceptions of their instructors’ inquiry-based practices (Authentic Activities; Complex Tasks) predict their content knowledge in science and research and their intent to persist; confidence is a partially mediating variable.
- Improvement to the attitude measure is warranted.