Evaluating Impacts at Multiple Levels: The Viewpoint from Project and Program Evaluation

Mack Shelley
University Professor of Statistics, Political Science, and School of Education
Iowa State University

National Alliance for Broader Impacts Summit
April 30, 2015
Madison, Wisconsin

Broader Impacts

- How well does the activity advance discovery and understanding while *promoting teaching, training, and learning*?
- How well does the proposed activity *broaden the participation of underrepresented groups* (e.g., gender, ethnicity, disability, geographic, etc.)?
- To what extent will it *enhance the infrastructure for research and education*, such as facilities, instrumentation, networks, and partnerships?
- Will the *results be disseminated broadly* to enhance scientific and technological understanding?
- What may be the *benefits of the proposed activity to society*?

How do I Know What Broader Impacts Look Like in Practice?

- Examples of specific activities to demonstrate broader impacts are available at:

  http://www.nsf.gov/pubs/2004/nsf042/bicexamples.pdf

  - Proposal authors should be creative in demonstrating the broader impacts of their projects.
  - Try to link similar kinds of activities you already may have underway to the research and education projects you are proposing for funding.
  - Proposers also should consider what types of activities best suit their interests, while enhancing the broader impacts of the project being proposed.

The NSF *User-Friendly Handbook for Project Evaluation*

http://informalscience.org/documents/TheUserFriendlyGuide.pdf

Why conduct evaluations?

- Evaluation produces information that can be used to make continuous improvements in the project.
- An evaluation can document what has been achieved.
  - extent to which goals are reached and desired impacts are attained
- Evaluation frequently provides new insights or new unanticipated information.
- There is an inherent interrelationship between evaluation (formative and summative) and project implementation.
- Provides information for communicating to stakeholders about the worth of the project to the public and "up the line" to senior decisionmakers and funders.

From *The User-Friendly Handbook*

- Different Types of Evaluation
  - Feasibility evaluation
  - Process evaluation
  - Implementation evaluation
  - Formative evaluation
  - Progress evaluation
  - Summative evaluation

From *The User-Friendly Handbook*

- Six phases of project evaluation
  - Develop a conceptual model (focused on the broader impacts) of the program and identify key evaluation points
    - Logic model
  - Develop evaluation questions and define measurable outcomes
  - Develop an evaluation design
  - Collect data
  - Analyze data
  - Provide information to interested audiences (i.e., tell your story)

The Need for Rigorous Evaluation

- In HR 3801 [The Education Sciences Reform Act of 2002], "**scientifically valid education evaluation**"

https://www2.ed.gov/policy/rschstat/leg/PL107-279.pdf

  - adheres to the highest possible standards of quality with respect to research design and statistical analysis
  - examines the relationship between program implementation and program impacts
  - provides an analysis of the results achieved by the program with respect to its projected effects
  - employs experimental designs using random assignment when feasible, and other research methodologies that allow for the strongest possible causal inferences when random assignment is not feasible

The Need for Rigorous Evaluation

What is the "Gold Standard?"

- U.S. Department of Education, Institute of Education Sciences, *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*

http://www2.ed.gov/rschstat/research/pubs/rigorousevid/rigorousevid.pdf

The Need for Rigorous Evaluation

- Charles C. Ragin, Joane Nagel, and Patricia White. (2004). *Workshop on Scientific Foundations of Qualitative Research*. Arlington, VA: National Science Foundation.

http://www.nsf.gov/pubs/2004/nsf04219/nsf04219_1.pdf

  - Recommendations to improve the quality of qualitative proposals and for evaluating such proposals. E.g., what is an ideal qualitative proposal?

Examples of Contexts for Broader Impacts

- Alliances for Graduate Education and the Professoriate

http://www.nsfagep.org/evaluation-resources/

- Louis Stokes Alliances for Minority Participation (LSAMP)

http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=13646&org=HRD&from=home

- Bridges to the Baccalaureate (B2B)
  - Connected to LSAMP
- Integrative Graduate Education and Research Traineeship (IGERT; now NRT—National Science Foundation Research Traineeship Program)

http://www.nsf.gov/crssprgm/igert/intro.jsp

An Example: ATOMS^{2XP}

- Advancing Teachers of Middle School Science (ATOMS^{2XP})
  - 3-year program aimed at increasing the number of highly qualified 5th–8th grade science teachers in school districts throughout the state of Mississippi, and the academic achievement levels among the students of those teachers
  - 120 hours of science content and pedagogical instruction
    - summer institute
    - video conferences
    - online discussions
    - attendance at state science events and conferences
  - 39 teachers comprised the first-year cohort
    - data were available on 35

ATOMS^{2XP} Outcomes

- Evaluation design focused on assessment of:
  - teachers' content knowledge
  - changes in pedagogy
  - student achievement
  - teachers' journaling of their experiences in the program
- Measures to evaluate progress and outcomes included
  - pre- and post- teacher content knowledge tests
  - classroom observation taxonomy
  - pretests and posttests measuring student content knowledge
  - teacher reflections

Variable Name	Variable Label
Didwell	Teacher did well (= 1) or did not do well (= 0)
Preoba	Teacher pre overall below (= 0) or above (= 1) mean
Postoba	Teacher post overall below (= 0) or above (= 1) mean
Timprove	Teacher improved score (1 = yes; 0 = no)
Gainba	Teacher gain below (= 0) or above (= 1) mean
Tprepba	Teacher pre Physics below (= 0) or above (= 1) mean
Tpostpba	Teacher post Physics below (= 0) or above (= 1) mean
Tphyimpr	Teacher Physics improved (post minus pre) (1 = yes; 0 = no)
Gphpctba	Teacher gain Physics percentage below (= 0) or above (= 1) mean
Tpreweba	Teacher pre Weather below (= 0) or above (= 1) mean
Tpoweba	Teacher post Weather below (= 0) or above (= 1) mean
Tweaimpr	Teacher Weather improved (post minus pre) (1 = yes; 0 = no)
Gwepctba	Teacher gain Weather percentage below (= 0) or above (= 1) mean
Tcpct	Student overall correct percentage pretest
Pcpct	Student Physics correct percentage out of 14 questions pretest
Wcpct	Student Weather correct percentage out of 16 questions pretest

HLM 2-Level Models

- Of particular interest was the nested structure of the data (students within teachers) and whether there were significant level-2 (teacher) impacts on student outcomes of gains in overall science content knowledge, physics content knowledge, and weather content knowledge.
- This multilevel analysis was conducted using the hierarchical linear modeling (HLM) software provided by Scientific Software International, Inc. For each of these analyses, non-missing observations were available on 639 students (level 1) and 35 teachers (level 2).

HLM 2-Level Model for Student Gain in Physics Correct Percentage (Posttest Minus Pretest)--GAINPCPC

The model specified for the fixed effects was

Level-1 Model

$GAINPCPC = B_0 + B_1*(DIDWELL) + B_2*(TPREPBA) + B_3*(TPOSTPBA) + B_4*(TPHYIMPR) + B_5*(GPHPCTBA) + B_6*(PCPCT) + R$

Level-2 Model

$B_0 = G_{00} + G_{01}*(GAINPCPC) + U_0$

$B_1 = G_{10} + G_{11}*(GAINPCPC)$

$B_2 = G_{20} + G_{21}*(GAINPCPC)$

$B_3 = G_{30} + G_{31}*(GAINPCPC)$

$B_4 = G_{40} + G_{41}*(GAINPCPC)$

$B_5 = G_{50} + G_{51}*(GAINPCPC)$

$B_6 = G_{60} + G_{61}*(GAINPCPC)$

Final estimation of fixed effects for Student Gain in Physics Correct Percentage (Posttest Minus Pretest)--GAINPCPC

- In no case did a level-2 variable have a statistically significant effect on the slope of the relationship of any level-1 effect on the outcome of individual students' gains in physics science content knowledge.
- However, there were significantly lower gain scores across classrooms with higher mean levels of student pretest physics content knowledge; this is reasonable, because there is less room for improvement for students in classes with higher initial levels of physics content knowledge.

HLM 2-Level Model for Student Gain in Weather Correct Percentage (Posttest Minus Pretest)--GAINWCPC

The model specified for the fixed effect is:

Level-1 Model

$GAINWCPC = B_0 + B_1*(DIDWELL) + B_2*(TPREWEBA) + B_3*(TPOWEBA) + B_4*(TWEAIMPR) + B_5*(GWEPCTBA) + B_6*(WCPCT) + R$

Level-2 Model

$B_0 = G_{00} + G_{01}*(GAINWCPC) + U0$

$B_1 = G_{10} + G_{11}*(GAINWCPC)$

$B_2 = G_{20} + G_{21}*(GAINWCPC)$

$B_3 = G_{30} + G_{31}*(GAINWCPC)$

$B_4 = G_{40} + G_{41}*(GAINWCPC)$

$B_5 = G_{50} + G_{51}*(GAINWCPC)$

$B_6 = G_{60} + G_{61}*(GAINWCPC)$

Final estimation of fixed effects for Student Gain in Weather Correct Percentage (Posttest Minus Pretest)--GAINWCPC

- Students of teachers whose pretest weather content knowledge was above the mean for all teachers had significantly higher weather content knowledge gains.
- The relationship between student weather content knowledge gains and teachers' improvement was greater for teachers who improve more than the average improvement for all teachers.
- Students had lower gain scores in weather content knowledge when they were in classrooms with higher mean levels of initial student weather content knowledge; this demonstrates a ceiling effect—students have less room for improvement when they already have scored well on initial weather content knowledge.

Conclusions and Recommendations

- Improved teachers' overall science content knowledge was associated with statistically significant improvement in students' science content knowledge overall and in Physics, but not necessarily in Weather.
- Targeted educational interventions may focus more effectively on specific content areas such as physics or weather than on attempts to increase student science knowledge more generally.
- The HLM result that level-2 effects are more pronounced for weather than for physics suggests that targeted interventions may have a greater impact for helping students to learn about weather than about physics. Gains are not equal across content areas.

Wrapping Up

- This ATOMS^{2XP} discussion is one example of how we can evaluate on multiple levels (here, students and teachers) simultaneously. The approach can be generalized rather easily for just about any evaluation context for which broader impacts are a focus.
- Project and program evaluation methods more generally provide a helpful framework for knowing what works, for whom, and with what possible broader impacts.

Thank you

Questions?
Comments?