

**The Trachtenberg School of Public Policy and Public Administration**  
**Fall 2018**

**Course Number:** PPPA 6085-14

**Course Title:** Advanced Techniques in Impact Evaluation

**Description:** This course is an alternative to PPPA 6016 for students who have taken at least two semesters of statistics and econometrics and would like to cover impact evaluations in more depth. Although the focus of the course is on impact evaluations, the course also includes several modules on other topics that are important for MPP, MPA, and Ph.D. students to learn. Readings on various topics generally include articles on the approaches and applications of the approaches.

**Prerequisites:** PPPA 6013 or an equivalent course in regression analysis

**Professor:** Dr. Burt S. Barnow  
Suite 601T  
Telephone: 202-994-6379 (O)  
202-427-7928 (C)  
E-mail: [barnow@gwu.edu](mailto:barnow@gwu.edu)  
Office hours: Monday and Tuesday 10:30 am to noon and by appointment. NOTE: I am here most days, so please feel free to drop by anytime or call/email me to tell me when you want to meet.

**Required and Optional Readings:**

Joshua D. Angrist and Jorn-Steffen Pischke, Mastering Metrics, Princeton University Press. ***Required and in the book store***  
Allan Kimmel, Ethics and Values in Applied Social Research, Sage 1988. ***Required and available on Blackboard***  
Paul J. Gertler et al. Impact Evaluation in Practice, the World Bank. ***Required and available on Blackboard***  
Shahidur R. Khandker, Gayatri B. Koolwal, and Hussain A. Samad, Handbook on Impact Evaluation, the World Bank. ***Required and available on Blackboard***  
Michael Lewis, Moneyball. 2004. ***Optional, but worth reading.***

And chapters from Newcomer, Hatry, and Wholey *The Handbook of Practical Program Evaluation*, Jossey-Bass, 4th Edition. ***Required and available on Blackboard***

All required readings except *Mastering Metrics* are on blackboard or will be emailed to you.

### **Student Learning Outcomes:**

Through course discussions, readings, and assignments, students will develop knowledge and skills to enable them to:

- 1) develop program logic models;
- 2) identify pertinent professional standards and ethical principles affecting specific dilemmas confronting evaluators in the field;
- 3) design implementation, outcome, and particularly impact evaluations;
- 4) assess impact evaluations for quality and relevance;
- 5) identify useful performance measures and design performance measurement systems;
- 6) learn how to distinguish performance measurement from impact evaluations;
- 7) learn how to combine results from multiple evaluations
- 8) develop useful recommendations based on evaluation findings.

### **Method of Instruction:**

The tasks and constraints facing professionals involved in the design and implementation of program evaluations are explored by class participation in both in-class and written exercises. Questions and problems facing both evaluators and managers of programs being evaluated are examined. Note: Attendance is required for successful completion of this course. Active class participation can enhance your grade significantly.

### **Assigned Readings:**

Assigned readings are selected to give students a representative sample of the professional evaluation literature, as well as to expose them to the sorts of issues which arise in the context of real life evaluations. Note that some of the reading assignments have flaws or limitations, so do not assume that all assigned readings are exemplary!

### **Assignments:**

1. Readings: Students are expected to have read required readings prior to the class meeting for which they are listed. Class discussion on the required readings will affect course grades, especially in borderline cases.

**NOTE: ALL written assignments must be submitted in hard copy and electronic copy on or before the due date unless other arrangements have been made in advance. Due dates are firm for all written assignments – late papers will be penalized by lower grades.**

2. Logic model: Students will prepare a logic model for a program of interest. The program should have inputs, activities, outputs, and outcomes, and likely will have contextual factors that will affect the outcomes (15% of grade). More detailed instructions will be provided one week before the due date. ***Due September 17.***

3. One Critique of an article: Students will review critically a scholarly article of their choice. (30% of the grade). ***Due November 5.***

There is no mandatory length for the paper, but in my experience papers 10-20 pages double spaced are optimal. The paper should include:

- A brief description of the focus and findings;
- Identification of the key policy questions addressed;
- A brief summary of the research design and data collection methods used;
- A systematic discussion of threats to the evaluation. The threats should be labeled as those the authors acknowledged and threats the authors did not acknowledge. The discussion should include an assessment of potential specification errors in the models estimated and the suitability of the comparison group (if there is one).
- A discussion of whether the author's conclusions are appropriate and why or why not.
- Suggestions on how the evaluation could be improved.
- An assessment of whether any policy recommendations made by the author are warranted and what are appropriate recommendations;

4. In-Class Exercises and Debates: In-Class exercises will be held in class throughout the semester. Class debates over ethical issues in program evaluation also will be held throughout the semester and require oral presentations. Students will be graded on their participation in the exercises and debates (accounting for 10% of course grade).

5. Performance Measures Project. Students should select a program of interest and identify the performance measures used for the program. The main idea of the assignment is to assess the current performance measures in terms of strengths and weaknesses and suggest additional or alternative measures. Additional information will be provided 2 weeks before the due date (15% of grade) ***due November 19.***

6. Applied Evaluation Project: Members of the class will respond to a request for proposals (RFP) for an impact evaluation. Students should form teams of two to four students to prepare the response. The groups should discuss their plan with the instructor as soon as you have identified a potential RFP to respond to. (30% of grade) ***Due December 10.***

7. Class participation: Students are encouraged to participate in class discussions. Class participation can count up to 20% of the grade in addition to the criteria listed above.

**Student Expected Effort.**

Over 14 weeks, students will spend 1 hour and 50 minutes (110 minutes) per week in class. Required readings, written assignments, and the debates are expected to take up, on average, 8 hours (480 minutes) per week. Over the course of the semester, students will spend 25.67 hours in instructional time and 112 hours preparing for class, for a total of 137.67 hours.

## **APPLIED PROJECT EVALUATION DESIGN**

This project is designed to provide you with on-the-job training. You are asked to identify a request for proposals (RFP) from a government agency, nonprofit organization, foundation, or international organization calling for an *impact evaluation*. You will only propose the evaluation; you are not expected to conduct the actual evaluation itself. Also, you need not write how you would conduct other activities called for in the RFP, such as an implementation study, but you should note how the implementation study will relate to the impact evaluation. The RFP need not be currently open.

**The report should have all of the components identified in the list below.**

### **Required Elements of the Report for the Applied Project**

The suggested contents and order of presentation for the report are as follows:

- I. Executive Summary:** This should generally be 1-2 pages.
- II. Introduction and Background:** An introduction to the project, including the names of the team should be given along with a description of the scoping activities, including a brief description of the program and a synthesis of relevant past research and evaluation findings. Be sure to cite relevant literature on the program.
- III. Evaluation Questions:** The issues that have been identified in the RFP and the specific questions that are addressed, or should be addressed should be provided. If you believe the RFP misses some important aspects of how the study should be conducted, include these as well and explain why they are useful to include.
- IV. Evaluation Design:** The design(s) to be undertaken, including the concepts and variables, the theory underlying the policy/program, etc. should be provided. A **logic model** of the program/policy must be developed and presented in the body of the report with an appropriate discussion. Be sure to be specific about the method(s) you have chosen and why you have chosen them.

- V. **Data Collection:** The sources of data available, measures used to address the research questions, data collection methods, and sampling procedures should be discussed. Also, there should be a discussion of limitations to validity and reliability, as well as actions undertaken to reduce the impact of the limitations identified. Use of a **design matrix** to cover all of these issues is strongly recommended.
- VI. **Data Analysis:** Proposed analytic strategies should be discussed. Appropriate tables and figures should be constructed in accordance with guidance given in class for projects that are completed. Describe the outcome and explanatory variables to be used, the statistical techniques to be used, and how you will be able to draw inferences about the program's impacts. Discuss the size of the sample to be analyzed and indicate if it is large enough to obtain statistically significant findings if the program has the desired impact (i.e., **conduct minimum detectable impact analysis**.) If the RFP prescribes a sample size, you should determine if it is adequate.
- VII. **Potential Problems and Fall-back Strategies:** Identify the potential problems that may arise in conducting the evaluation and the strategies that will be used to either avoid the problem or deal with its occurrence.
- VIII. **Suggested Deviations from the RFP:** If you believe the evaluation could be improved by modifying the outcome variables, analytical method, alternative data sources, sample size, or anything else, present your alternative strategies here and explain the pros and cons.
- IX. **Proposed Budget, Budget Narrative, and Workplan:** Budgetary estimates may range from specific to general depending upon the complexity of the proposed project. (This section should be brief.)
- X. **Conclusion:** A brief conclusion should be provided.

## **Class Schedule and Assignments**

### **August 27 (Session 1)**

#### **Introduction to the Course and Overview of the Field of Program Evaluation**

##### Readings:

Newcomer, Wholey, and Hatry, Chapter 1  
Gertler et al. Chapter 1  
Angrist and Pischke Introduction  
Khandker, Koolwal, and Samad, Chapters 1 and 2 (skim)

##### Questions:

- ☐ What is program evaluation? What types of studies and analytical support fall under this rubric?
- ☐ How does program evaluation differ from other forms of analysis?
- ☐ What are the different approaches to evaluation?
- ☐ How did evaluation evolve?
- ☐ Where does evaluation take place and who conducts evaluations?
- ☐ What are some of the more critical issues that face the evaluation profession?
- ☐ What role does program evaluation play for international funders, e.g. the World Bank?
- ☐ How do current performance measurement efforts relate to program evaluation?
- ☐ How does organizational culture shape evaluation capacity?

### **September 10 (Session 2) Logic Models and Theory of Change**

##### Readings:

Read the excerpt from Studemund, especially the material on omitted variables and measurement error. The handwritten handout is optional and shows how the results are derived. The Barnow article is optional.

McLaughlin and Jordan Chapter 3 in Newcomer et al.  
Wholey Chapter 4 in Newcomer et al.

Gertler et al. Chapter 2

American Evaluation Association Evaluation Standards

<http://www.uwex.edu/ces/pdande/evaluation/evallogicmodel.html>. (Visit this site and note the templates you can download.)

W.E. Kellogg Foundation (2004). Logic Model Development Guide. (Optional)

James P. Connell et al., eds. *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*. Washington, D.C.: The Aspen Institute, 1995. pp. 1-21.

Kathleen Wells and Julia H. Littell. "Study Quality Assessment in Systematic Reviews of Research on Intervention Effects." *Research on Social Work Practice*, January 2009, pp. 52-62.

Karen Fulbright-Anderson et al., eds. *New Approaches to Evaluating Community Initiatives, Volume 2: Theory, Measurement, and Analysis*. Washington, D.C.: The Aspen Institute, 1995. pp. 15-44.

Questions:

- ☐ What is the guidance provided to evaluators by the AEA professional Standards?
- ☐ What role should staff and external stakeholders play in evaluation?
- ☐ What role can the evaluator play in program development and design?
- ☐ What pre-design steps are desirable for the evaluator to take?
- ☐ What is the program theory? How can it be developed and refined?
- ☐ What is logic modeling?
- ☐ How might logic models guide evaluation?
- ☐ What is evaluability assessment? What are the steps? How can it be used to guide evaluation? How can it be used as a management tool?
- ☐ What is meant by a "theory of change" evaluation?
- ☐ What is required for TOC studies to produce credible estimates of program impacts?

**September 17 (Session 3) Concepts of Validity and Reliability and Experimental Designs I)**

"Threats to Validity and Reliability" by Newcomer

Larry L. Orr. *Social Experimentation: Evaluating Public Programs with Experimental Methods*. Thousand Oaks, CA: Sage Publications. Part 1: Background and Rationale; Part 2: Basic Concepts and Principles; and Part 3: Alternative Random Assignment Models.

Angrist and Pischke, Chapter 1.

Khandker, Koolwal, and Samad, Chapter 3 (Optional)

Gary Burtless, "The Case for Randomized Field Trials in Economic and Policy Research," *Journal of Economic Perspectives* 9 (1995): 6384.

James J. Heckman and Jeffrey A. Smith, "Assessing the Case for Social Experiments," *Journal of Economic Perspectives* 9 (1995): 85110.

G.C. Smith and J.P. Pell. (2003). "Parachute Use to Prevent Death And Major Trauma Related To Gravitational Challenge: Systematic Review of Randomised Controlled Trials." *British Medical Journal*, 327, 14591461.

Richard Nathan and Robinson Hollister (2008). "The Role of Random Assignment in Social Policy Research." *Journal of Policy Analysis and Management*, Spring 2008. pp. 401-415. Also see responses in *Journal of Policy Analysis and Management*, Fall 2008. pp. 606-615 (optional).

- What are the most common threats to measurement validity and measurement reliability, and to internal and external validity?
- What are the advantages of using random assignment?
- What are the disadvantages of using random assignment?
- What role should staff and external stakeholders play in evaluation?
- What types of bias can arise in random assignment studies?

#### **September 24 (Session 4) Experimental Designs II**

Kevin A. Schulman et al. "The Effect of Race and Sex on Physicians' Recommendations for Cardiac Catheterization," *New England Journal of Medicine*, February 25, 1999. 618-626; also read Sounding Board from July 22, 1999 issue. (This article and the commentary are important for understanding how people misinterpret logit analysis.)

Larry L. Orr. *Social Experimentation: Evaluating Public Programs with Experimental Methods*. Part 5: Implementation and Data Collection.

Gertler et al. Chapter 4 (Skim, but note the detailed discussion of differences between ITT and TOT impact estimates and the large number of international examples)

Howard S. Bloom et al, *The National JTPA Study: Title IIA Impacts on Earnings and Employment at 18 Months*, Abt Associates Inc., Bethesda, MD, January 1993. Executive Summary required; rest is optional. (Read ES carefully; see if you can find problems with the presentation of findings as we will spend significant time discussing)

Dennis K. Benson, "Review of the National JTPA Study," Unpublished document. Appropriate Solutions, Columbus, Ohio, March 1992.

Fred E. Romero, "The National JTPA Study: Lessons Not Learned," Unpublished, April 1993.

General Accounting Office. [HEHS9640] *Job Training Partnership Act: LongTerm Earnings and Employment Outcomes pp. 1-15, 26-38.* (Skim, but see if you agree with GAO's conclusions; be sure to read DOL's comments and GAO's responses before you decide if you agree)

Burt S. Barnow (2010). "Setting Up Social Experiments: The Good, the Bad, and the Ugly." *Journal for Labor Market Research* 2010. 43:91-105.

David Greenberg and Burt S. Barnow. (2014). "Flaws in Evaluations of Social Programs: Illustrations from Randomized Controlled Trials." *Evaluation Review*. 38(5): 359-387. (Skim)

- What potential problems do you see in reporting the results from logit analysis? How can you avoid them?
- What are the specific problems that arose in evaluating the JTPA program even though the evaluation used an experimental design? How could the design be improved?
- How could GAO and the Department of Labor disagree so much on what the GAO evaluation of JTPA means? How would you interpret the GAO results?

### **October 1 (Session 5) Non-experimental Methods I (Selection on Observables and Difference in Differences)**

#### **General Non-Experimental Methods**

Edward E. Leamer. "Let's Take the Con Out of Econometrics." *American Economic Review*. January 1983, pp. 31-43. (Not an easy read, but skim and see what lessons you can draw)

Burt S. Barnow, Thomas Kaplan and Robert Moffitt "Introduction" in Burt S. Barnow, Thomas Kaplan, and Robert Moffitt, editors *Evaluating Comprehensive State Welfare Reform: The Wisconsin Works Program*, Albany, NY: Rockefeller Institute Press, 2000. (Skim)

W. Norton Grubb, "The Varied Economic Returns to Postsecondary Education," *Journal of Human Resources* 28 (1993): 365-382. (Read carefully some interesting points, even if problems. **Bring to class—we will discuss in detail.**)

Thomas J. Kane and Cecilia Elena Rouse, Comment on W. Norton Grubb, "The Varied Economic Returns to Postsecondary Education," *Journal of Human Resources* 30 (1995) 205-221. (Optional—but note that they point out errors in reading data by Grubb!)

Howard Bloom, Carolyn Hill, and James Riccio. "Linking Program Implementation and Effectiveness: Lessons from a Pooled Sample of Welfare to Work Experiments." *Journal of Policy Analysis and Management*, Fall 2003. (Skim this to see an interesting study that combines process and impact analysis.)

Howard Bloom, Charles Michalopoulos, Carolyn Hill, Ying Lei (2002). *Can Nonexperimental Comparison Group Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare to Work Program?* New York: Manpower Demonstration Research Corporation. Chapters 1,2, and 4 required; rest is optional.

### **Difference in Difference**

Angrist and Pischke, Chapter 5.

Gertler et al. Chapter 6.

Khandker, Koolwal, Samad, Chapter 5.

Coady Wing et al. (forthcoming). "Designing Difference in Difference Studies: Best Practices for Public Health Research." *Annual Review of Public Health*. (Skim)

Melissa S. Kearney and Phillip B. Levine. (2015). "Media Influences on Social Outcomes: The Impact of MTV's *16 and Pregnant* on Teen Childbearing." *American Economic Review* 105(12): 3597-3632.

David A. Jaeger, Robert Kaestner, and Ted Joyce. A Cautionary Tale of Evaluating Assumptions: Did Reality TV Really Cause a Decline in Teenage Childbearing? *Journal of Business and Economic Statistics*.

Melissa S. Kearney and Phillip B. Levine. (2018). "Does Reality TV Induce Real Effects? Response to Jaeger, Joyce, and Kaestner." IZA DP 10318.

- ☐ What are the commonly used designs to measure program outcomes?
- ☐ What are the considerations in selecting a design to evaluate program impact?
- ☐ How do the evaluators weigh the tradeoffs in various designs?
- ☐ What strategies are available for controlling or ruling out various rival explanations?
- ☐ What designs are applicable for longitudinal data?
- ☐ What are the differences in how race/ethnicity and sex are treated in Grubb?
- ☐ Which of the three models used by Grubb is best and why?
- ☐ What is propensity score matching?

- What are the key assumptions underlying the use of propensity score matching?
- When does propensity score matching work well?
- When does propensity score matching not work well?
- How can you tell if propensity score matching is doing a good job, or can you?
- What are the situations where propensity score matching seems to work well, according to Cook et al.?

---

**October 15 (Session 6) Nonexperimental Methods II (Propensity Score Matching, and Regression Discontinuity Design)**

**Propensity Score Matching**

Gertler et al. Chapter 7.

Khandker, Koolwal, Samad, Chapter 4

Marco Caliendo and Sabine Kopeinig (2008). “Some Practical Guidance for the Implementation of Propensity Score Matching.” *Journal of Economic Surveys* 22:1, 31-72.

Jeffrey A. Smith and Petra E. Todd (2005). “Does Matching overcome LaLonde’s Critique of Nonexperimental Estimators?” *Journal of Econometrics* 125, 305-353. Skim this article. The reply by Dehejia and the rejoinder by Smith and Todd 355-375 are recommended but optional.

Peter R. Mueser, Kenneth R. Troske, and Alexey Gorisslavsky. “Using State Administrative Data to Measure program Performance.” (2007). *Review of Economics and Statistics*. 89(4): 761-783. (note how much more optimistic this article is than article above; also, the article shows that variations in PSM often does not have a big effect on findings; think about how they do DID—do you agree with the approach?.).

**Regression Discontinuity Design**

Gertler et al. Chapter 5

Angrist and Pischke Chapter 4

Khandker, Koolwal, Samad, Chapter 5.

Robin Jacob, Pei Zhu, Marie-Andree Somers, and Howard Bloom (2012). “A Practical Guide to Regression Discontinuity.” New York: MDRC. (Skim)

Thomas D. Cook, William R. Shadish, and Vivian C. Wong. "Three Conditions under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons." *Journal of Policy Analysis and Management*. Autumn 2008. pp. 724-750.

Guido Imbens and Thomas Lemieux (2007). "Regression Discontinuity Designs: A Guide to Practice." Cambridge, MA: National Bureau of Economic Research, Working Paper 13039. Optional. (Not easy, but read to see how the RDD is implemented and especially the graphic tests that can be done to see if it works.)

Howard Bloom (2009). Modern Regression Discontinuity Analysis." (optional)

- ☐ What are the key assumptions underlying the use of the methods described in this week's readings?
- ☐ Which of these methods is the strongest? Why?
- ☐ Which approaches are weak?
- ☐ Are there ways you can tell if the assumptions for these models are valid?

### **October 22 (Session 7) Nonexperimental Methods III: Time Series Regression Models**

Burt S. Barnow, "The Impact of Welfare Reform in New Hampshire on the TANF Caseload: Final Report." Baltimore, MD: Institute for Policy Studies, January 2003.

Travis St. Clair, Kelly Hallberg, and Thomas Cook (201). "The Validity and Precision of the Comparative Interrupted Time Series Design: Three Within Study Comparisons." *Journal of Educational and Behavioral Statistics* 41(3): 269-299. (Skim)

Steven Garasky and Burt S. Barnow, "Demonstration Evaluations and Cost Neutrality: Using Caseload Models to Determine the Federal Cost Neutrality of New Jersey's REACH Demonstration," *Journal of Policy Analysis and Management* 11 (1992): 624-636 (Skim).

Jacob Alex Klerman and Steven J. Haider. (2004). "A Stock-Flow Analysis of the Welfare Caseload." *Journal of Human Resources* 39(4): 865-886.

Rebecca Blank. (2002). "Evaluating Welfare Reform in the United States." *Journal of Economic Literature* 40(4): 1105-1166. (Skim except read pages 1127-1139)

- ☐ What are the key assumptions underlying the use of the methods described in this week's readings?
- ☐ Which of these methods is the strongest? Why?
- ☐ Which approaches are weak?

□ Are there ways you can tell if the assumptions for these models are valid?

## **October 29 (Session 8) Instrumental Variables and Two stage Least Squares**

Angrist and Pischke Chapter 3

Khandker, Koolwal, Samad, Chapter 6

Joshua D. Angrist and Alan B. Krueger. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives*, Fall 2001. Pp. 69-85.

Michael p. Murray. "Avoiding Invalid Instruments and Coping with Weak Instruments." *Journal of Economic Perspectives*, Fall 2006. Pp. 111-132.

## **November 5 (Session 9) Synthetic Control Groups, Performance Measurement**

### **Synthetic Control Groups**

Robert McClelland and Sarah Gault (2017). "The Synthetic Control Method as a Tool to Understand Policy." Washington, DC: The Urban Institute.

Alberto Abadie et al. (2010). "Synthetic Control Methods for Comparative Case studies: Estimating the Effects of California's Tobacco Control Program." *Journal of the American Statistical Association* 105(490): 493-505. (Optional)

Justin M. Ross. (2017). "Unfunded Mandates and Fiscal Structure: Empirical Evidence from a Synthetic Control Model." *Public Administration Review*. 78(1): 92-103. (Skim)

Sian Mughan and Geoffery Propheter (2017). "Estimating the Manufacturing Employment Impact of Eliminating the Tangible Personal Property Tax: Evidence from Ohio." *Economic Development Quarterly* 31(4): 299-311. (Skim)

### **Performance Measurement**

*Moneyball (optional)*

Poister Chapter 5 in Newcomer et al.

Ann B. Blalock and Burt S. Barnow. "Is the New Obsession with 'Performance Management' Masking the Truth about Social Programs?" in Dall Forsythe editor. *Quicker, Better Cheaper: Managing Performance in American Government*. Albany, NY: Rockefeller Institute Press, 2001.

Burt S. Barnow. "The Role of Performance Management in Workforce Investment Programs" in *The Workforce Investment Act: Implementation Experiences and Evaluation Findings*. Douglas Besharov and Phoebe Cottingham, editors. Kalamazoo, MI: Upjohn Institute for Employment Research, 2011.

Burt Barnow and Carolyn Heinrich (2010). "One Standard Fits All? The Pros and Cons of Performance Standard Adjustments." *Public Administration Review*, January/February 2010.

GAO overview on GPRA Modernization (Skim)

IBM Center for Excellence in Government explanation of GPRA modernization act at <http://www.scribd.com/doc/47464749/GPRA-Modernization-Act-of-2010-Explained> (SKIM)

- ☐ What is the synthetic control method?
- ☐ When is it useful?
- ☐ What are the benefits and limitations of the synthetic control method?
- ☐ What is performance measurement?
- ☐ What is program monitoring?
- ☐ What are the challenges to measuring performance?
- ☐ What is meant by performance management?
- ☐ How might performance measurement and program evaluation be effectively coordinated?
- ☐ Should performance standards be adjusted?
- ☐ What is the "balanced score card?"
- ☐ What are community indicators and how do they relate to governmental performance measurement?
- ☐ What bad incentives are sometimes produced by performance measurement systems? What can be done?
- ☐ What are the lessons from Moneyball for performance measurement (and baseball)?

### **November 12 (Session 10) Implementation and Process Studies**

Holcomb, P.A. and Nightingale, D.S. 2003. Conceptual Underpinnings of Implementation Analysis. In *Policy into Action: Implementation Research and Welfare Reform*, eds. Mary Clare Lennon and Thomas Corbett. Washington, DC: Urban Institute Press, 39-55.

John Trutko, "Process Evaluation." Chapter 2 in *Evaluability Assessment of Responsible Fatherhood Programs*. (peruse the rest of the volume for a look at a comprehensive evaluability assessment)

Krueger and Casey Chapter 20 in Newcomer et al. (skim)

Goodrick and Rogers Chapter 22 in Newcomer et al.

Questions:

- ☐ How should formative evaluations be designed?
- ☐ How do you measure program implementation?
- ☐ How should feedback be incorporated in an implementation study?
- ☐ How should an implementation study be linked with an outcome study?
- ☐ What are common problems that threaten impact evaluations, even in RCTs?

### **November 19 (Session 11) Ethics, Minimum Detectable Effects, and Power**

#### **Ethics**

Kimmel book

The Belmont Report

Jan Blustein. "Toward A More Public Discussion of the Ethics of Federal Social Program Evaluation." *Journal of Policy Analysis and Management*, Fall 2005. pp. 823-852. Also comments by Rolston, Schochet, and Barnow are required

#### **Minimum Detectable Effects and Statistical Power**

MDRC reading on minimum detectable effects on Blackboard

Larry L. Orr. *Social Experimentation: Evaluating Public Programs with Experimental Methods*. Part 4: Sample Design.

Also read spreadsheet on determining minimum detectable effects that I have posted.

Questions:

- ☐ What protections should be given to participants in an evaluation?
- ☐ What procedures are possible in ensuring confidentiality?
- ☐ What procedures can be developed for maintaining the credibility and fairness of the evaluation?
- ☐ What are the essential elements and desired format for informed consent agreements?
- ☐ What are Institutional Review Boards and why are they important?
- ☐ What special ethical problems does random assignment pose?
- ☐ What is the concept of minimum detectable effect, and how does it relate to statistical power?

- How can you use MDE to determine whether a particular sampling design is adequate?
- How can you use MDE to determine how large a sample you should select for an evaluation?

## **November 26 (Session 12) Evidence for Policy Making**

Abigail A. Fagan, Koren Hanson, J. David Hawkins, and Michael W. Arthur (2008). “Bridging Science to Practice: Achieving Prevention Program Fidelity in the Community Youth Development Study.” *American Journal of Community Psychology*. 41: 235-249

Commission on Evidence-Based Policymaking (2017). *The Promise of Evidence-Based Policymaking*.

GAO Report “Program Evaluation: A Variety of Rigorous Methods Can Help Identify Effective Interventions” (GAO -10-30) (Skim)

OMB May 18, 2012 Memorandum on Evidence-Based Budgeting

Students should visit one or more of the following sites and assess the approach used by the site:

1. IES What Works Clearinghouse (education) <http://ies.ed.gov/ncee/wwc/>
2. Office of Juvenile Justice and Delinquency Prevention Programs OJJDP Model Programs Guide  
[http://www.crimesolutions.gov/about\\_starttofinish.aspx](http://www.crimesolutions.gov/about_starttofinish.aspx) and
3. Agency for Healthcare Research and Quality, Effective Health Care  
<http://www.effectivehealthcare.ahrq.gov/index.cfm/what-is-comparative-effectiveness-research1/>
4. The Campbell Collaboration <http://campbellcollaboration.org/>
5. Clearinghouse for Labor Evaluation and Research (CLEAR)  
<http://clear.dol.gov/>

Questions:

- What is evidence-based policymaking?
- What is evidence?
- Is all evidence of equal importance?
- How do we determine how much weight to give to each piece of evidence?

## **December 3 (Session 13)**

### **Systematic Reviews and Meta-Evaluation**

Readings:

Boruch, Petrosino, and Morgan Chapter 25 in Newcomer et al.

AND for systematic reviews, go to  
<http://homvee.acf.hhs.gov/document.aspx?rid=4&sid=19>  
And click your way through the review process.

Questions:

- ☐ What is meta-evaluation and how is it best conducted?
- ☐ What is “evidence-based” policy/management/practice?
- ☐ When are findings from evaluations sufficient to constitute such “Evidence?”
- ☐ Why is it difficult to transfer evaluation and research findings into practice?
- ☐ What is practice-based evidence?
- ☐ What is an evidence-based systematic review and how does it differ from a literature review?

### **December 10 (Session 14)**

#### **Utilization of Evaluation Results and Course Overview**

Wholey Chapter 30 in Newcomer et al.

Hatry, Newcomer, and Wholey Chapter 31 in Newcomer et al. optional.

David Greenberg, Marvin Mandell, and Matthew Onstott. “The Dissemination and Utilization of Welfare-to-Work Experiments in State Policymaking.” *Journal of Policy Analysis and Management*. Vol. 19, 2000, pp. 367-382.

Charles E. Metcalf. "Presidential Address: Research Ownership, Communication of Results, and Threats to Objectivity in Client-Driven Research." *Journal of Policy Analysis and Management*. Vol. 17, 1998, pp. 153-163. (optional)

Politics of Knowledge (several short articles) (optional)

Questions:

- ☐ What factors influence utilization of evaluation results?
- ☐ What are the various types of utilization? How can they be measured?
- ☐ What can be done during evaluation design and implementation to enhance utilization?
- ☐ What are emerging and continuing significant issues in the evaluation profession?

**December 10: Applied Evaluation Project due – hard copy and electronic copies to my office by 6pm.**

## **Policies in the Trachtenberg School Courses**

1. Incompletes: A student must consult with the instructor to obtain a grade of I (incomplete) no later than the last day of classes in a semester. At that time, the student and instructor will both sign the CCAS contract for incompletes and submit a copy to the School Director. Please consult the TSPPPA Student Handbook or visit our website for the complete CCAS policy on incompletes.
2. Submission of Written Work Products Outside of the Classroom: It is the responsibility of the student to ensure that an instructor receives each written assignment. Students can submit written work electronically only with the express permission of the instructor.
3. Submission of Written Work Products after Due Date: Policy on Late Work: All work must be turned in by the assigned due date in order to receive full credit for that assignment, unless an exception is expressly made by the instructor.
4. Academic Integrity: I personally support the GW Code of Academic Integrity. It states: "Academic dishonesty is defined as cheating of any kind, including misrepresenting one's own work, taking credit for the work of others without crediting them and without appropriate authorization, and the fabrication of information." For the remainder of the code see: <http://www.gwu.edu/~ntegrity/code.html>  
Note especially the definition of plagiarism: "intentionally representing the words, ideas, or sequence of ideas of another as one's own in any academic exercise; failure to attribute any of the following: quotations, paraphrases, or borrowed information." Whenever you make use of the words or ideas of others, it is important to reference the work consulted. All examinations, papers, and other graded work products and assignments are to be completed in conformance with the George Washington University Code of Academic Integrity.
5. Changing Grades After Completion of Course: No changes can be made in grades after the conclusion of the semester, other than in cases of clerical error.
6. The Syllabus: This syllabus is a guide to the course for the student. Sound educational practice requires flexibility and the instructor may therefore, at her/his discretion, change content and requirements during the semester.

## **SUPPORT FOR STUDENTS OUTSIDE THE CLASSROOM**

### *DISABILITY SUPPORT SERVICES (DSS)*

Any student who may need an accommodation based on the potential impact of a disability should contact the Disability Support Services office at 202-994-8250 in the Marvin Center, Suite 242, to establish eligibility and to coordinate reasonable accommodations. For additional information please refer to:  
<https://disabilitysupport.gwu.edu/>

*UNIVERSITY COUNSELING CENTER (UCC) 202-994-5300*

The University Counseling Center (UCC) offers 24/7 assistance and referral to address students' personal, social, career, and study skills problems. Services for students include:

- crisis and emergency mental health consultations
- confidential assessment, counseling services (individual and small group), and referrals

<https://counselingcenter.gwu.edu/>

## **SECURITY**

In the case of an emergency, if at all possible, the class should shelter in place. If the building that the class is in is affected, follow the evacuation procedures for the building. After evacuation, seek shelter at a predetermined rendezvous location.