

**Evaluating Development Programs****INTL\_ST 383-2**Andre Nickow – [a-nickow@kellogg.northwestern.edu](mailto:a-nickow@kellogg.northwestern.edu)TUESDAY & THURSDAY [**Example days**]11:00am-12:20pm (CST) [**Example time**]Office hours:Mondays 1:00-3:30pm (CST) [**Example time**]

*Readings are subject to change. Please be sure to pay attention to any announcements that go out on email or via Canvas.*

This course builds a systematic and intuitive grasp of evaluation as applied to economic and social development programs. Students will be provided with the tools to become informed and critical consumers of evaluations, and those with the inclination to do so will receive stepping stones to develop skills for producing evaluations. They will engage with debates on the strengths and weaknesses of alternative approaches to evaluation and consider how different combinations of methods may be best suited to particular research questions or practical goals.

The first section of the course focuses on the building blocks of evaluation research, with a focus on global development. We begin by exploring a range of outcomes that development projects attempt to influence. Prominent examples include poverty, consumption, income, assets, physical and mental health, education, intra-household dynamics, empowerment, social capital, governance, service delivery, and environmental impact. Underlying each is a complex set of conceptual and measurement challenges. We then motivate the role of evaluation in development by exploring its primary purposes, including choosing between alternative intervention models, performance monitoring, learning, and adaptation, and contributions to social science. Additionally, this section of the course will explore the connections between evaluation and theory, including theories of change that structure evaluation designs as well as the broader social theories from anthropology, economics, political science, and sociology that shape hypotheses and assumptions surrounding evaluations as well as the programs they evaluate.

The second section of the course focuses on evaluation design, with discussions of measurement and validity providing a bridge from the first section. In particular, we will explore the measurement strategies that form the building blocks of evaluations, including quantitative surveys and direct measurement (e.g., audits and satellite imagery) as well as qualitative interviews, focus group discussions, and observation. Next, we will explore how these building blocks are brought together to create evaluations.

Finally, we explore the role of evaluation in the wider world, including building partnerships, managing fieldwork, building partnerships, and synthesizing research implications. We will end by exploring the political economy of

evaluation, including the institutional and cultural forces that shape evaluation and the range of consequences that evaluations exert on the world of development praxis.

### **COURSE OBJECTIVES**

This course attempts to prepare students to:

- Critique evaluations by identifying the research design, unpacking methodological tools, and assessing how and to what extent the evaluation's design and analysis succeed in soundly addressing the research question through rigorous analysis of the evaluation's assertions, assumptions, evidential basis, and utility (both scientific and practical)
- Understand key workhorse quantitative and qualitative methodological tools used in contemporary development economics evaluations, elaborate their strengths and weaknesses in context, and identify the most appropriate methodologies to understand human and institutional behavior and other outcomes
- Identify and show understanding of the social science theories and theoretical frameworks and assumptions—as well as the power relations embedded in these—that (often implicitly) underpin an evaluation and attempt to elucidate the influence of economic, political, and socio-cultural power on individual and group outcomes
- Interpret evaluations—including the forces shaping them and the effects they exert on development praxis—within the social, political, environmental, and cultural systems that give rise to target populations, programs, and evaluations, including recognition of the reciprocal relationships between broader societal forces on one hand and the subjective psycho-social conditioning of evaluators on the other
- Develop skills for foregrounding multidimensional social phenomena like colonialism, diaspora, education, empire, the environment, ethnicity, health, indigeneity, immigration, migration, nationality, refugees, cultural reception, sustainability, statelessness, and war in the interpretation and design of evaluations.

### **EVALUATING STUDENT PERFORMANCE**

Student performance will be evaluated based on class discussions, two evaluation synthesis papers, and one evaluation research design. Class discussion will include thinking of discussion questions and bringing them to class, as well as reflecting on the course topics and readings in dialogue with other students. Thorough completion of course readings will be an important prerequisite for successful participation in discussions. The two evaluation synthesis papers will involve identifying a series of evaluations on a related topic (with the topic and final evaluation list to be approved by the instructor in advance), summarizing the key methods, findings, and assumptions of each evaluation, and adjudicating between the evaluations to elaborate implications for praxis in light of the evaluations' respective strengths and weaknesses. The evaluation research design will include a high-level plan for a development evaluation that integrates multiple methodological tools discussed in the class, with realistic budgetary constraints taken into account. Depending on student interest and logistical practicalities, the

research design project can be formulated as an individual or small-group project. The evaluation synthesis papers and evaluation research design will require in-depth familiarity and engagement with material covered in course lectures, discussions, and readings.

**Participation in class:** 25%

**Evaluation synthesis paper 1:** 25%

**Written assignments:** 25%

**Evaluation research design:** 25%

## COURSE BREAKDOWN

### Week 1, Session 1: Course Overview and Introduction to Development Outcomes

This session will begin with an orientation to the course and an overview of the topics to be covered and main requirements. It will then move into a discussion of some of the main physical, social, and economic realities that development programs aim to impact and that evaluations aim to measure and make sense of. With just enough philosophical discussion of the meaning of development to orient the practicalities, we will focus on a range of relatively concrete outcomes rather than abstract ontology. To paraphrase a popular title form, we will ask: what are we talking about when we talk about development impact?

Readings:

- Drèze, Jean. (2016). *Sense and Solidarity*. "Introduction: Economics Among the Road Scholars." Permanent Black.
- Freire, Paulo. (1978). *Pedagogy of the Oppressed*. Selections.
- Banerjee, Abhijit & Duflo, Esther. (2007). "Economic lives of the poor." *Journal of Economic Perspectives* 21(1): 141-167.
- Krishna, Aniruddh. (2010). *One Illness Away*. Selections. Oxford University Press.
- Ngozi Adichie, Chimamanda. "The Danger of a Single Story." TEDGlobal 2009. Speech presented at the TEDGlobal 2009, July 2009 [8:40]
- Escobar, Arturo. (1995). *Encountering Development*. Chapter 1. Princeton University Press.

### Week 1, Session 2: Motivating Development Evaluations

Having explored the outcomes that development programs most commonly attempt to shape, we will define the domain of "evaluation" and situate it within the broader world at the intersection of development practice and social science research. We will discuss and broadly define several types of evaluations—such as impact, process, and formative evaluations—and discuss the diverse purposes they serve for development actors including government agencies, civic nongovernmental organizations, businesses, and social movements.

Readings:

- Freire, Paulo. (1968). *Pedagogy of the Oppressed*. Selections. Bloomsbury.
- Glennerster, R. & Takavarasha, K. (2013). *Running Randomized Evaluations*. "Chapter 1: The Experimental Approach." Princeton University Press.

- Patton, Michael Quinn. (2014). *Qualitative Research and Evaluation Methods*. Selections. Sage Publications.
- Patton, M. Q. (2017). Pedagogical principles of evaluation: Interpreting Freire. In M. Q. Patton (Ed.), *Pedagogy of Evaluation. New Directions for Evaluation*, 155, 49–77.

### **Week 2, Session 1: Development Theory, Social Science, and Evaluation**

This session will explore implicit and explicit ways that social science theory structures evaluation. We will look at where the field of evaluation stands relative to traditional social science disciplines (e.g., economics, political science, sociology, and anthropology) as well as frameworks—normative and analytic—that cut across disciplines (e.g., neoclassical, Marxist/conflict, feminist, and environmental). We will also explore concepts, constructs, and validity, as well as causal inference with a focus on the potential outcomes framework.

Readings:

- Banerjee, Abhijit & Duflo, Esther. (2011). *Poor Economics*. “Forward” and “Chapter 1: Think Again, Again.” PublicAffairs.
- Easterly, William. (2014). *The Tyranny of Experts*. Selections. Basic Books.
- McGovern, Michael (2011) “Popular Development Economics--An Anthropologist among the Mandarins.” *Perspectives on Politics*, 9(2): 345-355.
- Robinson, James A. (2019) “How different social scientists think.” Unpublished manuscript.

### **Week 2, Session 2: Program Theories and Measurement**

Beginning with the core social science methodology notions introduced in the previous session, this course will introduce program theories of change, i.e., the variety of ways that theories of change for development programs are formulated and depicted. This will include a discussion of causal links and mediation, background assumptions, moderators, and necessary/sufficient conditions. Additionally, students will be introduced to measurement and instruments, including a basic overview of quantitative survey measurement, direct measurement (e.g., audits, satellite data, and crop cuts), and qualitative measurement.

Readings:

- Rogers, Patricia. (2014). *Theory of change: Methodological briefs-impact evaluation*. 2(747).
- Glennerster, R. & Takavarasha, K. (2013). *Running Randomized Evaluations*. “Chapter 5: Outcomes and Instruments.” Princeton University Press.
- Goertz, Gary. (2020). *Social Science Concepts*. Selections. Princeton University Press.
- Marchais, G., Bazuzi, P., & Amani Lameke, A. (2020). “‘The data is gold, and we are the gold-diggers’: Whiteness, race and contemporary academic research in eastern DRC.” *Critical African Studies*, 12(3), 372–394.
- King, Julian. (2021). "Expanding theory-based evaluation: Incorporating value creation in a theory of change." *Evaluation and Program Planning* 89: 101963.

**Week 3, Session 1: Introduction to Impact Evaluations**

Having established the building blocks of development evaluation research, we will begin to focus in on survey-based impact evaluations as the reigning hegemonic core approach to development. We will explore how surveys and other quantitative measurement instruments can be used to estimate the causal impact of programs. This session will include a discussion of the types of questions that impact evaluations are well- vs. poorly-suited to address.

## Readings:

- Glennerster, R. & Takavarasha, K. (2013). *Running Randomized Evaluations*. “Chapter 3: Asking the Right Questions” and “Chapter 7: Threats.” Princeton University Press.

**Week 3, Session 2: Randomized Trials**

This session will focus in on randomized controlled trials (RCTs), a family of designs that represents the emblematic approach to impact evaluation and has transformed the face of development since the late 1990s. We will explore the rise and proliferation of RCTs over the past three decades, accolades and criticisms, and areas of overlap and distinction between RCTs on one hand and quasi-experiments on the other (since critiques of these fields tend to convolute the two). The Banerjee & Duflo chapter provides an exemplary review in which a series of RCTs are used to support a robust policy implication.

## Readings:

- Glennerster, R. & Takavarasha, K. (2013). *Running Randomized Evaluations*. “Chapter 2: Why Randomize?” and “Chapter 4: Randomizing.” Princeton University Press.
- Banerjee, Abhijit & Duflo, Esther. (2011). *Poor Economics*. “Forward” and “Chapter 3: Low-Hanging Fruit for Better (Global) Health.” PublicAffairs.
- Banerjee, Abhijit et al. (2015). “A multifaceted program causes lasting progress for the very poor.” *Science*. 348(6236), 1260799. Guest speaker(s):

**Week 4, Session 1: Quasi-Experiments, Historical Research, and Time Horizons**

While RCTs represent the purest form of impact evaluations—the hegemonic approach to development evaluation—they can be costly and disruptive. This session will explore several examples of quasi-experiments, which can under some circumstances come close to replicating RCTs with greater efficiency. Additionally—particularly given that many development trends play out over very long time horizons, we consider the role of historical research in policy-oriented development evaluation and also distinguish define the concept of impact trajectories.

## Readings:

- Woolcock, Michael. (2009). “Towards a Plurality of Methods in Project Evaluation: A Contextualised Approach to Understanding Impact Trajectories and Efficacy.” *Journal of Development Effectiveness* 1(1): 1-14.

- Bayley, C.A., Rao, Vijayendra, Szreter, Simon & Woolcock, Michael, (Eds.). (2011). *History, Historians, and Development Policy*. Selections. Manchester University Press.

#### **Week 4, Session 2: Qualitative Methods**

This session shifts gears from quantitative to qualitative methodologies. We will discuss the division of labor between qualitative and quantitative methods in development research, including differences in priorities and ways that the two streams can function synergistically. Additionally, we will discuss core qualitative methods, including semi-structured interviews, focus group discussions, and observations, as use for measurement and causal inference.

Readings:

- Patton, Michael Quinn. (2014). *Qualitative Research and Evaluation Methods*. Selections. Sage Publications.
- Krishna, Aniruddh. (2010). *One Illness Away*. Selections. Oxford University Press.
- Widner, Jennifer, Woolcock, Michael, and Nieto, Daniel Ortega. (2022). *The Case for Case Studies*. Selections. Cambridge University Press.

#### **Week 5, Session 1: Process Evaluations and Monitoring**

With key quantitative and qualitative measurement and causal inference techniques under our belt, we will go deeper into the types of evaluation aside from impact evaluation discussed in the second session of week one – including varieties of process and formative evaluations. We will explore how these evaluation types can supplement impact evaluations and serve purposes of their own. A key example will be the World Bank’s Social Observatory.

Readings:

- Patton, Michael Quinn. (2014). *Qualitative Research and Evaluation Methods*. Selections. Sage Publications.

#### **Week 5 Session 2: Mixed Methods Designs**

Development researchers have developed a variety of innovative modes of combining quantitative and qualitative research questions to maximize synergy in improving the validity and interpretability of results. This session will explore areas of potential alongside risks of muddled results, using some examples that have made it close to the mainstream of development research: the “participatory econometrics” and “Q-squared approach”.

Readings:

- Rao, Vijayendra. (2002). “Experiments in 'Participatory Econometrics.'” *Economic and Political Weekly*: 1887-1891.
- Shaffer, Paul. (2013). *Q-squared: Combining Qualitative and Quantitative Approaches in Poverty Analysis*. Selections. Oxford University Press.

- Bamberger, Michael, Vijayendra Rao, & Woolcock, Michael. (2010) "Using mixed methods in monitoring and evaluation: experiences from international development." *World Bank Policy Research Working Paper* 5245.

### **Week 6, Session 1: Field Issues, Part I**

Having gained exposure to the major categories of evaluation design used in development, we will turn to practical issues in implementing evaluations. This session will focus in particular on research ethics—including deciding when an RCT is appropriate, and when an evaluation is appropriate at all—as well as building partnerships with government agencies, nongovernmental organizations, businesses, and other development actors. Additionally, development evaluations—like the programs they analyze—are often complex operations involving numerous people from different groups. We will therefore explore issues that come up in field management.

Readings:

- White, H., & Raitzer, D. A. (2017). *Impact evaluation of development interventions: A practical guide*. Asian Development Bank.
- Gertler, Paul J., et al. (2016). *Impact Evaluation in Practice*. Selections. World Bank Group.

### **Week 6, Session 2: Scaling Programs, Scaling Evaluations**

Development programs often work differently at small scales—e.g., when run by a dedicated nonprofit working only in certain areas—than they do at larger scales or in more diverse settings. A key challenge for development policy is identifying program approaches that work at scale. Evaluators must also deal with these issues, both in that they must sometimes evaluate programs at large scales or otherwise make inferences about scaling, and also because evaluations face similar scaleup challenges. Moreover, findings appear to systematically differ depending on who is implementing the evaluation and who is implementing the program. This session will explore key issues of scaling. The debate between Duflo and Drèze on values and priorities in political economy form a bridge between this session and the following one on political economy, opening a series of potential critiques of the technocratic approach to development embodied in the hegemonic impact evaluation perspective and pointing toward the ever-present importance of power and struggles over the distribution of resources—as well as values.

Readings:

- List, John A. (2022). *The Voltage Effect*. Selections. Crown Currency.
- Duflo, Esther. (2020). "Field Experiments and the Practice of Policy." *American Economic Review*. 110(7): 1952-1973.
- Drèze, Jean. (2020). "Policy Beyond Evidence." *World Development* 127: 104797

### **Week 7, Session 1: Evaluating Political Economy – and the Political Economy of Evaluations**

This session builds on the challenges of evaluating at scale introduced in the previous section and raises some challenges associated with evaluating political economy programs and examples of how researchers have gotten around them. The session will pair these issues with distinct but related considerations in the political economy of

evaluations. Here we will explore the systems that create evaluations: the overlapping sets of organizational actors who, through their collaboration, competition, and negotiations fund, commission, produce, review, critique, and consume evaluations. We will also begin an exploration of how evaluations are used, e.g., adjudicating between multiple policy options, fine-tuning programs, or simply signaling to stakeholders that something is being done about quality control—what organizational sociologists refer to as “myth and ceremony”.

**Readings:**

- Banerjee, Abhijit & Duflo, Esther. (2011). *Poor Economics*. “Chapter 7: The Men from Kabul and the Eunuchs of India” and “Chapter 10: Policies, Politics, and Political Economy.” PublicAffairs.
- Fanon, Franz (1952) *Black Skin, White Masks*. Selections. Grove Press.
- Nunn, Nathan and Nancy Qian (2014) “US Food Aid and Civil Conflict,” *American Economic Review*, VOL. 104, NO. 6, (pp. 1630-66).
  
- Andersen, Jorgen Juel Johannesen, Niels; Rijkers, Bob. 2020. “Elite Capture of Foreign Aid : Evidence from Offshore Bank Accounts.” Policy Research Working Paper;No. 9150. World Bank

**Week 7, Session 2: Research Synthesis and Policy Inference**

Virtually no decisions on development practice should be made based on the results of a single evaluation. Instead, evaluations contribute to bodies of evidence that, once they become sufficiently robust, can be leveraged in partial and contextual ways. This session will explore informal and formal methods for aggregating evaluation evidence and using it to make policy inferences.

**Readings**

- Glennerster, R. & Takavarasha, K. (2013). *Running Randomized Evaluations*. “Chapter 9: Drawing Policy Lessons.” Princeton University Press.
- Widner, Jennifer, Woolcock, Michael, and Nieto, Daniel Ortega. (2022). *The Case for Case Studies*. “Will it work here? Using case studies to generate 'key facts' about complex development programs.” Cambridge University Press.
- Woolcock, Michael. (2013). “Using case studies to explore the external validity of ‘complex’ development interventions.” *Evaluation*, 19(3): 229-48.

**Week 8, Session 1: Politics, Civic Movements, and Action Research**

Although development evaluations prototypically cater to development policymakers in government agencies, nongovernmental organizations and, to a lesser extent, for-profit firms that style themselves as “social enterprises”, a much wider array of institutional and extra-institutional actions may use evaluation to improve their ends. This session explores ways that evaluations have and are being used to foment social change that is more radical than the orientations of mainstream, technocratically-oriented development evaluations.

**Readings**

- Khemani, Stuti et al. (2016). *Making Politics Work for Development*. Selections. World Bank Group.



- Heller, Patrick & Rao, Vijayendra (Eds.). (2015). *Deliberation and Development*. Selections. World Bank Group.
- Cornish, Flora, et al. (2023). "Participatory action research." *Nature Reviews Methods Primers*, 3(1): 34.
- Thomas, Emma F., et al. (2022). "Mobilise: A higher-order integration of collective action research to address global challenges." *Political Psychology* 43 (2022): 107-164.

**Week 8, Session 2: Conclusion**

This final session will include discussions of the final research designs and debriefing on the course content and ways to move forward within the fields of development and/or evaluations.