Using RCTs in Economic Education Research

November 2024

Todd Pugatch Elizabeth Schroeder

Abstract: Randomized controlled trials (RCTs) have become an essential tool for economists. The credibility revolution in empirical economics emphasizes research designs that identify casual effects, and random assignment of treatment is seen as the gold standard. Implementation can, however, be a challenge in many applications. The field of economic education is in a unique position to learn from RCTs, given the ability to test interventions in the classroom or at educational institutions. We discuss what is needed to effectively run an RCT in an educational setting, drawing from the experimental literature on topics such as student success in higher education and diversity in undergraduate economics. We additionally outline quasi-experimental approaches that can be used when treatment cannot be randomized.

JEL Classification: I21; I23
Keywords: randomized controlled trial; economics of education; higher education

---

Pugatch: University at Buffalo, tpugatch@buffalo.edu.
Schroeder: Oregon State University, liz.schroeder@oregonstate.edu.

1 Introduction

Economists have a keen interest in evaluating the effectiveness of their own teaching methods and institutional policies, as shown by recent surveys of the literature on economic education (Allgood, Walstad, and Siegfried 2015; Fernandez, Yetter, and Holder 2021; Birdi et al. 2023). This literature includes studies of inputs to learning such as classroom games, experiments, and other active learning techniques. The use of classroom technology and online resources has been a topic of interest, as well as curricular changes that add new modules to courses or change how some topics are taught. Another strand of the literature looks at student decisions, for example choosing a college major or how and when to study, and how classroom policies affect those choices. Across these topics, issues of diversity and inclusion have also gained attention in recent years. Whether deciding to introduce a new classroom technology or encouraging more students to major in economics, social scientists with training in econometrics and causal inference have a variety of tools from which they can draw when assessing their efforts.

In many cases, when instructors or other decision-makers enact reforms to the way economics is taught or promoted at their institutions, it is possible to introduce these interventions in a way that allows evaluation of their causal impact. Given the ability to assign a new intervention to specific students or class sections, it is often possible to conduct a randomized controlled trial (RCT). This opportunity to cleanly establish treatment and control groups affords instructors and other researchers the ability to estimate the causal impact of a change. In this paper, we draw from the literature to highlight the potential of RCTs in economic education and provide a guide for effectively running an RCT in this context. We then briefly discuss other methods for causal inference that can be used when randomization is not possible.

2 RCTs in economic education

2.1 Why RCTs are the gold standard

We can never directly observe the effect of an intervention on an individual student. This is the fundamental problem of causal inference; we will never observe the counterfactual state in which a treated student did not receive the intervention, just as we will never see what would have happened to an untreated student had they received the treatment. Instead, we must compare outcomes for treated students to outcomes for some control group. Students who were not subject to a curriculum change or other new policy are unlikely to be a valid control group for students who did experience the reform, however, as both students and instructors are likely to select into treatment based on the benefits they expect.

For example, consider the problem of estimating the effect of active-learning classroom activities on student grades. We might expect that instructors who spend time and effort designing and implementing such activities were already among the most dedicated teachers, in which case their students may have done better than students in other classes even without the new activities. This type of selection bias will be present for any pedagogical intervention that is more likely to attract instructors who care the most about teaching, leading to overestimation of the effects of the interventions. At the institution level, initiatives to attract diverse students to economics are more likely to be undertaken by departments that already care about promoting diversity and inclusion. As a result, any specific initiative may appear more effective than it was if the comparison group is departments that did not focus on diversity. Students also have some ability to choose which course sections to take, in addition to their overall choice of major, leading to selection on the student side. In evaluating online learning or flipped classrooms, for example, students who are less engaged and have low attendance records might be more likely to choose a course section with an online component. In this case, negative selection into these sections could make the use of technology appear less effective. Even if the same instructor implements a new technique in some of their course sections but not others, it can be difficult to control for all other differences between the sections, such as class meeting time or the term during which each section meets.

RCTs solve the fundamental problem of causal inference by ensuring that there can be no selection into treatment assignment. The treatment and control groups should therefore be comparable on unobservable characteristics such as motivation, interest in economics, or commitment to equity. If this is true, then the control group is a valid comparison for the treatment group, and it is possible to estimate the effect of treatment by comparing the two groups. The advantages of using RCTs in economic education research can be illustrated with two branches of the literature: the literature on student success and the literature on interest and diversity in economics.

2.2 Student success

Universities offer a variety of services aimed at improving student academic outcomes. Tutoring services and academic coaching can not only help with course material, but may also assist students in overcoming procrastination, setting goals, and learning to study effectively. Evaluation of these types of services faces a selection problem, however, as students who choose to use support services are likely to differ in a variety of ways from students who do not. Students may seek academic support because of low grades, inducing negative selection and making the programs appear less successful. Alternatively, students who seek services may be particularly motivated to do well, making such programs appear more effective than they are. Drawing from the behavioral economics literature, several papers have studied light-touch interventions, or nudges, that remind or inform students about available academic services and best study practices. Introductory economics classes are often large, so reaching these students with methods that are low-cost and do not require a much extra time from instructors are particularly attractive. In addition, nudges such as email reminders are straightforward to randomize, meaning education researchers are able to study both the effectiveness of nudges and, in some cases, induce exogenous variation that can be used to study the impact of the services that are being recommended.

RCTs in this literature typically randomize treatments among students in economics classes and combine administrative data on student outcomes with survey data obtained by instructors. One large-scale effort took place at the University of Toronto's Student Achievement Lab, which collaborated with instructors of first-year economics courses. These instructors teach around 5,000 students per year, resulting in a sample of 25,000 students over a five-year period. The instructors set aside a small portion of their grade requirements for students to complete one- to two-hour surveys at the beginning of the school year. Student were then randomly assigned to treatment or control groups. Researchers linked the survey data to university administrative records on academic outcomes and conducted follow-up surveys about other outcomes such as study habits or mental health. They designed and tested a variety of interventions using emails, texting, online coaching, and face-to-face coaching. Overall, none of the interventions produced significant improvements in student outcomes such as grades or retention. The randomized designs allowed the researchers to look for heterogeneity in the effectiveness of different interventions, however, and they find some positive effects of in-person coaching on time spent studying, but not enough to impact grades. Lighter-touch interventions such as online planning exercises and text or email nudges were less effective (Oreopoulos and Petronijevic 2018; Oreopoulos et al. 2019; Oreopoulos and Petronijevic 2019).

A series of RCTs at Oregon State University similarly worked with instructors to have students in introductory economics courses complete surveys as a course assignment at the beginning and end of the term. The researchers randomly assigned approximately 2,100 students to treatment and control groups, collected data on whether students opened treatment emails and clicked on links within those emails, and records of visits to academic support services, use of economics tutoring, and completion of extra practice problems. Email and text messages encouraging the use of academic support services were effective at encouraging students to complete extra practice problems, but less effective at encouraging them to attend academic coaching or economics tutoring on campus. Effects on grades were limited. Here again, the study design allowed the authors to compare effectiveness of different types of messaging. They found evidence that emails were more effective than text messages; messages were more effective later in the term; and students began to experience message fatigue after two messages (Pugatch and Wilson 2024). Another experiment designed a commitment contract that was randomly offered among nearly 900 introductory economics students. By signing up for the contract, students committed to attend tutoring if their midterm grades were sufficiently low. The authors found about 10 percent of students were willing to pay for such commitment devices, in the form of foregone chances to win a textbook scholarship if they failed to comply with the contract. However, contract take-up did not lead to significant increases in tutoring attendance or course grades (Pugatch, Schroeder, and Wilson 2024). Other RCTs find similar results. At University of North Carolina at Chapel Hill, randomly assigned email messages to remind students of the availability of economics support programs on campus increased use of academic support services among women, first-year students, and students of color, but did not have an effect on grades (Balaban and Conway 2020). At Florida Gulf Coast University, a commitment pledge designed to help introductory economics students overcome procrastination led to more time spent studying, but overall test scores did not improve (Wright, Arora, and Wright 2023).

In an intervention focused on underrepresented college students, Carrell and Kurlaender (2023)recruited faculty members from 20 different course subjects to participate, resulting in around 3,000 students. Faculty sent personalized emails to randomly assigned students about their course performance and ways to improve. Not only did the treatment improve student perceptions of their professors, it also led to higher grades. An experiment at West Point randomized 551 students into either online or in-person sections of introductory economics, finding significantly lower grades for online students (Kofoed et al. 2024).

2.3 Interest and diversity in Economics

Students self-select into courses and majors based on their preferences and beliefs. Efforts to attract students to economics by providing information to current students or introducing new material into introductory classes must therefore be carefully designed if we wish to evaluate their causal impact. In recent years, a growing literature has focused on increasing diversity in undergraduate economics by attracting more students from underrepresented groups. Much of this work was motivated by the Undergraduate Women in Economics (UWE) Challenge, founded by Claudia Goldin and Tatyana Avilova in

2014 to learn about why women are underrepresented in economics and what types of interventions might narrow the gap (Avilova and Goldin 2018; 2024). The challenge included 88 schools, of which 20 were randomly chosen to receive grants aimed at increasing the number of female economics majors. Each treatment school designed its own intervention, and some implemented their own individual RCTs, providing evidence on information nudges, role models, and mentoring. Much like the literature on student success, students in the studies discussed here are matched to administrative data on the courses they take in subsequent terms and the major they declare, and often take surveys at the beginning and end of the term as part of class assignments or extra credit.

One branch of this literature looks at informational nudges by randomly assigning students to receive email messages containing different types of information about economics. At the University of Illinois Urbana-Champaign, Halim, Powers, and Thornton (2022)sent messages to students in introductory economics classes that emphasized either the range of careers or the earnings potential of majoring in economics. Similar studies that were not part of the UWE Challenge were conducted at Oregon State University (Pugatch and Schroeder 2021; 2024) and among incoming underrepresented minority and female students across nine liberal arts colleges (Bayer, Bhanot, and Lozano 2019). These studies find insignificant effects for female students, but the Oregon State studies find positive and significant effects of informational nudges on majoring in economics for first-generation college students.

Other interventions emerging from the UWE Challenge introduced female role models, mentoring for female students, or information about grade distributions. At Southern Methodist University, Porter and Serra (2020)randomly assigned sections of economics principles courses to receive career talks from female alumnae of the economics major. They found that female students who received the role model visits were significantly more likely to take more economics classes and to major in economics. At Colorado State University, Li (2018)randomly assigned different treatments across recitation sections of principles classes, providing information in class about economics, inviting female students to participate in peer mentoring, and sending an encouraging message to female students whose grades were above the class median. Information combined with the encouraging nudge increased the probability of majoring in Economics for female students whose grades were above the median. At the University of Colorado Boulder, Antman, Flores, and Skoy (2020)sent randomly assigned students a Qualtrics survey eliciting their beliefs about their relative performance in the class. For some treatment groups, they also provided information about their actual relative performance, contributing to our understanding that female students are more sensitive to grades.

3 What is needed to effectively run an RCT

This section covers how to effectively run an RCT in economic education. We do not intend to replace more comprehensive guides, but instead to list issues and tips for novice experimenters.[1]

3.1 Research question

As with any research project, the question is central. The design, execution, and ultimate success of your project follow from the research question(s). Begin with the "big picture" question: for instance, why do women major in economics at lower rates than men? Progressively narrow the question until it suggests an RCT which can contribute an answer. For instance, do women major in economics at lower rates than men because they are more sensitive to course grades? A single RCT cannot answer these questions. However, a single RCT might answer variations of the question, such as: does providing students with information about the completion rate of economics majors with their course grade change their subsequent course enrolment? Does the response differ between men and women? An RCT could answer these questions.

3.2 Intervention design

3.2.1 Types of interventions

A well-posed research question should lead naturally into the intervention design. In the example above, students might be randomly assigned to receive information about the completion rate of economics majors with their course grade (treatment group), while the control group does not receive this information. Variations on the treatment are also often possible. For instance, treatment group A might get this information, while treatment group B might get the information plus encouragement to continue taking economics courses.

The range of possible interventions in economic education is large, but constrained by resources and feasibility. We might wish to test whether full scholarships for economics majors or an overhauled economics curriculum influence behavior, but having access to scholarship funds or the patience for curricular overhaul is another matter. Instead, more scaled-down interventions – a lottery for a modest scholarship if one meets a grade threshold, or an alternative version of a popular course, for instance – might be possible.

Another common constraint in economic education RCTs concerns the control group. In many cases, denying students access to a treatment is not possible for regulatory or ethical reasons. For instance, a tutoring service must be open to all students, or all students must be eligible for the same extra credit

[1] For further reading on how to design and conduct RCTs, we recommend Orr 1999; Duflo, Glennerster, and Kremer 2007; Gerber and Green 2011; and Glennerster and Takavarasha 2014, and the J-PAL Research Resources webpage `povertyactionlab.org/research-resources` (J-PAL, n.d.).

assignments in a course. In these cases, interventions often take the form of *randomized encouragement designs*, in which the treatment is a nudge or other encouragement to change behavior, while the control group does not receive the encouragement. These designs expand the range of feasible interventions, but at the cost of changing what the RCT measures. For instance, instead of measuring the effect of *receiving tutoring* on grades, your RCT measures the effect of *receiving a nudge to attend tutoring* on grades.[2]

As discussed in Section 2, interventions tested via RCT in economic education often fall into several categories:

1. *Information:* treated students receive information. Control students receive different, less, or no information. Example: Pugatch and Schroeder (2021; 2024) and Halim, Powers, and Thornton (2022) randomly assigned students to receive different messages about studying economics, such as salary information and career information.
2. *In-class experiences*: treated students receive a different form of pedagogy than control students. Kofoed et al. (2024) randomly assigned students into in-person or remote sections of introductory economics.
3. *Support services*: treated students receive mentoring or other additional support. Balaban and Conway (2020) randomly assigned students to receive invitations to academic support sessions.
4. *Incentives*: treated students receive an incentive to change behavior. Pugatch and Wilson (2018) randomly assigned students to receive a voucher for the campus coffee shop if they attended tutoring.

3.2.2 Unit of randomization

A key consideration in intervention design is the *unit of randomization*, i.e., the level at which the intervention is assigned. The unit of randomization is distinct from the *unit of analysis*, which is the level at which the outcomes are measured. In economic education, the unit of analysis is often the student. For instance, Felkey et al. (2021) and Pugatch, Schroeder, and Wilson (2024) randomly assigned individual students the offer of commitment contracts to study more. However, the unit of randomization is not always the student. If all students in a group (for instance, a course section) are randomly assigned to treatment or control, then this group is the unit of randomization. For instance, Porter and Serra (2020) assigned course sections to visits from female role models. Experiments in which groups are the unit of randomization are called *cluster RCTs*.

[2] One might still estimate the effect of tutoring on grades via instrumental variables estimation, with the nudge instrumenting for tutoring attendance. The estimate then represents the local average treatment effect (LATE) for students who complied with the nudge.

Choosing between students and course sections as unit of randomization involves tradeoffs between statistical power, risk of spillovers, and feasibility:[3]

Table 1: tradeoffs between different units of randomization

	Statistical power	**Spillover risk**	**Feasibility**
Student	*High* Randomly assigning students within a course section to treatment and control holds constant all idiosyncratic factors common to the course section.	*High* Students can share information about their treatment status with section classmates, potentially contaminating the control group.	*Low* The type of intervention, academic regulations, or ethical concerns might prevent instructors from offering treatment to some students but not others within the same course section.
Course section	*Low* By definition, treatment assignment correlates with other attributes of the course section. Requires many course sections to balance these attributes between treatment and control groups.	*Low* Spillovers are limited to exchange of information about treatment status across course sections.	*High* The course section is often the natural unit to test an intervention, such as a new form of pedagogy.

The optimal unit of randomization will vary from case to case. In the case of the commitment contract in Pugatch, Schroeder, and Wilson (2024), the researchers prioritized power based on their uncertainty about take-up of the contract offer. They mitigated the spillover risk by preventing students in the control group from accessing the contract. In Porter and Serra (2020), female role models visited physical classrooms; splitting the classroom or scheduling individual visits with hundreds of students would be infeasible.

3.2.3 Power analysis

Suppose you are an economics instructor for a class of 50 students. You create a new interactive online simulation of a course concept and want to test if it can increase scores on your upcoming exam. You expect it will increase scores by 5 percentage points, from a mean score of 80 percent. You randomly

[3] Other units of randomization are possible. For instance, instructors could be randomly assigned to treatment and control across all their classes.

assign the simulation to half the class. Supposing the assignment is as effective as intended, how likely is your RCT to find a statistically significant effect?

This question can be answered via power analysis. Statistical power is the likelihood of finding a statistically significant effect when the effect exists. Equivalently, power is one minus the probability of Type II error (failing to reject the null hypothesis when the alternative hypothesis is true). In the example above, the RCT has only a 7 percent chance of finding the expected effect![4]

The example demonstrates the value of conducting power analysis before running an RCT. The research design seems reasonable, with a relatively modest treatment effect, a sample size exceeding the old undergraduate statistics rule of thumb of n = 30, and the benefits of student-level randomization. Yet to reach power of 80 percent – the common standard among RCT practitioners – would require a sample size of 1,812 students! When designing RCTs, power analyses are often humbling experiences.

The picture is even bleaker when conducting cluster RCTs. In our running example, suppose a sample of 1,850 students, split into 37 sections of 50 students each. But now suppose randomization is at the course section level, perhaps because you worry treated students will share access to the simulation with control students, or control students will complain about unfairness. Now instead of 80 percent power, as under student-level randomization, power falls to 14 percent![5] Why? When treatment is assigned to entire course sections, the influences common to a course section – instructor, time of day, classroom conditions, etc. – lead to correlated outcomes among students in the section. This makes it more difficult to disentangle the effect of treatment from the common influences of the section. The effective sample size is closer to the number of sections (37) than the number of students (1,850). Low power may explain the common finding of null effects on student grades in economic education interventions.

In addition to calculating power, power analysis can also calculate required sample sizes and minimum detectable effect sizes (MDEs). The `power` suite of commands in Stata or other statistical software can facilitate the analysis.

The sobering reality of power analysis leads to several rules of thumb when designing an economic education RCT:

1. *Whenever possible, randomize at the student level.* Some treatments might even allow within-student randomization, for instance randomly assigning different practice problems to the same student during a course.
2. *Increase sample size any way you can.* Include all sections of a course taught over a year, include online sections, recruit other departments or campuses, etc.
3. *Consider more proximate outcomes*. The more distal the outcome, the more difficult to detect effects, because of the lengthy causal chain from short to long run outcomes. Outcomes such

[4] We assume a two-sided test at the 5 percent level. Even under the more generous assumptions of a one-sided test at the 10 percent level, the RCT has only a 21 percent chance of detecting the effect.

[5] We assume an intracluster correlation of 0.2, a reasonable value in education interventions.

as graduation are slow to move and have long gestation periods. Outcomes such as take-up of an intervention are more immediate and amenable to change.

3.2.4 Concept notes

In addition to power analysis, we strongly recommend drafting a concept note before launching an RCT. In some cases, concept notes or similar documents are required by potential funders or implementation partners. Even when not required, concept notes help clarify thinking, improve the research design, or discover flaws which could weaken or derail the project. Gathering feedback on the concept note from collaborators and colleagues inevitably results in a better RCT than the first draft.

Concept notes vary in form and length. From our experience, one to four pages suffice, covering the following topics:

- Research question(s)
- Anticipated contribution to literature
- Description of intervention
    - Theory of change: why should the intervention have an effect?
- Experimental design, including:
    - Description of treatments
    - Unit and timing of randomization
- Data collection
    - Data sources (i.e., administrative, survey, observation)
    - Timing of data collection
- Analysis
    - Primary and secondary outcomes
    - Methodology, including main estimating equation(s)
- Project timeline

3.3 RCT logistics

3.3.1 Cooperation from implementing partners

Running an RCT takes a village. Researchers new to the method may be surprised to learn how much planning and maintenance are required to conduct a successful experiment. Unlike analysis of secondary data, RCTs require significant investment in project management. Securing and maintaining cooperation from implementing partners are necessary for success.

Implementing partners will vary across projects, but common partners in economic education RCTs include:

- *Instructors*: course instructors and teaching assistants are often responsible for implementing interventions and providing data.
- *University administrators*: the registrar, institutional research office, academic advisors and support services, information technology, or other administrators can provide data not available through other channels.
- *Department chairs*: department chairs can help convene other partners, identify available resources, and advocate for the importance of the project when required. Even if your department chair's cooperation is not strictly necessary, it is a good idea to keep them informed to avoid any detrimental actions, such as rescheduling instructors or sections after random assignment.

In most cases, you will rely on these implementing partners' voluntary cooperation. Treat them well, set clear expectations, and show appreciation for their efforts.

3.3.2 Funding

Many economic education RCTs can be conducted without funding. Pedagogical interventions and information experiments (particularly when using electronic communications) often have nearly zero marginal costs. Nonetheless, many RCTs require funding, and even those which are nominally zero cost could still benefit from funds to improve implementation or compensate participants.

Your university's research office is a good place to inquire about potential sources of funding, both internal and external. Sponsors of large research projects in higher education include the US Department of Education Institute of Education Sciences (IES), National Science Foundation, and Spencer Foundation. These funds tend to be highly competitive, but smaller funders also exist.

Be entrepreneurial when seeking funding. For instance, we (Pugatch and Schroeder 2021; 2024; Pugatch and Wilson 2024) won an internal grant competition from a technology innovation fund at Oregon State University, for our proposal to diversify and promote student success among economics students via electronic communications. Because preparing proposals takes considerable time, we also recommend developing a "Plan B" for how to conduct a (likely scaled-down) version of the project if the funding proposals are unsuccessful.

3.3.3 Setup

<u>Data collection</u>

Plan ahead for data collection. Your concept note should clarify the data necessary to analyze results of your RCT. Follow the causal pathway(s) hypothesized in your theory of change.

Data in economic education RCTs tend to come from two sources: administrative data and surveys. Administrative data are generally considered superior, as they are less prone to biases associated with self-reporting and less susceptible to attrition. However, institutions may be unable, unwilling, or slow to provide administrative data (about which more below). Plan accordingly.

Survey data provide more control and flexibility to researchers, as survey design can align with your project goals. College students are accustomed to completing surveys on Qualtrics and other online platforms. But like older adults, they are increasingly inundated with survey requests. Make it easy for students to respond by designing surveys which are clear, short, and accessible. Compensating students (for instance, with modest extra credit if in a classroom setting) helps increase response rates and reduce attrition.

Human subjects research approvals

Most economic education RCTs meet the formal definition of research with human subjects, and therefore must be approved by an Institutional Review Board (IRB). Your university's IRB can provide details on its requirements. If collaborating with colleagues at other institutions, check if your IRB allows for reliance on another IRB's determination. If so, you can avoid duplicated efforts across multiple IRBs.

In addition to IRB approval, institutions will have their own safeguards for use of student data. In the United States, education records are covered under the Family Educational Rights and Privacy Act (FERPA). Although FERPA is a federal law, its interpretation may vary across institutions. Your university may require separate agreements with entities providing education records, such as the registrar, to ensure FERPA compliance.

Trial registration and pre-analysis plan

It is increasingly common for social scientists to publicly register their RCTs, as required for medical trials. RCT registration promotes research transparency and a complete scientific record regardless of publication. The American Economic Association maintains an RCT Registry at `socialscienceregistry.org`. Registering an RCT takes about 10 minutes using information from your concept note.

A pre-analysis plan (PAP) is a document detailing your plan to analyze results from your RCT. Although it is an optional field in the AEA RCT Registry, the research community increasingly expects pre-analysis plans for RCTs, to combat selective reporting of results. Researchers remain free to report results outside the PAP, as long as they label these results "exploratory" and report all results specified in the PAP (Banerjee et al. 2020). Some researchers complete a PAP prior to any data collection, while others wait until data collection is complete but analysis has not yet begun.

3.3.4 Implementation

Before running your RCT, we strongly recommend running a pilot study. A pilot tests the intervention in a smaller sample and/or shorter duration than the full study. Pilots provide valuable information about obstacles to implementation, data collection, and other logistics. Due to their smaller sample size, pilots are usually underpowered to detect effects, so do not despair if your pilot reveals null effects. But be sure to adjust other aspects of your RCT based on lessons from the pilot.

Before the full RCT launches, clarify tasks and timelines among your collaborators and implementing partners. Unlike much other academic research in economics, RCTs run on firm deadlines. Investing in project management has high returns.

3.3.5 After the experiment

When the intervention concludes, the fun of data analysis begins. (In truth, you will likely wish to take some time off. You or a research assistant must also clean the data. But you are closer to the fun part than when you started!) This phase of RCTs is much like other empirical economics research, except that pre-analysis plans are more common for initial stages of inquiry, and implementing partners will be eager to learn the results. Honor the efforts of your implementing partners and other stakeholders by sharing results with them. Also ask for their feedback. Because these groups know the setting so well, they often deliver insights which can greatly enrich the completed research.

4 What to do when one cannot run an RCT

RCTs are not always feasible. The previous section underscores the extensive preparation and careful execution required to answer a research question in economic education via RCT. When an RCT is not possible, the applied economist toolkit might still provide the means to answer the same or similar research question. This section lists some of those tools, with examples.

4.1 Quasi-experimental approaches

1. *Natural experiments*. Sometimes random variation in treatment assignment exists without its deliberate creation via RCT. Canaan and Mouganie (2021) use random assignment of faculty academic advisors – the institutional policy at American University of Beirut, the setting of the study – to test whether advisor gender influences student outcomes. They find female economics majors randomly assigned to a female advisor drop out at lower rates and are more likely to graduate with an economics degree.
2. *Regression discontinuity (RD) designs.* In academic settings, resources or achievements are often rationed according to threshold rules. For instance, numerical course averages determine letter grades, or admissions to selective programs rely on grade point average (GPA) cutoffs. Such settings lend themselves to regression discontinuity (RD) designs, which compare outcomes among students just above and below the relevant cutoff. Using the GPA cutoff for economics majors at University of California-Santa Cruz, Bleemer and Mehta (2022) find that students who major in economics after scoring barely above the cutoff earn $22,000

more in average early career income than students just below the cutoff. Using course letter grade cutoffs, McEwan, Rogers, and Weerapana (2021) find that female students just above a cutoff are 18 percentage points more likely to major in economics than female students just below. These examples answer important research questions about treatments (majoring in economics, course letter grades) which could not be randomly assigned.

3. *Difference in differences (DD)*. Settings with panel or repeated cross sectional data on interventions and outcomes often use difference in differences to estimate treatment effects. Antman, Flores, and Skoy (2020) complement their information RCT with DD estimates of whether economics grades in the A or B ranges influence student outcomes. Other studies have used DD methods to study the effect of exposure to economics on gender attitudes (Paredes, Paserman, and Pino 2023); of alumni speaker gender on interest in economics (Patnaik et al. 2024); and of email reminders on course grades (Shakya and Levinstein, n.d.).
4. *Instrumental variables*. Many studies rely on instrumental variables to complement other research designs. For instance, RCTs adopting randomized encouragement designs may use random assignment to instrument for take-up of a treatment, delivering a local average treatment effect (LATE) estimate of the effect of the intervention among compliers. Fuzzy regression discontinuity designs use threshold rules with imperfect compliance as instrumental variables for treatment take-up. The Bleemer and Mehta (2022) study cited above uses the GPA cutoff to instrument for majoring in economics, the treatment of interest.

4.2 Descriptive approaches

Although not reliable for causal inference, descriptive studies can uncover patterns for later testing using RCTs, or document stylized facts of independent interest. Descriptive studies have played an important role in advancing the knowledge frontier in several areas of economic education covered in this paper, including:

1. *Gender gaps*: to our knowledge, Rask and Tiefenthaler (2008) were the first to document the apparently greater sensitivity of female students to economics course grades, a topic later pursued in several RCTs and other studies. Emerson, McGoldrick, and Siegfried (2018) analyzed the relationship between gender gaps and the presence of role models and quantitative requirements. Butcher, McEwan, and Weerapana (2023) find that students at women's colleges are more likely to major in economics compared to female students at coeducational colleges, providing additional descriptive evidence on the roles of peers and role models.
2. *Diversity*: Bayer et al. (2020) measure perceptions of relevance, belonging, and growth mindsets among students in introductory economics. They find significant differences across demographic groups. Krafft et al. (2024) explore differences in persistence in economics

across demographic groups, focusing on the roles of institutional policies and support systems.

3. *Content of introductory economics courses*: Bayer, Bruich, et al. (2020) and Owen and Hagstrom (2021) describe new approaches to introductory economics courses and their reception among students at Harvard University and Hamilton College, respectively.

5 Conclusion

Randomized controlled trials (RCTs) can provide clear insights into the causal effects of interventions in economic education. RCTs in economic education have demonstrated the efficacy of interventions to change student behavior, improve academic outcomes, and increase diversity within economics programs, among other topics. RCTs in these areas have also uncovered which approaches fail to work well or have unintended consequences. However, the implementation of RCTs requires careful planning, including considerations of the unit of randomization, power analysis, and the cooperation of various stakeholders. Despite the challenges, the benefits of using RCTs, such as the ability to control for selection bias and accurately measure the impact of interventions, make them a valuable tool for researchers and educators.

When RCTs are not feasible, quasi-experimental approaches and descriptive studies can still offer valuable insights. At best, alternative methods such as natural experiments, regression discontinuity design, and difference in differences can deliver credible causal inference in the absence of experimental evidence. Descriptive approaches can complement RCTs and other causal evidence by identifying patterns and generating hypotheses for future research. The growing literature on economic education underscores the importance of using a variety of research methods to address complex questions about teaching and learning. Creative use of RCTs and complementary approaches can contribute a deeper understanding of how to improve economic education.

6 References

Allgood, Sam, William B. Walstad, and John J. Siegfried. 2015. "Research on Teaching Economics to Undergraduates." *Journal of Economic Literature* 53 (2): 285–325. https://doi.org/10.1257/jel.53.2.285.

Antman, Francisca M., Nicholas E. Flores, and Evelyn Skoy. 2020. "Can Better Information Reduce College Gender Gaps? The Impact of Relative Grade Signals on Academic Outcomes for Students in Introductory Economics."

Avilova, Tatyana, and Claudia Goldin. 2018. "What Can UWE Do for Economics?" *AEA Papers and Proceedings* 108:186–90. https://doi.org/10.1257/pandp.20181103.

———. 2024. "Seeking the 'Missing Women' of Economics with the Undergraduate Women in Economics Challenge." *Journal of Economic Perspectives* 38 (3): 137–62. https://doi.org/10.1257/jep.38.3.137.

Balaban, Rita, and Patrick Conway. 2020. "A Test of Enhancing Learning in Economics through Nudges." *AEA Papers and Proceedings* 110 (May):289–93. https://doi.org/10.1257/pandp.20201050.

Banerjee, Abhijit, Esther Duflo, Amy Finkelstein, Lawrence F Katz, Benjamin A Olken, and Anja Sautmann. 2020. "In Praise of Moderation: Suggestions for the Scope and Use of Pre-Analysis Plans for RCTs in Economics." Working Paper 26993. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w26993.

Bayer, Amanda, Syon P. Bhanot, Erin T. Bronchetti, and Stephen A. O'Connell. 2020. "Diagnosing the Learning Environment for Diverse Students in Introductory Economics: An Analysis of Relevance, Belonging, and Growth Mindsets." *AEA Papers and Proceedings* 110 (May):294–98. https://doi.org/10.1257/pandp.20201051.

Bayer, Amanda, Syon P. Bhanot, and Fernando Lozano. 2019. "Does Simple Information Provision Lead to More Diverse Classrooms? Evidence from a Field Experiment on Undergraduate Economics." *AEA Papers and Proceedings* 109 (May):110–14. https://doi.org/10.1257/pandp.20191097.

Bayer, Amanda, Gregory Bruich, Raj Chetty, and Andrew Housiaux. 2020. "Expanding and Diversifying the Pool of Undergraduates Who Study Economics: Insights from a New Introductory Course at Harvard." Working Paper 26961. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w26961.

Birdi, Alvin, Steve Cook, Caroline Elliott, Ashley Lait, Tesfa Mehari, and Max Wood. 2023. "A Critical Review of Recent Economics Pedagogy Literature, 2020–2021." *International Review of Economics Education* 43 (June):100264. https://doi.org/10.1016/j.iree.2023.100264.

Bleemer, Zachary, and Aashish Mehta. 2022. "Will Studying Economics Make You Rich? A Regression Discontinuity Analysis of the Returns to College Major." *American Economic Journal: Applied Economics* 14 (2): 1–22. https://doi.org/10.1257/app.20200447.

Butcher, Kristin F., Patrick McEwan, and Akila Weerapana. 2023. "Women's Colleges and Economics Major Choice: Evidence from Wellesley College Applicants." Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w31144.

Canaan, Serena, and Pierre Mouganie. 2021. "Does Advisor Gender Affect Women's Persistence in Economics?" *AEA Papers and Proceedings* 111 (May):112–16. https://doi.org/10.1257/pandp.20211122.

Carrell, Scott E., and Michal Kurlaender. 2023. "My Professor Cares: Experimental Evidence on the Role of Faculty Engagement." *American Economic Journal: Economic Policy* 15 (4): 113–41. https://doi.org/10.1257/pol.20210699.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. "Using Randomization in Development Economics Research: A Toolkit." *Handbook of Development Economics* 4:3895–3962.

Emerson, Tisha LN, KimMarie McGoldrick, and John J. Siegfried. 2018. "The Gender Gap in Economics Degrees: An Investigation of the Role Model and Quantitative Requirements Hypotheses." *Southern Economic Journal* 84 (3): 898–911.

Felkey, Amanda J., Eva Dziadula, Eric P. Chiang, and Jose Vazquez. 2021. "Microcommitments: The Effect of Small Commitments on Student Success." *AEA Papers and Proceedings* 111 (May):92–96. https://doi.org/10.1257/pandp.20211043.

Fernandez, Jose M., Erin A. Yetter, and Kim Holder. 2021. "What Do Economic Education Scholars Study? Insights from Machine Learning." *The Journal of Economic Education* 52 (2): 156–72. https://doi.org/10.1080/00220485.2021.1887027.

Gerber, Alan S., and Donald P. Green. 2011. "Field Experiments and Natural Experiments." https://academic.oup.com/edited-volume/35474/chapter/303823967?searchresult=1&utm_source=TrendMD&utm_medium=cpc&utm_campaign=Oxford_Academic_Books_TrendMD_0.

Glennerster, Rachel, and Kudzai Takavarasha. 2014. *Running Randomized Evaluations: A Practical Guide*. Princeton University Press. https://doi.org/10.1515/9781400848447.

Halim, Daniel, Elizabeth T. Powers, and Rebecca Thornton. 2022. "Gender Differences in Economics Course-Taking and Majoring: Findings from an RCT." *AEA Papers and Proceedings* 112 (May):597–602. https://doi.org/10.1257/pandp.20221120.

J-PAL. n.d. "Research Resources." The Abdul Latif Jameel Poverty Action Lab (J-PAL). Accessed October 10, 2024. https://www.povertyactionlab.org/research-resources.

Kofoed, Michael S., Lucas Gebhart, Dallas Gilmore, and Ryan Moschitto. 2024. "Zooming to Class? Experimental Evidence on College Students' Online Learning during COVID-19." *American Economic Review: Insights* 6 (3): 324–40. https://doi.org/10.1257/aeri.20230077.

Krafft, Caroline, Kristine West, Allen Bellas, Ming Lo, Joy Moua, Natalie Chandler, J Dombroski, Najma Hussein, and Ayah Makled. 2024. *Persistence in Economics: Differences across Identities and Institutions*.

Li, Hsueh-Hsiang. 2018. "Do Mentoring, Information, and Nudge Reduce the Gender Gap in Economics Majors?" *Economics of Education Review* 64 (June):165–83. https://doi.org/10.1016/j.econedurev.2018.04.004.

McEwan, Patrick J., Sheridan Rogers, and Akila Weerapana. 2021. "Grade Sensitivity and the Economics Major at a Women's College." *AEA Papers and Proceedings* 111 (May):102–6. https://doi.org/10.1257/pandp.20211045.

Oreopoulos, Philip, Richard W. Patterson, Uros Petronijevic, and Nolan G. Pope. 2019. "Low-Touch Attempts to Improve Time Management among Traditional and Online College Students." *Journal of Human Resources*, November. https://doi.org/10.3368/jhr.57.1.0919-10426R1.

Oreopoulos, Philip, and Uros Petronijevic. 2018. "Student Coaching: How Far Can Technology Go?" *Journal of Human Resources* 53 (2): 299–329. https://doi.org/10.3368/jhr.53.2.1216-8439R.

———. 2019. "The Remarkable Unresponsiveness of College Students to Nudging And What We Can Learn from It." Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w26059.

Orr, Larry L. 1999. *Social Experiments: Evaluating Public Programs with Experimental Methods*. Sage. https://books.google.com/books?hl=en&lr=&id=fEDr7roP9dkC&oi=fnd&pg=PR11&dq=orr+social+experiments&ots=qlcQTXKINt&sig=G3S17H_A0y2QiJbbTwj6GtWF-jg.

Owen, Ann L., and Paul Hagstrom. 2021. "Broadening Perceptions of Economics in a New Introductory Economics Sequence." *The Journal of Economic Education*, 1–17.

Paredes, Valentina, M. Daniele Paserman, and Francisco J. Pino. 2023. "Does Economics Make You Sexist?" *The Review of Economics and Statistics*, July, 1–47. https://doi.org/10.1162/rest_a_01356.

Patnaik, Arpita, Gwyn Pauley, Joanna Venator, and Matthew Wiswall. 2024. "The Impacts of Same and Opposite Gender Alumni Speakers on Interest in Economics." *Economics of Education Review* 102 (October):102579. https://doi.org/10.1016/j.econedurev.2024.102579.

Porter, Catherine, and Danila Serra. 2020. "Gender Differences in the Choice of Major: The Importance of Female Role Models." *American Economic Journal: Applied Economics* 12 (3): 226–54. https://doi.org/10.1257/app.20180426.

Pugatch, Todd, and Elizabeth Schroeder. 2021. "Promoting Female Interest in Economics: Limits to Nudges." *AEA Papers and Proceedings* 111 (May):123–27. https://doi.org/10.1257/pandp.20211024.

———. 2024. "A Simple Nudge Increases Socioeconomic Diversity in Undergraduate Economics." *Economic Inquiry* 62 (1): 287–307. https://doi.org/10.1111/ecin.13172.

Pugatch, Todd, Elizabeth Schroeder, and Nicholas Wilson. 2024. "Study More Tomorrow." *Educational Evaluation and Policy Analysis*, January, 01623737231218730. https://doi.org/10.3102/01623737231218730.

Pugatch, Todd, and Nicholas Wilson. 2018. "Nudging Study Habits: A Field Experiment on Peer Tutoring in Higher Education." *Economics of Education Review* 62 (Supplement C): 151–61. https://doi.org/10.1016/j.econedurev.2017.11.003.

———. 2024. "Nudging Demand for Academic Support Services: Experimental and Structural Evidence from Higher Education." *Journal of Human Resources* 59 (5): 1637–82.

Rask, Kevin, and Jill Tiefenthaler. 2008. "The Role of Grade Sensitivity in Explaining the Gender Imbalance in Undergraduate Economics." *Economics of Education Review* 27 (6): 676–87.

Shakya, Shishir, and Michael Levinstein. n.d. "Grading Policy, Email Nudges, and Grade Outcomes." *The Journal of Economic Education* 0 (0): 1–14. https://doi.org/10.1080/00220485.2024.2372306.

Wright, Nicholas, Puneet Arora, and Jesse Wright. 2023. "I Promise to Work Hard: The Impact of Commitment Pledges on Student Performance." SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.4610399.