

مروری نظام‌مند بر آزمایش‌های میدانی در مدیریت دولتی^۱

جاسپر آسرینگ هانسن^۲، لارس تامرس^۳
مترجم: شیرین خوشنود

شواهدی برای تمرین

چکیده

آزمایش‌های میدانی برای محققان و متخصصان مدیریت دولتی ارزش بالایی دارد؛ زیرا ممکن است امکان استنتاج علی را در محیط‌های واقعی فراهم کند. انجام آزمایش‌های میدانی اغلب دشوار است. محققان و متخصصانی که آزمایش‌های میدانی انجام می‌دهند باید هزینه‌ها، امکان پذیر بودن، اخلاقی بودن و اعتبار مطالعات خود را در نظر بگیرند. آزمایش‌های میدانی در مدیریت دولتی رایج شده است (جیمز، جان، و موزلی،^۴ ۲۰۱۷). محققانی که موضوعاتی مانند تولید مشترک (جکوبسون^۵، ۲۰۱۳)، تبعیض (نیل، آدام و گروس^۶، ۲۰۱۶) و رهبری (بل^۷، ۲۰۱۴) را بررسی می‌کنند، از این روش استفاده می‌کنند. انقلاب اعتبار در علم اقتصاد (انگریست و پیشکه^۸، ۲۰۱۰) به مدیریت دولتی نفوذ کرده است؛ جایی که تمرکز بر شناسایی اثرهای علی، همراه با ترجیحات برای درمان‌های واقع‌بینانه، گسترش آزمایش‌های میدانی را توضیح می‌دهد. واقع‌گرایی تجربی مزیتی است که آزمایش‌های میدانی را از سایر روش‌های دیگر دارای اعتبار داخلی مانند آزمایش‌های پیمایشی و آزمایشگاهی متمایز می‌کند (باکگارد و همکاران^۹، ۲۰۱۵).

آزمایش‌های میدانی در مدیریت دولتی رایج شده است. با مجوز دادن برای شناسایی اثرهای علی در محیط‌های واقع‌گرایانه، آزمایش‌های میدانی ممکن است در چندین برنامه تحقیقاتی مرتبط با این حوزه به مرکزیت تبدیل شوند. انجام آزمایش‌های میدانی دشوار است و مشکلات اغلب در طول فرایند رخ می‌دهند. با این حال، محققانی که تازه به این روش دست یافته‌اند، منابع کمی در مدیریت دولتی دارند برای اینکه مشکلاتی که هنگام انجام آزمایش‌های میدانی ایجاد می‌شوند را بررسی کنند. این بررسی سیستماتیک، ۴۲ آزمایش میدانی در مدیریت دولتی را شناسایی می‌کند و به‌عنوان مقدمه‌ای برای آزمایش‌های میدانی در مدیریت دولتی عمل می‌کند. این مقاله، چگونگی توسعه آزمایش‌های میدانی در طول زمان را مورد بحث قرار می‌دهد و روندهای آزمایش میدانی در مدیریت دولتی را برجسته می‌کند؛ سپس، مسائلی را که باید در هنگام طراحی آزمایش‌های میدانی در نظر گرفته شود را مورد بحث قرار می‌دهد. از جمله این موارد می‌توان به هزینه، امکان پذیر بودن، اخلاق و اعتبار اشاره کرد. در نهایت، نویسندگان؛ یک دستور کار تحقیقاتی آینده برای آزمایش‌های میدانی مدیریت دولتی را پیشنهاد می‌کنند.

1. A Systematic Review of Field Experiments in Public Administration, *Public Administration Review*, Vol. 80, Iss. 6, pp. 921–931. DOI: 10.1111
2. Jesper Asring Hansen
3. Lars Tummers
4. James, John, and Moseley
5. Jakobsen
6. Grohs, Adam & Knill
7. Bell
8. Angrist and Pischke
9. Baekgaard et al.

میدانی بودن، «میزانی است که یک آزمایش در یک میدان واقعی انجام می‌شود.» (ادن^۴، ۲۰۱۷: ۹۷). مرزهای مبهم بین آزمایش‌های میدانی و پیمایشی وجود دارد؛ اما متوجه شدیم که آزمایش‌های میدانی مدیریت دولتی خوب عمل می‌کنند.

بخش دوم تجزیه و تحلیل هزینه‌ها؛ امکان پذیر بودن، اخلاقی بودن و اعتبار آزمایش‌های میدانی را در نظر می‌گیرد؛ زیرا اکثر آزمایشگران میدانی باید این مسائل را در نظر بگیرند. آزمایش‌های میدانی معمولاً نیازمند تلاش بیشتر محققان نسبت به سایر انواع مطالعات است؛ به همین دلیل است که بررسی نحوه مدیریت این موارد توسط محققان مدیریت دولتی مفید است. علاوه بر این، ماتون مختلفی را متمایز می‌کنیم که در مورد نحوه مدیریت این موارد به طور مناسب اطلاع می‌دهند؛ به طوری که آزمایش‌های میدانی، هزینه‌های کمتری را به بخشی از محققان و شرکت کنندگان تحمیل می‌کنند و نتیجه‌گیری تا حد امکان معتبر خواهد بود.

در این مقاله مباحث زیر مطرح می‌شود: ما توصیفی از چیستی آزمایش میدانی ارائه می‌دهیم. پس از آن، نحوه انجام بررسی سیستماتیک و جستجوی مطالعات را شرح می‌دهیم؛ سپس به تحلیل می‌رویم. بخش اول، روندها و موضوعات مورد مطالعه را بررسی می‌کند، در حالی که بخش دوم هزینه‌ها، امکان پذیر بودن، اخلاقیات و اعتبار را مورد بحث قرار می‌دهد. ما با بحث در مورد آینده آزمایش میدانی در مدیریت دولتی، مقاله را پایان می‌دهیم.

آزمایش‌های میدانی و مدیریت دولتی

آزمایش‌های میدانی مانند سایر آزمایش‌ها، واحدها را به طور تصادفی به گروه‌های رفتار و کنترل اختصاص می‌دهند. آنها از نظر میدانی با انواع دیگر آزمایش‌ها متفاوت هستند (ادن، ۲۰۱۷). میدانی بودن با حداقل چهار معیار تعیین می‌شود: (۱) مداخله واقع بینانه است، (۲) شرکت کنندگان با رفتار در دنیای واقعی مواجه می‌شوند، (۳) زمینه، طبیعی است و (۴) معیارهای نتیجه، بازتابی از نتیجه مورد علاقه هستند (ادن، ۲۰۱۷). زیبور، جیمز و لیست^۵ (۲۰۱۹)، معیارهای دیگری را پیشنهاد می‌کنند: اینکه آیا جمعیت شامل دانش آموزان است یا جمعیت مورد علاقه، خواه محیط مصنوعی یا طبیعی است، و آیا رفتار آشکار یا پنهان است؟ این معیارها چهار نوع آزمایش را ایجاد می‌کنند: آزمایش‌های آزمایشگاهی، آزمایش‌های میدانی مصنوعی، آزمایش‌های میدانی چارچوبی و آزمایش‌های میدانی طبیعی. (زیبور، جیمز و لیست، ۲۰۱۹)

معیارهای میدانی، تفاوت بین آزمایش‌های میدانی و آزمایش‌های

آزمایش‌های میدانی ممکن است به مرکز برنامه‌های تحقیقاتی جدید در مدیریت دولتی تبدیل شوند. یک ادبیات نوظهور با ترکیب بینش‌های روان‌شناسی با مدیریت دولتی، سوالات پژوهشی را در مدیریت دولتی بررسی می‌کند (به الگوی پژوهشی با عنوان مدیریت دولتی رفتاری؛ به باتاگلیو و همکاران^۱، ۲۰۱۹ و گریملیخویسن و همکاران^۲، ۲۰۱۷ مراجعه کنید). از آنجا که بسیاری از سوالات اصلی این دستور کار تحقیقاتی (و به طور کلی، مدیریت دولتی)، مربوط به رفتار شهروندان، کارمندان دولت و سیاستمداران است، آزمایش‌های میدانی ممکن است سوالات مهمی را روشن کند که در مدیریت دولتی نادیده گرفته شده‌اند. با این حال، تا به امروز، آزمایش‌های میدانی زیادی در مدیریت دولتی وجود ندارد؛ اگرچه، این روش در حال کسب محبوبیت است (جیمز، جان، و موزلی، ۲۰۱۷). ما گمان می‌کنیم که این ممکن است به این دلیل باشد که انجام آزمایش‌های میدانی دشوار است و تحقیقات بسیار کمی در مدیریت دولتی برای راهنمایی محققان در کنترل مشکلاتی وجود دارد که هنگام انجام آزمایش‌های میدانی به وجود می‌آید. این بررسی سیستماتیک به عنوان مقدمه‌ای برای آزمایش‌های میدانی در مدیریت دولتی با برجسته کردن موضوعات مورد بحث و سوالاتی که باید در هنگام انجام آزمایش‌های میدانی در نظر گرفته شوند، عمل می‌کند. ما نمونه‌هایی از مدیریت دولتی و ادبیات مرتبط را برای ایجاد یک منبع دقیق در مورد چگونگی انجام آزمایش‌های میدانی ایجاد می‌کنیم.

یک بررسی سیستماتیک از آزمایش‌های میدانی در مدیریت دولتی برای بررسی تحقیقات تجربی میدانی انجام دادیم و از دستورالعمل‌های PRISMA (اقلام گزارش ترجیحی برای بررسی‌های سیستماتیک و متاآنالیز) پیروی کردیم که یک چک‌لیست ۲۷ موردی برای گزارش مرورهای سیستماتیک ارائه می‌دهد. پیروی از PRI SMA منجر به مطالعات شفاف‌تر و قابل تکرارتر می‌شود (موهر و همکاران^۳، ۲۰۰۹). بخش اول تجزیه و تحلیل به روندها و موضوعات در آزمایش‌های میدانی می‌پردازد. این نشان می‌دهد که آزمایش‌های میدانی در مدیریت دولتی محبوب‌تر می‌شوند و عملکرد، رهبری و تولید مشترک، بیشترین موضوعات مورد مطالعه هستند.

پس از تجزیه و تحلیل اکتشافی، در مورد اینکه آیا مطالعات دربرگیرنده شاخص‌های «میدانی بودن» هستند و چگونگی تشخیص آزمایش میدانی از یک آزمایش پیمایشی بحث می‌کنیم.

1. Battaglio et al.
2. Grimmlikhuijsen et al.
3. Moher et al.
4. Eden
5. Czibor, Jimenez-Gomez, and List

در بررسی، سال انتشار و موضوعاتی که آنها را پوشش می‌دهند، توضیح می‌دهیم. همه داده‌ها را می‌توان در آدرس <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/LF9H7Q>

یافت.

معیارهای تطبیق‌پذیری

مطالعات در صورتی که تمام معیارهای زیر را داشته باشند، در مرور سیستماتیک گنجانده شدند:

- طراحی مطالعه: مطالعات باید به عنوان آزمایش‌های میدانی یا مطالعات ممیزی شناسایی شوند. آنها باید حاوی کلمات «آزمایش میدانی»، «کارآزمایی کنترل شده تصادفی» یا «مطالعه حسابرسی» در عنوان یا چکیده باشند.

- موضوع و نوع شرکت‌کنندگان: موضوع باید با مدیریت دولتی مرتبط باشد. موضوعات مرتبط شامل: تعاملات شهروند و دولت، مداخلات در سازمان‌های عمومی و مواردی از این نوع است. شرکت‌کنندگان در این مطالعه باید علاقه‌مند به مدیریت دولتی باشند. برای مثال، شرکت‌کنندگان سیاستمداران، کارمندان دولتی یا شهروندان هستند.

- زبان: ما تنها مطالعات گزارش شده به زبان انگلیسی را در نظر گرفتیم. - وضعیت انتشار: ما فقط مقالات مجلات با داور هم‌تای منتشر شده در مجلات مدیریت دولتی را در نظر گرفتیم.

- سال انتشار: تمام سال‌های انتشار را برای یافتن مطالعات، تا سال ۲۰۱۸ (بدون محدودیت) در نظر گرفتیم. ما تنها مطالعات منتشر شده از سال ۲۰۱۰ تا ۲۰۱۸ و برخی از مطالعاتی که در سال ۲۰۱۸ ارائه شده بودند و در سال ۲۰۱۹ منتشر شدند را یافتیم.

راهبرد جستجو

ما از سه استراتژی جستجو استفاده کردیم. ابتدا، ما پایگاه‌های اطلاعاتی الکترونیکی را برای عبارات «آزمایش میدانی»، «مطالعه حسابرسی» و «عمومی» [زشته جستجو: «آزمایش میدانی» یا «مطالعه حسابرسی» و «عمومی»] جستجو کردیم. در پایگاه اسکپوس جستجو و ۱۰۴۰ نتیجه به دست آمد. دوم، ما مجلات پیشرو و تخصصی مدیریت دولتی را با استفاده از عبارت جستجوی «آزمایش میدانی» پیدا کردیم. نشریاتی که در آنها جستجو انجام شد، این موارد بود:

Journal of Public Administration Research and Theory- (مطالعه بالقوه) ۴۸)

پیمایشی را برجسته می‌کند. از آنجایی که آزمایش‌های پیمایشی در نظرسنجی‌ها گنجانده شده‌اند، معمولاً نمی‌توانند رفتارهایی مشابه آنچه در دنیای واقعی با آن مواجه می‌شوند داشته باشند. زمینه‌های آنها، نظرسنجی‌هایی است که شبیه موقعیت‌های زندگی واقعی نیستند و نتایج اغلب پرسش‌ها هستند (اگرچه، برخی از نظرسنجی‌ها شامل شاخص‌های نیمه‌رفتاری هستند). تفاوت‌های زیاد بین نظرسنجی‌ها و تنظیمات واقع‌گرایانه باعث می‌شود که مشخص نباشد نتیجه‌گیری‌های آنها چگونه بر رفتار در دنیای واقعی منطبق می‌شوند (لویت و لیست^۱، ۲۰۰۷). همین امر در مورد آزمایش‌های آزمایشگاهی نیز صدق می‌کند. آزمایش‌های آزمایشگاهی، شرکت‌کنندگانی را جذب می‌کنند که به انواع مختلف درمان در محیط آزمایشگاه اختصاص داده می‌شوند. اغلب، تفاوت‌های عمیقی بین محیط آزمایشگاه و دنیای واقعی وجود دارد؛ اگرچه آزمایش‌های آزمایشگاهی مزایایی را از نظر کنترل تجربی ارائه می‌کنند (فالك و هلکمن^۲، ۲۰۰۹)؛ از این رو، آزمایش‌های میدانی به محققان اجازه می‌دهند تا دانشی در مورد چگونگی تجلی اثر علی‌علاقه در دنیای واقعی به دست آورند. یک آزمایش میدانی در مدیریت دولتی در یک محیط طبیعی (به عنوان مثال، یک شهرداری) شامل موضوعات مرتبط (به عنوان مثال، کارمندان دولت)، استفاده از درمان‌های معتبر (به عنوان مثال، جمعیت‌شناسی شهروندان) و سنجش نتایج واقعی (مانند کیفیت پاسخ‌ها) است. نمونه‌ای که تمام معیارها را برآورده می‌کند، مطالعه جیلک، ون دورن و ریس^۳ (۲۰۱۸) است. این آزمایش میدانی نشان داد که چگونه خانه‌های مراقبت از سالمندان دولتی و خصوصی در بلژیک به ایمیل‌های بومیان نسبت به مهاجران مغربی پاسخ می‌دهند. در نتیجه، محققان توانستند پاسخ‌ها را مقایسه کنند و نشان دهند که آیا تبعیض براساس قومیت با نرخ‌های متفاوتی در سازمان‌های دولتی و خصوصی رخ می‌دهد یا خیر؟ تصور اینکه آزمایش‌های نظرسنجی یا آزمایشگاهی بتواند واقع‌گرایانه مورد نیاز برای نتیجه‌گیری در مورد تبعیض توسط ارائه‌دهندگان خدمات دولتی و خصوصی را ارائه دهد، دشوار است.

مروری سیستماتیک از آزمایش‌های میدانی

ما یک بررسی سیستماتیک برای ارزیابی آزمایش‌های میدانی در مدیریت دولتی انجام داده‌ایم. چک‌لیست PRISMA و نحوه پابندی ما به موارد موجود در اطلاعات پشتیبانی را به صورت آنلاین گزارش می‌کنیم. در آنجا، ما در مورد مقالات گنجانده شده

1. Levitt and List

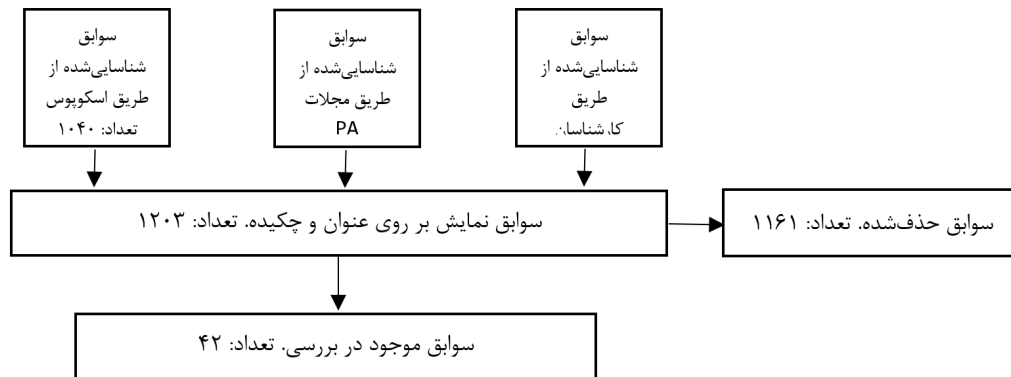
2. Falk and Heckman

3. Jilke, van Dooren, and Rys

را شناسایی کردند که ما دو مورد از آنها را در بررسی قرار دادیم.
 ۱. Textbox. ۱. رشته جستجوی مورد استفاده برای Scopus.
 String used for Scopus: TITLE-ABS-KEY (“field exper-
 iment” OR “audit study” AND “public”) AND (LIMIT-
 TO(PUBSTAGE, “final”) OR LIMIT-TO(PUBSTAGE,
 “aip”)) AND (LIMIT-TO (ACCESSTYPE[OA]) OR
 LIMIT-TO (ACCESSTYPE[OTHER])) AND (LIMIT-
 .[“TO [LANGUAGE, “English

انتخاب رکورد

پس از غربالگری، عناوین و چکیده‌های همه سوابق شناسایی شده
 از طریق بررسی سیستماتیک، ۴۲ مطالعه واجد شرایط برای ورود
 پیدا کردیم. شکل (۱)، فرایند انتخاب را به تصویر می کشد.



شکل ۱: فرایند انتخاب

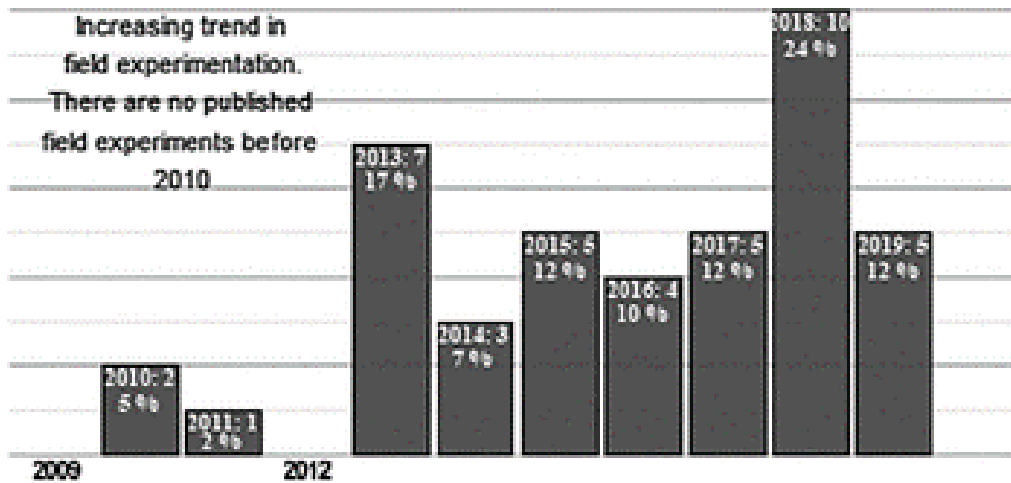
نتایج

ویژگی‌های آزمایش‌های میدانی در مدیریت دولتی
 بخش اول این تحلیل، ویژگی‌های آزمایش‌های میدانی
 شناسایی شده را برجسته می‌کند. اولین نتیجه، روند افزایشی است
 که در شکل (۲) نشان داده شده است. قبل از سال ۲۰۱۰، هیچ
 آزمایش میدانی در مجلات مدیریت دولتی وجود نداشت. از سال
 ۲۰۱۰ به بعد، آزمایش‌های میدانی محبوبیت بیشتری پیدا کردند.
 بیشترین آزمایش‌های میدانی در سال ۲۰۱۸ منتشر شده است.
 مجلات مدیریت دولتی، ۱۰ آزمایش میدانی را در سال ۲۰۱۸
 منتشر کردند (به عنوان مثال، اندرسن و موبنیهان^۱، ۲۰۱۸؛ جیلک،
 ون دورن و ریس^۲، ۲۰۱۸؛ پورتر و روگوسکی^۳، ۲۰۱۸). این
 روند، ارتباط در نظر گرفتن مسائلی را که هنگام انجام آزمایش‌های
 میدانی به وجود می‌آیند، برجسته می‌کند.

کدنویسی

ما مطالعات را براساس موضوعات تحت پوشش آنها کدگذاری
 کردیم تا سؤالات برجسته شده با استفاده از آزمایش‌های میدانی را
 بررسی کنیم. فرایند کدگذاری تا حدی اکتشافی بود؛ زیرا هر بار که
 مقاله‌ای در بررسی با هیچ‌یک از مضامین مطابقت نداشت، دسته
 جدیدی را اضافه می‌کردیم. ما در نهایت به هفت موضوع رسیدیم
 (انگیزه، تولید مشترک، عملکرد، رهبری، شفافیت، تبعیض، و موارد
 دیگر) که متقابل نیستند. این بدان معناست که یک مقاله می‌تواند
 دو موضوع را همزمان مطالعه کند. به عنوان مثال، یک مطالعه
 می‌تواند به ادبیات رهبری و انگیزه به‌طور همزمان مرتبط باشد (به
 عنوان مثال، بل، ۲۰۱۴).

1. Andersen and Moynihan
2. Jilke, van Dooren, and Rys
3. Porter and Rogowski



شکل ۲: آزمایش میدانی در طول زمان

جدول ۱. موضوعات در آزمایش‌های میدانی

مطالعه نمونه	تعداد/درصد	موضوع مدیریت دولتی
Andersen and Moynihan (2016)	۸/۱۹٪	عملکرد
Bellé (2014)	۷/۱۷٪	رهبری
Thomsen and Jakobsen (2015)	۶/۱۴٪	تولید مشترک
Grimmelikhuijsen and Klijn (2015)	۶/۱۴٪	شفافیت
Linos (2018)	۴/۱۰٪	انگیزه
Jilke, van Dooren, and Rys (2018)	۴/۱۰٪	تبعیض
Hess, Hanmer, and Nickerson (2016)	۹/۲۱٪	سایر موارد
۴۴/۱۰۵٪		جمع

جدول ۱، موضوعات مورد مطالعه را نشان می‌دهد. موضوعی که اغلب مورد مطالعه قرار می‌گیرد، عملکرد است (به‌عنوان مثال، اندرسن و لارسن^۱؛ اندرسن و موینیهان^{۲۰۱۸})، با ۱۹ درصد از آزمایش‌های میدانی منتشر شده و به دنبال آن رهبری که به دلیل پروژه LEAP افزایش یافت و در آن، محققان دانشگاه آرهوس؛ آموزش رهبری را به مدیران دولتی ارائه کردند تا تأثیرات رهبری تحول‌آفرین و تعاملی را بر نتایج سازمانی بررسی کنند (به‌عنوان مثال، آن و همکاران^{۲۰۱۹}؛ جنسن^{۲۰۱۸}). ۱۷ درصد دیگر از آزمایش‌های میدانی شناسایی شده در رهبری مطالعه مورو و ۱۴ درصد آزمایش‌های میدانی بر تولید مشترک متمرکز شده‌اند

(به‌عنوان مثال، یاکوبسن^{۲۰۱۳}؛ یاکوبسن^۴ و اندرسن^۵ ۲۰۱۳). موضوعاتی مانند تبعیض (به‌عنوان مثال، جیلک، ون دورن و ریس^۶، ۲۰۱۸)، شفافیت (به‌عنوان مثال، گرملیخویسن و کلینج^۷، ۲۰۱۵)، و انگیزه (به‌عنوان مثال، بل، ۲۰۱۹) نیز محبوب هستند، اما تا حدودی کمتر. مقوله سایر (حاوی ۲۱ درصد از آزمایش‌های میدانی مدیریت دولتی) شامل مطالعاتی در مورد موضوعاتی مانند تشویق (جان و بلوم^۸، ۲۰۱۸)، بودجه دولتی و کمک‌های خیریه (جیلک و همکاران^۹، ۲۰۱۹) و ثبت نام رأی‌دهندگان (هس، هانمر، و نیکرسون^{۱۰}، ۲۰۱۶).

1. Andersen and Larsen
2. An et al.
3. Jensen
4. Jakobsen

5. Jakobsen and Andersen
6. Jilke, van Dooren, and Rys
7. Grimmelikhuijsen and Klijn
8. John and Blume

9. Jilke et al.
10. Hess, Hanmer, and Nickerson

میدانی بودن

به همین دلیل باید انتظار داشت که همه معیارهای میدانی برآورده نشوند و این مطالعه، عناصری از آزمایش های پیمایشی و میدانی را ترکیب می کند (بله، کانتارلی و بلاردینلی^۱، ۲۰۱۷). از این رو، تعیین واضح اینکه آزمایش میدانی چیست و چگونه از آزمایش پیمایشی متمایز می شود، دشوار است؛ اما آزمایش های میدانی در مدیریت دولتی به طور کلی با آزمایش های پیمایشی، حداقل در برخی موارد متفاوت است. محققانی که آزمایش های میدانی را انجام می دهند باید در نظر بگیرند که چگونه مطالعات میدانی خود را افزایش دهند تا مطالعات را تا حد امکان واقعی کنند و در نتیجه، از قدرت اولیه آزمایش های میدانی بهره مند شوند. حوزه های که بیش از همه در ادبیات مدیریت دولتی در این زمینه نیاز به بهبود دارد، میدانی بودن معیار نتیجه است. شاخص های رفتاری، بیشتر نشان دهنده رفتار دنیای واقعی هستند تا سوالات نظرسنجی؛ به همین دلیل است که آزمایش های میدانی مدیریت دولتی باید در آینده به دفعات بیشتری از چنین شاخص هایی استفاده کنند. نمونه هایی از معیارهای نتیجه رفتار در ادبیات مدیریت دولتی عبارت است از: عملکرد وظیفه (بله، ۲۰۱۴)، کیفیت پاسخ ها به ایمیل ها (جیلکه، ون دورن و ریس، ۲۰۱۸) و برنامه های شغلی (لینوس، ۲۰۱۸).

نتیجه گیری هایی که ما در مورد میدانی بودن به دست می آوریم ممکن است تا حدی به طبقه بندی مورد استفاده برای ارزیابی میدانی بودن بستگی داشته باشد. طبقه بندی توسعه یافته توسط زیبور، جیمز جمز و لیست^۲ (۲۰۱۹) ممکن است به نتایج کمی متفاوت در مورد میدان داری نسبت به نتایج به دست آمده با معیارهای مورد استفاده در این مقاله منجر شود. میدانی بودن آزمایش های میدانی مدیریت دولتی، موضوعی است که ممکن است در آینده مورد بحث بیشتر قرار گیرد.

ملاحظات در آزمایش های میدانی

اکنون به ملاحظاتی می پردازیم که محققان هنگام انجام آزمایش های میدانی با آنها مواجه می شوند. ما این کار را انجام می دهیم تا به محققان کمک کنیم تا چگونه یک آزمایش میدانی را تنظیم کنند تا در آینده شاهد تعداد بیشتری از آنها باشیم. ما در مورد هزینه ها، امکان پذیر بودن، اخلاقیات و اعتبار بحث می کنیم. دو موضوع اول برای راهنمایی محققانی که تازه وارد آزمایش های میدانی شده اند، برجسته می شوند؛ زیرا چنین مطالعاتی پرهزینه و اجرای آن دشوار است. با نشان دادن چگونگی کاهش هزینه ها (برو کم، کالا، و سخون^۳، ۲۰۱۷) و نحوه در نظر گرفتن امکان پذیر

بخش قبل، افزایش تعداد آزمایش های میدانی در تحقیقات مدیریت دولتی را نشان داد. اکنون می توانیم میزان امتیاز مطالعات شناسایی شده را در درجه میدانی بودن آنها بررسی کنیم. این مهم است؛ زیرا نشان می دهد که مطالعات مدیریت دولتی در کدام جنبه های میدانی امتیاز بالایی دارند و پیشرفت بالقوه در کجا قرار دارد. واضح ترین مورد آزمایش میدانی زمانی است که محققان، مطالعه ای را در محیط های طبیعی با واحدهای تحلیل مربوطه انجام می دهند که در معرض مداخله ای واقع بینانه با اندازه گیری متغیرهای مربوطه به عنوان نتیجه است. ناگفته نماند مطالعه ای که به عنوان یک آزمایش میدانی شناخته می شود، به تمام معیارها پایبند است. با این حال، بیشتر آزمایش ها را می توان روی یک پیوستار قرار داد و مرز بین آزمایش میدانی و آزمایش پیمایشی همیشه مشخص نیست.

مطالعات مدیریت دولتی که به عنوان آزمایش های میدانی شناسایی می شوند، هنگام اعمال چهار معیار بررسی شده در بخش تعریف آزمایش های میدانی (ادن، ۲۰۱۷)، امتیاز بالایی از نظر میدانی دارند. برخی آزمایش های میدانی در مجلات مدیریت دولتی وجود دارد که به وضوح همه معیارها را برآورده می کنند (بله، ۲۰۱۴، جاکوبسن، ۲۰۱۳، جیلکه، ون دورن و ریس، ۲۰۱۸). این مطالعات، آزمودنی های مرتبط را به گروه های آزمایشی با درمان های واقع بینانه اختصاص و در محیط های واقعی رخ می دهند و معیارهایی از رفتار را به عنوان پیامدهای خود دارند. سایر مطالعات، همه معیارها را برآورده می کنند؛ اما سوالات نظرسنجی را به عنوان نتایج دارند (مانند اندرسن و لارسن، ۲۰۱۶؛ تامسن و جاکوبسن، ۲۰۱۵). این امر، شباهت به آزمایش های پیمایشی را افزایش می دهد؛ اگرچه آنها هنوز چندین ویژگی دارند که در آنها به وضوح با آزمایش های پیمایشی تفاوت دارند. از جمله این واقعیت است که آنها در زمینه های واقعی رخ می دهند و رفتارهای واقع بینانه دارند. در مطالعات دیگری که نتایج را با پرسشنامه های نظرسنجی اندازه گیری می کنند، رفتارها واقعی تر از آزمایش های پیمایشی هستند؛ اگرچه، رفتارها شباهت هایی به روش انجام درمان ها در آزمایش های پیمایشی دارند (جیمز، ۲۰۱۱؛ جیمز و موزلی، ۲۰۱۴). با این حال، آنها هنوز با آزمایش های پیمایشی با انجام در محیط های طبیعی متفاوت هستند که میدانی بودن آنها را نسبت به آزمایش های پیمایش معمولی افزایش می دهد. در نهایت، برخی مطالعات به عنوان آزمایش های میدانی کیفی شناخته می شوند؛

1. Bellé, Cantarelli, and Belardinelli
2. Czibor, Jimenez-Gomez, and Lišt
3. Broockman, Kalla, and Sekhon

اطمینان حاصل می‌کردند، در مورد محتوای آموزش رهبری به توافق می‌رسیدند، راه‌حل را به مدیران دولتی ارائه می‌کردند (که شامل چهار جلسه آموزش رهبری بود) و در نهایت، سازمان‌ها را دنبال می‌کردند تا بررسی کنند که آیا آموزش، نتایج را تغییر داده است یا خیر؟ طبق دانش ما، این مطالعه با جمع‌آوری یک موج پایه قبل از ارائه راه‌حل و با ترکیب چندین معیار در یک شاخص، هزینه‌ها را کاهش داد (جنسن ۲۰۱۸). نمونه دیگر، مداخله پرهزینه مطالعه اندرسن و موینیهان (۲۰۱۸) است. آنها بررسی می‌کنند که آیا تازه‌واردان متمایز اجتماعی بر بازتاب تیم تأثیر می‌گذارند یا خیر و اینکه تیم‌ها در مورد تازه‌واردان متمایز اجتماعی چه فکری می‌کنند. محققان، استخدام‌های جدیدی را به برخی از مدارس که در آزمایش شرکت کردند، اختصاص دادند که دوباره تفاوت بین هزینه‌های درمان در آزمایش‌های میدانی و آزمایش‌های پیمایشی را نشان می‌دهد. معیارهای نتیجه نیز در این مورد در شاخص‌ها ترکیب می‌شوند و در نتیجه، هزینه‌ها را کاهش می‌دهند.

همه آزمایش‌های میدانی به اندازه مطالعات فوق‌گرا نیستند. نمونه‌ای از یک نوع آزمایش میدانی که هزینه بیشتری نسبت به سایر انواع مطالعات ندارد، مطالعه ممیزی است. در مطالعات ممیزی، محققان درخواست‌های جعلی ارسال می‌کنند تا بررسی کنند که آیا ویژگی‌های پس‌زمینه متفاوت بر نرخ پاسخ‌دهی تأثیر می‌گذارد یا خیر؟ تفاوت در نرخ پاسخ مشروط به ویژگی‌های پس‌زمینه، نشان‌دهنده تبعیض است. این بدان معنی است که مطالعات باید برای افراد تجربی، ایمیل‌هایی ارسال کنند که ویژگی‌های پس‌زمینه متفاوتی دارند. ریس، ون دورن، جیلکه (۲۰۱۸)، در مطالعه خود در مورد تبعیض براساس قومیت در میان ارائه‌دهندگان مراقبت از سالمندان فلاندی این کار را انجام می‌دهند. آنها دریافتند که نرخ پاسخ برای ارائه‌دهندگان دولتی و خصوصی متفاوت نیست، اما ارائه‌دهندگان عمومی تمایل دارند پاسخ‌هایی با کیفیت بالاتر ارائه دهند. بنابراین، محققان دانشی از تفاوت‌های عمومی - خصوصی را در یک محیط واقعی ارائه می‌کنند.

مطالعات حساسی ممکن است برای انجمن‌ها و سازمان‌های تحقیقاتی پرهزینه باشد، ولی برای محققانی که آنها را انجام می‌دهند ارزان باشد. در واقع، پاسخ‌دهندگان ناخواسته برای شرکت در مطالعات حساسی استخدام می‌شوند (پیجر ۲۰۰۷). این

بودن (لیست، ۲۰۱۱)، امیدواریم محققان در آینده شانس بیشتری برای موفقیت در انجام آزمایش‌های میدانی مدیریت دولتی داشته باشند. موضوعات اخلاق و اعتبار با هدف حفظ ارتباط آزمایش‌های میدانی مدیریت دولتی مورد بحث قرار می‌گیرد. آزمایش‌های میدانی در مدیریت دولتی باید اخلاق و اعتبار را در آینده در نظر بگیرند تا اعتبار خود را حفظ کنند که این یکی از مزایای اولیه آنها به ادبیات مدیریت دولتی است (انگیست و پیشکه^۱، ۲۰۱۰).

هزینه‌ها

یک تفاوت مهم بین آزمایش‌های میدانی و سایر انواع مطالعات این است که آزمایش‌های میدانی می‌توانند هزینه‌های قابل توجه بیشتری را برای محققان به همراه داشته باشند. این امر به‌ویژه در هنگام مقایسه آزمایش‌های میدانی با آزمایش‌های پیمایشی که معمولاً انجام آن آسان است و هزینه‌های پایینی را به دنبال دارد، صادق است (موتز^۲، ۲۰۱۱). هزینه‌های آزمایش‌های میدانی، حاکی از آن است که محققان باید از تحلیل‌های هزینه-فایده برای ارزیابی میزان قدرتی که می‌توانند با سطح معینی از منابع به دست آورند، استفاده کنند. چندین روش به کاهش هزینه‌ها کمک می‌کند (به عنوان مثال، بروکمن، کالا، و سخون^۳ ۲۰۱۷) که در این بخش به برخی از آنها می‌پردازیم.

یکی از روش‌های کاهش هزینه‌ها، همکاری با سازمان‌هاست. سازمان‌ها ممکن است به ارائه راه‌حل‌ها کمک کنند؛ بنابراین برخی از هزینه‌های مربوط به آزمایش میدانی را متقبل شوند. با این حال، همکاری‌های پژوهشی - عملی، نوشدارویی نیستند؛ زیرا ممکن است بسته به زیرساخت‌ها و منافع سازمان، مشکلاتی ایجاد کنند (گلنرستر^۴ ۲۰۱۷؛ لویت و لیست^۴ ۲۰۰۹). ما این مشکلات را در بخش عملی مورد بحث قرار می‌دهیم.

در هزینه‌های آزمایش‌های میدانی، تفاوت‌هایی وجود دارد، حتی اگر آنها گران باشند. این با در نظر گرفتن برخی از آزمایش‌های میدانی مدیریت دولتی نشان داده شده است. در یک سوی این پیوستار، ما مداخلات سازمانی در مقیاس بزرگ داریم و بررسی می‌کنیم که چگونه پیامدها در هنگام تغییرات عمیق در شیوه‌ها یا رویه‌های سازمانی تغییر می‌کنند. پروژه LEAP (به عنوان مثال، جنسن^۵ ۲۰۱۸؛ آن و همکاران^۶ ۲۰۱۹)، در این انتهای پیوسته قرار می‌گیرد؛ زیرا محققان باید از همکاری با شهرداری آرهوس

1. Angrist and Pischke
2. Mutz
3. Glennerster
4. Levitt and List
5. Jensen
6. An et al.
7. Pager

محقق داشته باشند (لویت و لیست ۲۰۰۹). از جمله این مسائل می توان به اهمیت تصادفی سازی (ادن، ۲۰۱۷)، مطالعه سوالات با بازده اجتماعی بالاتر به جای بازده خصوصی به سازمان (لویت و لیست ۲۰۰۹) و اطمینان از اینکه سازمان از استراتژی هایی برای جلوگیری از عدم انطباق موضوعات تجربی حمایت می کند اشاره کرد (گلنرستر ۲۰۱۷). ادن (۲۰۱۷)، چندین توصیه در مورد نحوه مدیریت اختلافات بین محققان و سازمان ها ارائه می دهد.

همکاری ها ممکن است در اجرای ارزیابی های مقیاس بزرگ برنامه ها به محققان کمک کند، اما به قیمت انعطاف پذیری است. با این حال، گلنرستر بیان می کند که مزایای خود پیاده سازی به ندرت بیشتر از هزینه های گلنرستر (۲۰۱۷) است.

موضوع دیگر، مربوط به امکان پذیر بودن کنترل است. آزمایش های میدانی کنترل کمتری را ارائه می دهند، به ویژه زمانی که با آزمایش های آزمایشگاهی مقایسه می شوند که در آن متغیرهایی مانند برهمکنش های مکرر را می توان دستکاری کرد و می توان آزمایش تئوری کنترل شده را انجام داد (فالک و هکمن ۲۰۰۹؛ لوناتی و همکاران ۲۰۱۸). محیط میدانی واقع گرایی بیشتری را ارائه می کند، اما با این واقع گرایی، پیچیدگی بیشتری به وجود می آید، زیرا روابط بین کارفرمایان و کارمندان ممکن است اثرهای درمان را تعدیل کند (فالک و هکمن ۲۰۰۹). فالک و هکمن (۲۰۰۹)، این موضوع را به عنوان موضوعی برای آزمایش های میدانی نمی دانند، بلکه به عنوان ویژگی ای که اهمیت در نظر گرفتن تأثیر علی مورد نظر را قبل از تصمیم گیری در مورد روش تحقیق برجسته می کند، در نظر گرفته می شود.

اخلاق

ملاحظات اخلاقی در آزمایشات میدانی مهم هستند. آزمایش های میدانی، مداخلات واقعی هستند، به این معنی که ممکن است بر زندگی شرکت کنندگان تأثیر بگذارند. این نشان می دهد که ملاحظات اخلاقی در آزمایش های میدانی مدیریت دولتی مهم هستند.

این ملاحظات ابتدا به این مفهوم مربوط می شود که راه حل ها نباید وضعیت هیچ فردی را بدتر کند. ادبیات مدیریت دولتی در مورد آزمایش های میدانی به این معیار پایبند است. هیچ یک از آزمایش های میدانی، متغیرهایی را دستکاری نمی کنند که انتظار می رود وضعیت کسی را بدتر کند. در عوض، راه حل ها پیامدهای مثبت بالقوه مانند شفافیت (گریملیخویسن و کلیجن، ۲۰۱۵)، مشارکت (هاک، آندرسون و پوتوسکی ۲۰۱۳) و تجهیزات تحریک کننده تولید مشترک را تشویق می کنند (تامسن و جکوبسن، ۲۰۱۵).

ممکن است اعتماد به جامعه علمی را کاهش دهد. علاوه بر این، مطالعات حسابرسی ممکن است برای سازمان هایی که ناخواسته در آن شرکت می کنند، به دلیل زمان لازم برای پاسخ دادن به ایمیل ها پرهزینه باشد. در مطالعات بسیار قوی، این ممکن است به معنای مقدار قابل توجهی، ازدست دادن ساعات کاری باشد. ما در مورد اخلاق مطالعات حسابرسی بیشتر در بخش اخلاق بحث می کنیم.

امکان پذیر بودن

دشواری های عملی انجام آزمایش های میدانی ممکن است دلیل مهمی باشد که چرا تا به امروز آزمایش های میدانی کمی در مدیریت دولتی انجام شده است. به طور کلی، اجرای آزمایش های میدانی دشوار است (به عنوان مثال، لیست ۲۰۱۱ را ببینید). ما با خواندن مطالعات منتشر شده قادر به درک این مشکلات عملی نیستیم؛ زیرا به ندرت مسائلی را که با آن مواجه شده اند توصیف می کنند. علاوه بر این، مطالعات منتشر شده؛ آزمایش های میدانی هستند که بر موانع آزمایش میدانی غلبه کردند. این نشان می دهد که ما نمی توانیم به ادبیات مدیریت دولتی در مورد آزمایش های میدانی روی آوریم تا میزان مشکلات عملی مربوط به انجام آزمایش های میدانی را بسنجیم.

برخی از آزمایشگران مطالعات میدانی، توصیه هایی در مورد چگونگی غلبه بر موانع عملی برای آزمایش میدانی ارائه کرده اند؛ از جمله گلنرستر که آزمون آزمایشی و همکاری با مجریان را توصیه می کند (گلنرستر ۲۰۱۷؛ گلنرستر و تاکاواراشا ۲۰۱۳). آزمون آزمایشی به یافتن مشکلات در اجرای راه حل، همچنین شناسایی دلایل عدم انطباق و فرسایش کمک می کند (گلنرستر ۲۰۱۷).

همکاری ها باعث می شود که مطالعات در مقیاس بزرگ در مورد مداخلات برای محقق امکان پذیرتر و کم هزینه تر باشد. این موضوع برای محققان مدیریت دولتی که امکان همکاری با سازمان های دولتی و غیرانتفاعی را در هنگام انجام آزمایش های میدانی دارند، مرتبط است. گلنرستر پیشنهاد می کند که محققان می توانند از معیارهای زیر برای تعیین امکان پذیری مشارکت استفاده کنند: سازمان باید مقیاس کافی، انعطاف پذیری، تخصص فنی، تخصص و شهرت محلی، جایجایی کم کارکنان و تمایل به دانستن حقیقت داشته باشد (برای بحث کامل در مورد معیارها، به گلنرستر ۲۰۱۷، ۱۸۰-۱۷۸ مراجعه کنید). یافتن شراکت های خوب مهم است، زیرا برخی از محققان هنگام همکاری با سازمان ها با مشکلاتی مواجه شده اند. سازمان ها ممکن است منافع متفاوتی از علایق

1. Glennerster

2. Glennerster and Takavarasha

3. Falk and Heckman

4. Lonati et al.

5. Hock, Anderson, and Potoski

درونی است. در صورت نقض این مفروضات، اعتبار درونی آزمایش میدانی به خطر می‌افتد. با این حال، از آنجایی که آنها از طریق تغییرات نتایج شرکت کنندگان تعریف می‌شوند، نمی‌توان مستقیماً مشاهده کرد که آیا آنها در آزمایش نقض شده‌اند یا خیر. اکنون به بحث در مورد چگونگی تطبیق مفروضات مربوط به اعتبار داخلی می‌پردازیم.

تهدید ۱: استثناپذیری

اولین تهدیدی که اعتبار داخلی را تهدید می‌کند، نقض فرض استثناپذیری است. این فرض بیان می‌کند که تصادفی‌سازی، پیامدها را از طریق متغیرهای دیگری غیر از دریافت راه‌حل تحت تأثیر قرار نمی‌دهد (انگریست و پیشکه، ۲۰۰۹). اگر این فرض برآورده نشود، اثر علی‌شناسایی شده در یک مطالعه ترکیبی از راه‌حل و سایر متغیرها است. فرض حذف‌پذیری را نمی‌توان به‌طور کلی نشان داد و محققان باید این فرض را «بر اساس مورد به مورد» ارزیابی کنند (آتی و ایمبسن، ۲۰۱۷). علاوه بر این، فرض استثناپذیری بستگی به چگونگی تعریف راه‌حل توسط محققان دارد. راه‌حل را می‌توان به‌عنوان اختصاص به رفتار یا دریافت راه‌حل تعریف کرد و ممکن است با تعریف دوم راه‌حل، اما نه با تعریف اول، فرض استثنایی نقض شود. این بدان معناست که محققان باید از قبل ثبت‌نام کنند که آیا به انجام رفتار یا دریافت راه‌حل علاقه‌مند هستند تا از انعطاف‌پذیری محقق که منجر به مثبت کاذب می‌شود، اجتناب کنند (سیمونز، نلسون و سیمونسون، ۲۰۱۱).

استثناپذیری، اهمیت مشاهده فرایندهای تولید داده برای نشان دادن بعید بودن تخلف را برجسته می‌کند. برخی از محققان خاطرنشان می‌کنند که مشاهدات فرایند علی به ادعاهایی در مورد علیت کمک می‌کند (برادی و کولیر، ۲۰۰۴). یک نمونه کلاسیک، جان اسنو است که شیوع وبا را در لندن مطالعه کرد. با ترکیب دانش دقیق در مورد منابع آب در خانواده‌ها با داده‌های یک آزمایش طبیعی، او توانست نشان دهد که وبا یک بیماری قابل انتقال از آب است و این نظریه رایج که از طریق هوای بد (میاسما) منتشر شده است را رد کرد (دانینگ، ۲۰۱۲). به‌طور مشابه، دانش دقیق در مورد شرایط آزمایشی میدانی و واکنش شرکت‌کنندگان به آنها ممکن است به ادعاهایی در مورد اعتبار مطالعه کمک کند. راه دوم برای افزایش معقول بودن تردیدپذیری این است که گروه‌های رفتار و کنترل را تا حد امکان مشابه انجام دهیم (گربر و گرین، ۲۰۱۲). محققان ممکن است با راه‌حل‌های دوگانه در انجام این کار موفق شوند که احتمال برخورد برابر با گروه‌های رفتار و کنترل را افزایش می‌دهد. ما نتوانستیم هیچ آزمایشی

دومین ملاحظات اخلاقی این است که تا حد امکان، زمان کمتری را از طرف شرکت کنندگان تلف کنید. جکوبسن (۲۰۱۳) در مطالعه خود در مورد تولید مشترک شهروندان، اخلاق را مورد بحث قرار می‌دهد که در آن بیان می‌کند ابداع یک رفتار کنترل دارونما اخلاقی نیست؛ حتی اگر این امکان را برای رد این فرضیه فراهم کند که درمان به دلیل اثر دارونما جواب داده است (جکوبسن، ۲۰۱۳). این ملاحظات در مورد مطالعات حسابرسی نیز اعمال می‌شود، به‌ویژه به این دلیل که شرکت کنندگان نمی‌توانند رضایت آگاهانه خود را اعلام کنند، زیرا آنها ناخواسته برای مشارکت استخدام می‌شوند (پیچر، ۲۰۰۷). در واقع، برخی از مطالعات حسابرسی به قدرت آماری بسیار بالایی دست می‌یابند که ممکن است به این معنی باشد که زمان زیادی را از طرف شرکت کنندگان صرف می‌کنند. به عنوان مثال، فاف و همکاران^۱ (۲۰۱۹) بررسی می‌کنند که آیا تعصب مذهبی، دسترسی به خدمات عمومی را شکل می‌دهد یا خیر؟ برای انجام این کار، آنها درخواست‌های جعلی را برای ۴۵۰۰۰ مدیر مدارس دولتی ارسال کردند. با فرض اینکه فرایند درمان به‌طور متوسط دو دقیقه طول کشید، این نشان می‌دهد که محققان از ۱۵۰۰ ساعت کاری مدیران در مدارس دولتی استفاده کردند.

محققان باید هنگام طراحی مطالعات حسابرسی به دلیل عدم رضایت آگاهانه، اهمیت سوال تحقیق و قدرت آماری را به‌طور کامل در نظر بگیرند. نمونه‌ای از این ملاحظات، مطالعه گریملیخویسن و همکاران است (۲۰۱۹) که درخواست‌های خود را تا حد امکان شفاف کردند و در عین حال، موضوع عدم انطباق را در مطالعه خود در مورد تأثیرات قوانین آزادی اطلاعات بر شفافیت کاهش دادند. ما به محققان توصیه می‌کنیم که به تأیید هیأت‌های بازبینی اخلاقی برسند، همان‌طور که گریملیخویسن و همکارانش (۲۰۱۹) در صورتی که احتمال نقض اخلاقی در مطالعه وجود داشته باشد، انجام دادند.

اعتبار داخلی

مانند سایر انواع مطالعات، آزمایش‌های میدانی باید در نظر بگیرند که آیا نتایج آنها از نظر داخلی و خارجی معتبر است یا خیر؟ ملاحظات اعتبار درونی عموماً بر این موضوع تمرکز می‌کند که آیا تأثیرات غیر از ارائه راه‌حل، نتایج شرکت‌کنندگان در مطالعه را تغییر می‌دهد؛ در حالی که اعتبار خارجی به تعمیم‌یافته‌های یک مطالعه به زمینه‌های مختلف مربوط می‌شود.

اعتبار درونی آزمایش‌های میدانی با مفروضات حذف‌پذیری و عدم تداخل مرتبط است. علاوه بر این، فرسایش؛ چالشی برای اعتبار

1. Pfaff et al.
2. Athey and Imbens
3. Simmons, Nelson, and Simonsohn
4. Brady and Collier
5. Dunning
6. Gerber and Green

نتایج ازدست‌رفته دارند، به‌طور سیستماتیک در نتیجه متفاوت هستند از واحدهایی که نتایج ازدست‌رفته ندارند و (۲) فرسایش در گروه‌های آزمایشی متفاوت است (جیمز، جان، و موزلی، ۲۰۱۷). پتانسیل فرسایش در آزمایش‌های میدانی بیشتر از آزمایش‌های آزمایشگاهی است (جیمز، جان، و موزلی، ۲۰۱۷). این تا حدی به این دلیل است که در آزمایش‌های میدانی، کنترل کمتری صورت می‌گیرد و اندازه‌گیری متغیرهای نتیجه را دشوارتر می‌کند. به‌طور کلی، دو نوع راه‌حل برای فرسایش وجود دارد: راه‌حل‌های مبتنی بر طراحی و راه‌حل‌های آماری.

طراحی و اجرای آزمایش‌ها با هدف به‌حداقل رساندن فرسایش، راه‌حلی است که محققان باید در نظر بگیرند (جیمز، جان، و موزلی، ۲۰۱۷). جنسن (۲۰۱۸) از اقداماتی برای به‌حداقل رساندن فرسایش استفاده می‌کند. این راه‌حل در نزدیکی جغرافیایی پاسخ‌دهندگان انجام شد و محققان در پروژه LEAP به‌دنبال شناسایی دلایل فرسایش از قبل برای یافتن راه‌هایی برای کاهش آن بودند.

گربر و گرین (۲۰۱۲)، چندین راه‌حل مبتنی بر طراحی و آمار را برای تطبیق با فرسایش پیشنهاد می‌کنند. یکی از راه‌حل‌های پیشنهادی آن‌ها، تماس با افراد غیرپاسخ‌دهنده تا حد امکان با سطح معینی از منابع است. یک راه‌حل مرتبط، نمونه‌برداری چندمرحله‌ای است. در نمونه‌گیری چند مرحله‌ای، پژوهشگران ابتدا به‌صورت خوشه‌ای از مناطق جغرافیایی نمونه‌برداری می‌کنند؛ سپس در مرحله دوم می‌توانند به‌طور تصادفی از افراد درون این خوشه‌ها نمونه‌برداری کنند (بررسی اجتماعی اروپا ۲۰۱۸).

راه‌حل‌های آماری که آن‌ها پیشنهاد می‌کنند، داده‌های محدودکننده با پرکردن معقول‌ترین داده‌ها یا داده‌هایی هستند که در رابطه با هرگونه اثرهای قابل توجه رفتار مشاهده‌شده محافظه‌کارانه‌تر هستند. چنین راه‌حل‌های آماری، مبتنی بر حدس هستند و سناریوهای خلاف واقع را در نظر نمی‌گیرند (پیرل و مک‌نزی^۲ ۲۰۱۸)؛ اما ممکن است زمانی مطلوب باشند که هدف، نشان دادن این باشد که فرسایش تفاضلی بر استنباط‌های علی تأثیر نمی‌گذارد. ما هیچ آزمایش میدانی مدیریت دولتی با استفاده از راه‌حل‌های آماری برای فرسایش پیدا نکردیم.

اعتبار خارجی

علاوه بر تهدیدهای شناسایی شده برای اعتبار داخلی، تهدیدهایی نیز برای اعتبار خارجی وجود دارد که از جمله آن‌ها می‌توان به عدم انطباق و سوگیری تصادفی اشاره کرد. تهدیدهای اعتبار خارجی، اهمیت دانستن مقدار علاقه را نشان می‌دهد. علاوه بر این، اعتبار

میدانی مدیریت دولتی را پیدا کنیم که به‌صراحت بیان کند که آنها دوگانه هستند. با این حال، متوجه شدیم که محققان در پروژه LEAP به‌دنبال استانداردسازی راه‌حل‌های خود بودند تا فرض استثنایی را محتمل‌تر کنند (به‌عنوان مثال، آن و همکاران، ۲۰۱۹ و جنسن، ۲۰۱۸). آزمایش‌کنندگان قبل از اینکه با مدیران دولتی با آموزش رهبری تحول‌گرا یا تعاملی رفتار کنند، درباره محتوا و ارائه آموزش رهبری توافق کردند (جنسن، ۲۰۱۸)؛ از این‌رو، حتی بدون کورکردن مضاعف، می‌توان اعتماد را افزایش داد که استثنایپذیری، یک فرض قابل اجرا در یک مطالعه است.

تهدید ۲: تداخل

تداخل، زمانی رخ می‌دهد که واحدهای آزمایشی نتایج یکدیگر را تغییر می‌دهند و در نتیجه، اثرهای رفتار را آلوده می‌کنند (باورز، فردریکسون و پاناگوپولس^۱، ۲۰۱۳). این یک سوگیری ایجاد می‌کند که ممکن است هم رو به بالا و هم رو به پایین باشد؛ به این معنی که محققان نمی‌توانند اثرهای علی را در صورت وجود تداخل تخمین بزنند.

مشاهدات فرایند علی ممکن است به دانستن اینکه آیا تداخل بر تخمین‌ها تأثیر می‌گذارد یا خیر کمک کند. محققان ممکن است بررسی کنند که آیا واحدها با یکدیگر تداخل دارند یا خیر و احتمالاً در کدام جهت سوگیری وجود دارد. راهکار دیگر طراحی مطالعاتی با هدف کاهش احتمال تخلف است. به عنوان مثال، محققان ممکن است رفتار را به‌صورت تصادفی دسته‌بندی کنند تا احتمال تداخل در نتایج سوگیری را کاهش دهند. اگر محققان بتوانند سطحی را شناسایی کنند که بعید است واحدها تداخل داشته باشند، تصادفی‌سازی درمان در این سطح ممکن است مناسب‌ترین گزینه باشد. این در چندین مطالعه در مورد تولید مشترک انجام می‌شود که در آن، محققان حمایت زبانی را در سطح مراکز مراقبت از کودکان به‌جای سطح خانواده‌ها تصادفی می‌کنند (جکوبسن، ۲۰۱۳، جکوبسن و آندرسن، ۲۰۱۳، تامسن و جکوبسن، ۲۰۱۵). اگرچه این قدرت آماری را کاهش می‌دهد؛ اما این مساله را نیز کاهش می‌دهد که خانواده‌ها ممکن است به دوستانی در مجاورت خود که پشتیبانی زبانی دریافت نمی‌کنند، کمک کنند.

تهدید ۳: فرسایش

فرسایش زمانی اتفاق می‌افتد که داده‌های نتیجه ازدست‌رفته باشند. هنگامی که دو شرط وجود داشته باشد، فرسایش به یک مشکل برای استنتاج علی تبدیل می‌شود: (۱) واحدهایی که

با این حال، حتی اگر ITT اثر برنامه را اندازه‌گیری می‌کند و شامل افراد ناسازگار می‌شود، اعتبار خارجی آن ممکن است هنگام تعمیم به یک مورد جدید که در آن مکانیسم‌ها ممکن است متفاوت عمل کنند، محدود شود. در این مورد، مورد کاوی^۱ توصیه می‌شود (آتی و ایمبنز، ۲۰۱۷).

مورد کاوی به اثرات درمان در میان پیروان به جای اثرات تخصیص درمان اشاره دارد. این توسط مورد کاوی اندازه‌گیری می‌شود، که اثر علیّ تخصیص درمان برای کسانی است که از درمان تعیین شده خود پیروی کردند. اگر مورد کاوی مورد علاقه باشد، محقق باید یک رگرسیون حداقل مربعات دو مرحله‌ای (2SLS) انجام دهد (گربر و گرین^۲، ۲۰۱۲). محقق مورد کاوی را به دست می‌آورد، که ITT تقسیم بر نسبت انطباق‌کننده‌ها است. هدف برخی از آزمایش‌های میدانی، شناخت مورد کاوی است. گریملیخویجسن و کلیجن (۲۰۱۵) بررسی می‌کنند که آیا شفافیت قضایی، اعتماد عمومی را افزایش و به‌طور تصادفی تشویق برای تماشای یک سریال تلویزیونی درباره قضاوت را گسترش می‌دهد؟ برخی از شهروندان ممکن است تشویق را دریافت نکنند. از آنجا که گریملیخویجسن و کلیجن به اثر شفافیت بر اعتماد (و نه اثر تخصیص تشویق) علاقه‌مند هستند، مورد کاوی را بازبایی می‌کنند.

مورد کاوی همچنین محدودیت‌هایی را برای اعتبار خارجی ایجاد می‌کند. این اثر رفتاری متوسط محلی (LATE) است، زیرا فقط اثرهای علیّ را در بین افرادی که از تکلیف رفتاری خود پیروی می‌کنند، بازبایی می‌کند (آرونو و کارنگی^۳، ۲۰۱۳)؛ بنابراین، مورد کاوی به کل جمعیت تعمیم نمی‌یابد، جایی که مشکلات اجرا ممکن است اثرهای درمان را کاهش دهد.

تهدید ۲: سوگیری تصادفی‌سازی

تهدید دیگر برای اعتبار خارجی، سوگیری تصادفی‌سازی است (هکمن و اسمیت^۴، ۱۹۹۵). سوگیری تصادفی‌سازی به این مفهوم اشاره دارد: کسانی که برای شرکت در کارآزمایی‌های تصادفی‌سازی و کنترل شده ثبت نام می‌کنند، ممکن است به‌طور قابل توجهی با کسانی که ثبت نام نمی‌کنند، متفاوت باشند. این امر، نوعی سوگیری در انتخاب ایجاد می‌کند و مفهوم آن این است که تحقیق در مورد مداخلات ممکن است به کل جمعیت تعمیم داده نشود. به عنوان مثال، واحدهای شرکت‌کننده ممکن است کسانی باشند که به احتمال زیاد، اثرهای مثبت درمان را تجربه می‌کنند، به این معنی که مداخله زمانی که برای کل جمعیت مقیاس‌بندی شود، اثرهای کمتری دارد (اندرسن و هویدمن^۴، ۲۰۱۹).

خارجی ممکن است به آزمایش‌کنندگان میدانی اجازه ندهد که مقیاس‌پذیری پروژه را به دلیل ویژگی‌های خاص واحدهای شرکت‌کننده یا مسائل اجرایی استنباط کنند. به زبان ساده: زمینه ممکن است مهم باشد. ما عدم انطباق و سوگیری تصادفی‌سازی را تهدیدی برای اعتبار خارجی می‌دانیم.

تهدید ۱: عدم انطباق

عدم انطباق زمانی رخ می‌دهد که آزمودنی‌ها با شرایط آزمایشی که به آن اختصاص داده شده‌اند، مطابقت نداشته باشند. هنگامی که عدم انطباق رخ می‌دهد، محققان باید انتخاب کنند که آیا می‌خواهند اثرهای مداخله را بررسی یا اثرهای مداخله را در میان افرادی که از درمان پیروی کرده‌اند، ارزیابی کنند؛ بنابراین، عدم انطباق هنگام انتخاب مقدار مورد علاقه مهم است و ارتباط نزدیکی با اعتبار خارجی دارد. با این حال، عدم انطباق نیز می‌تواند یک موضوع اعتبار داخلی باشد، زیرا ممکن است استنتاج‌های علیّ را بسته به اینکه چگونه محققان برای مدیریت آن انتخاب می‌کنند، سوگیری کند.

می‌توان اثرات علیّ را در صورت عدم انطباق به دو روش اندازه‌گیری کرد: یا از طریق قصد درمان اثر (ITT) یا از طریق میانگین اثر علیّ سازگار (CACE) (آتی و ایمبنز، ۲۰۱۷). ITT اثر تخصیص درمان است که هر گونه مسائل مربوط به عدم انطباق را در بین واحدهای آزمایشی نادیده می‌گیرد. این بدان معناست که ITT اثر کلی مداخله است. به آن اثر برنامه نیز می‌گویند. محققان ممکن است به این مقدار بسیار علاقه‌مند باشند و تصمیم‌گیرندگان ممکن است بخواهند بدانند که آیا یک مداخله همان‌طور که در نظر گرفته شده است کار می‌کند یا خیر. این ممکن است اعتبار خارجی را افزایش دهد؛ زیرا عدم انطباق احتمالاً در تنظیمات دنیای واقعی رخ می‌دهد.

جاکوبسن و اندرسن (۲۰۱۳a) این موضوع را در مقاله خود در مورد تولید مشترک در نظر و نتیجه می‌گیرند که ITT در مطالعه آن‌ها مرتبط‌تر است؛ زیرا آن‌ها به تأثیری که در محیط ارائه خدمات واقعی رخ می‌دهد، از جمله تمام مسائل اجرایی ممکن، علاقه نشان می‌دهند. این یک دلیل معتبر است، زیرا سوال مورد علاقه این است که آیا تولید مشترک باعث افزایش حقوق صاحبان سهام می‌شود یا خیر. توانایی ارائه دانش در مورد این موضوع به جای اینکه شهروندان رعایت کنند، به تکلیف راه‌حل ارایه شده بستگی دارد.

1. CACE

2. Gerber and Green

3. Heckman and Smith

4. Andersen and Hvidman

دارد که آزمایش های میدانی هنوز نتوانسته اند آنها را اثبات کنند. یکی از آنها، نوآوری در بخش دولتی است. از آنجا که نوآوری ها، مداخلات سازمانی هستند، به نظر می رسد چنین مطالعاتی پرهزینه خواهند بود، اگرچه احتمالاً مزایای زیادی برای علم و جامعه خواهند داشت (کلین اسمیت و شین^۳، ۲۰۱۸، اسکیرا و نبوت^۴، ۲۰۱۳). به طور مشابه، مسئولیت اجرایی؛ یک زمینه جدید تحقیق است که در آن آزمایش های میدانی ممکن است استنتاج هایی از ارتباط نظری و عملی بالا را نتیجه دهد. با بررسی چگونگی کاهش هزینه های روان شناختی، یادگیری و انطباق (مونیپهان، هرد و هاروی، ۲۰۱۴) که توسط شهروندان در هنگام دسترسی به خدمات عمومی انجام می شود، این مقالات ممکن است جهش بزرگی در اطلاع رسانی به سیاست گذاران داشته باشند.

د نهایت، اثرهای شبکه ها بر سطح اطلاعات فردی، نگرش ها و تغییر (به عنوان مثال، باندیرا و رسول^۵، ۲۰۰۶)، پتانسیل ارائه می دهد. تحقیقات اندکی در مدیریت دولتی، اهمیت شبکه ها را بر روی نگرش ها و رفتارهای فردی در آزمایش های میدانی ارزیابی می کند؛ اما ممکن است این ها به صورت علی در محیط های واقع بینانه مورد بررسی قرار گیرند تا چگونگی تأثیر جریان اطلاعات بر افراد در سازمان ها و جوامع را بررسی کنند. امیدواریم محققانی که به دنبال پاسخگویی به چنین سؤالاتی هستند به این بررسی روی آورند تا هزینه ها، امکان پذیر بودن، اخلاقی بودن و اعتبار مطالعات خود را در نظر بگیرند که می تواند آزمایش های میدانی بیشتری را در مدیریت دولتی در آینده امکان پذیر کند.

مساله دیگری که ممکن است از پتانسیل آزمایش های میدانی بیشتر در مدیریت عمومی سود ببرد، تحقیق در مورد مسائل روش شناختی در آزمایش های میدانی است. ما پیشنهاد می کنیم که تحقیقات آینده به بررسی چگونگی مشارکت سازمان ها در آزمایش های میدانی، چگونگی افزایش انطباق در آزمایش های میدانی و چگونگی و زمان تداخل بین واحدهای آزمایشی بپردازند. مشارکت دادن سازمان ها در آزمایش های میدانی بسیار دشوار است (ادن^۶، ۲۰۱۷)؛ زیرا آنها معمولاً منافع متفاوتی نسبت به منافع محقق دارند (لویت و لیست^۷، ۲۰۰۹). تحقیقات می توانند بررسی کنند که چگونه نشانه های اطلاعاتی مختلف بر پاسخ های سازمانی و فردی برای شرکت در آزمایش های میدانی تأثیر می گذارند. این امر به آزمایش های میدانی مدیریت دولتی کمک می کند تا سوگیری تصادفی را کاهش دهد (آندرسون و هیدمن^۸، ۲۰۱۹).

این موضوع مربوط به تحقیقات مدیریت دولتی است؛ زیرا مداخلات اغلب با سازمان ها انجام می شود و ممکن است تفاوت های سیستماتیک بین کسانی که مشارکت را انتخاب می کنند و کسانی که مشارکت نمی کنند، وجود داشته باشد. تحقیقات اندکی این مورد را بررسی می کند. به همین دلیل است که ما نمی دانیم آیا آزمایش های میدانی مدیریت دولتی به کل مجموعه موارد مورد علاقه تعمیم می یابد یا خیر. یک پروژه جدید توسط آندرسن و هیدمن (۲۰۱۹) به بررسی تفاوت های بین مدارس شرکت کننده و والدین در مداخله خواندن می پردازد و اینکه آیا تفاوت های پیش زمینه یا تفاوت در اجرا، تأثیرات مداخله را توضیح می دهد؟ به طور کلی، مداخله هیچ اثر آشکاری را نشان نمی دهد. با این حال، محققان داده هایی درباره کل جمعیت مدارس دارند که به آنها امکان می دهد شرکت کنندگان را با افراد غیر شرکت کننده در مطالعه خود مقایسه کنند. این مطالعه، یافته های جالبی را در مورد اینکه چه کسی در آزمایش های میدانی شرکت می کند و اینکه آیا مسائل پیاده سازی اثرهای درمان ناهمگن را توضیح می دهند یا خیر، فراهم می کند. نتایج نشان می دهد که پیاده سازی ضعیف یک مساله مهم تر از اثرهای درمان ناهمگن و انتخاب در آزمایش ها برای این نوع مداخلات است. چنین مطالعه ای ممکن است اولین پیشنهاد را از چگونگی تأثیر سوگیری تصادفی سازی و مسائل پیاده سازی بر مقیاس پذیری آزمایش های میدانی مدیریت دولتی ارائه دهد.

آینده آزمایش میدانی در مدیریت دولتی

این مقاله به بررسی آزمایش های میدانی مدیریت دولتی می پردازد. روند روبه افزایشی در آزمایش میدانی در مدیریت دولتی وجود دارد (مطالعات بسیار اخیر شامل لینوس و ریس^۱، ۲۰۲۰، سندرز و کیر کمین^۲، ۲۰۱۹). انقلاب اعتبار در علم اقتصاد به مدیریت عمومی نیز سرایت کرده است؛ جایی که تمرکز بر شناسایی اثرهای علی همراه با اولویت ها برای راه حل های واقعی، گسترش آزمایش های میدانی را توضیح می دهد. برخی از موضوعات به طور فزاینده ای با استفاده از آزمایش های میدانی مورد مطالعه قرار می گیرند. این ها شامل اطلاعات عملکرد (آندرسن و لارسن^{۱۶}، ۲۰۱۶)، رهبری (بله^{۱۴}، ۲۰۱۴)، تولید مشترک (تامسن و یاکوبسن^{۱۵}، ۲۰۱۵) و تبعیض (جیلکه، ون دورن، و ریس^{۱۸}، ۲۰۱۸) است. با این حال، هنوز چندین موضوع مختلف در مدیریت دولتی وجود

1. Linos and Riesch
2. Sanders and Kirkman
3. Clingsmith and Shane
4. Skiera and Nabout
5. Bandiera and Rasul

6. Eden
7. Levitt and List
8. Andersen and Hvidman

دوم، عدم انطباق موضوعی است که تعمیم‌پذیری ادعاهای تجربی میدانی را محدود می‌کند. اگر تحقیقات، چگونگی وادار کردن افراد را به تبعیت از راه‌حل ارائه‌شده و احتمالاً چگونگی جلوگیری از فرسایش بررسی کند، این امر به آزمایش‌های میدانی آینده در مدیریت دولتی کمک خواهد کرد. درنهایت، تداخل ممکن است نتایج در آزمایش‌های میدانی را منحرف کند؛ به همین دلیل است که اطلاعات در مورد اینکه چه زمانی تداخل یک مساله است، به آزمایش‌کنندگان میدانی کمک می‌کند تا بدانند چه زمانی خوشه‌بندی رفتارهای تصادفی ترجیح داده می‌شود و چه زمانی انجام این کار در سطح فردی مطلوب‌تر است. رشته‌های مرتبط، اثرهای سرریز را در کمپین‌های رأی‌گیری (سینکلر، مک کانل و گرین^۱، ۲۰۱۲) و تصمیمات ثبت‌نام مدرسه (بوبونیس و فاینان^۲، ۲۰۰۹) مدل کرده‌اند؛ اما تداخل ممکن است در زمینه‌های سازمانی عمومی متفاوت عمل کند، به همین دلیل است که آزمایش‌های میدانی مدیریت دولتی ممکن است از تحقیق بیشتر در این زمینه سود ببرد.

را پس از شناخته‌شدن نتایج کاهش می‌دهد که یک علت بسیار محتمل برای بحران تکرار است. انعطاف‌پذیری به محققان اجازه می‌دهد تا به اهمیت آماری دست یابند، بدون توجه به اینکه داده‌ها نشان‌دهنده تأثیرات هستند (سیمونز، نلسون و سیمونسون^۳، ۲۰۱۱). پیش‌ثبت‌نام به محققان این امکان را می‌دهد که قبل از جمع‌آوری داده‌ها در مورد انتخاب‌هایی که انجام می‌دهند فکر کنند و «درجات آزادی محقق» را کاهش دهند (سیمونز، نلسون و سیمونسون، ۲۰۱۱) که نتایج مثبت کاذب را در تحقیقات افزایش می‌دهد.

آزمایش‌های میدانی مدیریت دولتی به‌طور فزاینده‌ای محبوب می‌شوند و دلایل مناسبی هم دارند. آزمایش‌های میدانی ممکن است در مورد اثرهای علی در محیط‌های واقعی اطلاع‌رسانی کنند که انواع دیگر مطالعات به‌ندرت در مورد آن اطلاع‌رسانی می‌کنند. با این حال، آزمایش‌های میدانی کمی در مدیریت دولتی وجود دارد. ما امیدواریم که این بررسی سیستماتیک، محققان را در بررسی چگونگی انجام آزمایش‌های میدانی از مراحل اولیه در حصول اطمینان از همکاری با سازمان‌ها یا شرکت‌کنندگان، به مرحله طراحی که در آن می‌توان به مسائل اعتبار رسیدگی کرد، راهنمایی کند. با نشان دادن اینکه چگونه این ملاحظات در آزمایش‌های میدانی مدیریت دولتی مورد بحث قرار می‌گیرند و ارائه پیشنهادهایی در مورد چگونگی یافتن اطلاعات بیشتر، به دنبال ارائه مقدماتی در آزمایش‌های میدانی برای محققانی که تازه به این روش می‌پردازند، بوده‌ایم. علاوه بر این، ما برنامه‌های تحقیقاتی آینده را پیشنهاد داده‌ایم که امیدواریم آزمایش‌کنندگان میدانی در سال‌های آینده به آن‌ها توجه کنند.

اکنون به ارائه پیشنهادهای خود برای انجام آزمایش‌های میدانی در آینده می‌پردازیم. مطالعات معتبرتر، ابتدا با تمرکز بر پیش‌ثبت‌نام به‌دست می‌آیند. ما فقط دو مطالعه از پیش‌ثبت‌شده را در مرور خود پیدا کردیم (گریملیخوسن و همکاران^۳ ۲۰۱۹؛ منگر و استین^۴ ۲۰۱۸). پیش‌ثبت‌نام، دارای مزایایی است که آزمایش‌کنندگان میدانی مدیریت دولتی می‌توانند از آنها استفاده کنند (نوسک و همکاران^۵، ۲۰۱۸). پیش‌ثبت‌نام به محققان این امکان را می‌دهد تا قبل از انجام مطالعات در مورد انتخاب‌های طراحی فکر کنند. علاوه بر این، پیش‌ثبت‌نام؛ خطر فرضیه‌سازی

1. Sinclair, McConnell, and Green
2. Bobonis and Finan
3. Grimmlikhuijsen et al.
4. Menger and Stein
5. Nosek et al.
6. Simmons, Nelson, and Simonsohn

منابع و مآخذ: 

- Adman, Per, and Hanna Jansson. 2017. A Field Experiment on Ethnic Discrimination among Local Swedish Public Officials. *Local Government Studies* 43(1): 44–63.
- An, Seung-Ho, Kenneth J. Meier, Anne Bøllingtoft, and Lotte Bøgh Andersen. 2019. Employee Perceived Effect of Leadership Training: Comparing Public and Private Organizations. *International Public Management Journal* 22(1): 2–28.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- ———. 2010. The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *Journal of Economic Perspectives* 24(2): 3–30.
- Andersen, Simon Calmar, and Ulrik Hvidman. 2019. Implementing Parent-Aimed Programs at Scale: Evidence from a Randomized Controlled Trial. Working paper.
- Andersen, Simon Calmar, and Morten H. Larsen. 2016. Cognitive Biases in Performance Evaluations. *Journal of Public Administration Research and Theory* 26(4): 647–62.
- Andersen, Simon Calmar, and Donald P. Moynihan. 2016. How Leaders Respond to Diversity: The Moderating Role of Organizational Culture on Performance Information Use. *Journal of Public Administration Research and Theory* 26(3): 448–60.
- ———. 2018. How Do Socially Distinctive Newcomers Fare? Evidence from a Field Experiment. *Public Administration Review* 78(6): 874–82.
- Andersen, Simon Calmar, Helena S. Nielsen, and Mette Kjærgaard Thomsen. 2018. How to Increase Citizen Coproduction: Replication and Extension of Existing Research. *International Public Management Journal*. Published online December 20. <https://doi.org/10.1080/10967494.2018.1518851>.
- Aronow, Peter M., and Allison Carnegie. 2013. Beyond LATE: Estimation of the Average Treatment Effect with an Instrumental Variable. *Political Analysis* 21(4): 492–506.
- Athey, Susan, and Guido W. Imbens. 2017. The Econometrics of Randomized Experiments. In *Handbook of Field Experiments*, edited by Abhijit V. Banerjee and Es-



ther Duflo, 73–140. Oxford: Elsevier.

- Baekgaard, Martin, Caroline Baethge, Jens Blom-Hansen, Claire A. Dunlop, Marc Esteve, Morten Jakobsen, Brian Kisida, John Marvel, Alice Moseley, Søren Serritzlew, Patrick Stewart, Mette Kjærgaard Thomsen, and Patrick J. Wolf. 2015. Conducting Experiments in Public Management Research: A Practical Guide. *International Public Management Journal* 18(2): 323–42.
- Bandiera, Oriana, and Imran Rasul. 2006. Social Networks and Technology Adoption in Northern Mozambique. *Economic Journal* 116(514): 869–902.
- Battaglio, Paul R., Jr., Paolo Belardinelli, Nicola Bellé, and Paola Cantarelli. 2019. Behavioral Public Administration ad fontes: A Synthesis of Research on Bounded Rationality, Cognitive Biases, and Nudging in Public Organizations. *Public Administration Review* 79(3): 304–20.
- Bellé, Nicola. 2013. Experimental Evidence on the Relationship between Public Service Motivation and Job Performance. *Public Administration Review* 73(1): 143–53.
- ———. 2014. Leading to Make a Difference: A Field Experiment on the Performance Effects of Transformational Leadership, Perceived Social Impact, and Public Service Motivation. *Journal of Public Administration Research and Theory* 24(1): 109–36.
- ———. 2015. Performance-Related Pay and the Crowding Out of Motivation in the Public Sector: A Randomized Field Experiment. *Public Administration Review* 75(2): 230–41.
- *Bellé, Nicola, and Paola Cantarelli. 2019. Do Ethical Leadership, Visibility, External Regulation, and Prosocial Impact Affect Unethical Behavior? Evidence from a Laboratory and a Field Experiment. *Review of Public Personnel Administration* 39(3): 349–71.
- Bellé, Nicola, Paola Cantarelli, and Paolo Belardinelli. 2017. Cognitive Biases in Performance Appraisal: Experimental Evidence on Anchoring and Halo Effects with Public Sector Managers and Employees. *Review of Public Personnel Administration* 37(3): 275–94.
- Ben-Aaron, James, Matthew Denny, Bruce Desmarais, and Hanna Wallach. 2017. Transparency by Conformity: A Field Experiment Evaluating Openness in Local Governments. *Public Administration Review* 77(1): 68–77.
- Bhanot, Syon P., Gordon Kraft-Todd, David Rand, and Erez Yoeli. 2018. Putting Social Rewards and Identity Salience to the Test: Evidence from a Field Experiment with



- Teachers in Philadelphia. *Journal of Behavioral Public Administration* 1(1): 1–16.
- Bobonis, Gustavo J., and Frederico Finan. 2009. Neighborhood Peer Effects in Secondary School Enrollment Decisions. *Review of Economics and Statistics* 91(4): 695–716.
 - Bowers, Jake, Mark M. Frederickson, and Costas Panagopoulos. 2013. Reasoning about Interference between Units: A General Framework. *Political Analysis* 21(1): 97–124.
 - Brady, Henry E., and David Collier. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman & Littlefield.
 - Broockman, David E., Joshua L. Kalla, and Jasjeet S. Sekhon. 2017. The Design of Field Experiments with Survey Outcomes. A Framework for Selecting More Efficient, Robust, and Ethical Designs. *Political Analysis* 25(4): 435–64.
 - Clingsmith, David, and Scott Shane. 2018. Training Aspiring Entrepreneurs to Pitch Experienced Investors: Evidence from a Field Experiment in the United States. *Management Science* 64(11): 5164–79.
 - Czibor, Eszter, David Jimenez-Gomez, and John A. List. 2019. The Dozen Things Experimental Economists Should Do (More of). *Southern Economic Journal* 86(2): 371–432.
 - Darolia, Rajeev, Cory Koedel, Paco Martorell, Katie Wilson, and Francisco Perez-Arce. 2015. Do Employers Prefer Workers Who Attend For-Profit Colleges? Evidence from a Field Experiment. *Journal of Policy Analysis and Management* 34(4): 881–903.
 - Dudau, Adina, Georgios Kominis, and Melinda Szocs. 2018. Innovation Failure in the Eye of the Beholder: Towards a Theory of Innovation Shaped by Competing Agendas within Higher Education. *Public Management Review* 20(2): 254–72.
 - Dunning, Thad. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge: Cambridge University Press.
 - Eden, Dov. 2017. Field Experiments in Organizations. *Annual Review of Organizational Psychology and Organizational Behavior* 4: 91–122.
 - European Social Survey. 2018. European Social Survey Round 9 Sampling Guidelines: Principles and Implementation. https://www.europeansocialsurvey.org/docs/round9/methods/ESS9_sampling_guidelines.pdf [accessed February 27, 2020].
 - Falk, Armin, and James J. Heckman. 2009. Lab Experiments Are a Major Source of Knowledge in the Social Sciences. *Science* 23: 535–8.
 - Feigenberg, Benjamin, Erica Field, Rohini Pande, Natalia Rigol, and Shayak Sarkar.



2014. Do Group Dynamics Influence Social Capital Gains among Microfinance Clients? Evidence from a Randomized Experiment in Urban India. *Journal of Policy Analysis and Management* 33(4): 932–49.
- Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton.
 - Glennerster, Rachel. 2017. The Practicalities of Running Randomized Evaluations: Partnerships, Measurement, Ethics, and Transparency. In *Handbook of Field Experiments*, edited by Abhijit V. Banerjee and Esther Duflo, 175–243. Oxford: Elsevier.
 - Glennerster, Rachel, and Kudzai Takavarasha. 2013. *Running Randomized Evaluations*. Princeton, NJ: Princeton University Press.
 - Grimmelikhuijsen, Stephan, Sebastian Jilke, Asmus Leth Olsen, and Lars G. Tummers. 2017. Behavioral Public Administration: Combining Insights from Public Administration and Psychology. *Public Administration Review* 77(1): 45–56.
 - Grimmelikhuijsen, Stephan, Peter John, Albert Meijer, and Ben Worthy. 2019. Do Freedom of Information Laws Increase the Transparency of Government? A Pre-registered Replication of a Field Experiment. *Journal of Behavioral Public Administration* 1(2): 1–10.
 - Grimmelikhuijsen, Stephan, and Albert Klijn. 2015. The Effects of Judicial Transparency on Public Trust: Evidence from a Field Experiment. *Public Administration* 93(4): 995–1011.
 - Grohs, Stephan, Christian Adam, and Cristoph Knill. 2016. Are Some Citizens More Equal than Others? Evidence from a Field Experiment. *Public Administration Review* 76(1): 155–64.
 - Haynes, Laura C., Donald P. Green, R. Gallagher, Peter John, and David J. Torgerson. 2013. Collection of Delinquent Fines: An Adaptive Randomized Trial to Assess the Effectiveness of Alternative Text Messages. *Journal of Policy Analysis and Management* 32(4): 718–30.
 - Heckman, James J., and Jeffrey A. Smith. 1995. Assessing the Case for Social Experiments. *Journal of Economic Perspectives* 9(2): 85–110.
 - Hess, Douglas R., Michael J. Hanmer, and David W. Nickerson. 2016. Encouraging Local Compliance with Federal Civil Rights Laws: Field Experiments with the Na-

- tional Voter Registration Act. *Public Administration Review* 76(1): 165–74.
- Hock, Scott, Sarah Anderson, and Matthew Potoski. 2013. Invitation Phone Calls Increase Attendance at Civic Meetings: Evidence from a Field Experiment. *Public Administration Review* 73(2): 221–8.
 - Jacobsen, Christian Bøtcher, and Lotte Bøgh Andersen. 2019. High Performance Expectations: Concept and Causes. *International Journal of Public Administration* 41(11): 1–11.
 - Jakobsen, Morten. 2013. Can Government Initiatives Increase Citizen Co-production? Results of a Randomized Field Experiment. *Journal of Public Administration Research and Theory* 23(1): 27–54.
 - Jakobsen, Morten, and Simon Calmar Andersen. 2013a. Coproduction and Equity in Public Service Delivery. *Public Administration Review* 73(5): 704–13.
 - ———. 2013b. Intensifying Social Exchange Relationships in Public Organizations: Evidence from a Randomized Field Experiment. *Journal of Policy Analysis and Management* 32(1): 60–82.
 - James, Oliver. 2010. Performance Measures and Democracy: Information Effects on Citizens in Field and Laboratory Experiments. *Journal of Public Administration Research and Theory* 21(3): 399–418.
 - ———. 2011. Managing Citizens' Expectations of Public Service Performance: Evidence from Observation and Experimentation in Local Government. *Public Administration* 89(4): 1419–35.
 - James, Oliver, and Alice Moseley. 2014. Does Performance Information about Public Services Affect Citizens' Perceptions, Satisfaction, and Voice Behavior? Field Experiments with Absolute and Relative Performance Information. *Public Administration* 92(2): 493–511.
 - James, Oliver, Peter John, and Alice Moseley. 2017. Field Experiments in Public Management. In *Experiments in Public Management Research: Challenges and Contributions*, edited by Oliver James, Sebastian R. Jilke, and Gregg G. van Ryzin, 89–116. Cambridge, Cambridge University Press.
 - Jensen, Ulrich T. 2018. Does Perceived Societal Impact Moderate the Effect of Trans-



- formational Leadership on Value Congruence? Evidence from a Field Experiment. *Public Administration Review* 78(1): 48–57.
- Jilke, Sebastian, Wouter van Dooren, and Sabine Rys. 2018. Discrimination and Administrative Burden in Public Service Markets: Does a Public–Private Difference Exist? *Journal of Public Administration Research and Theory* 28(3): 423–39.
 - Jilke, Sebastian, Jiahuan Lu, Chengxin Xu, and Shugo Shinohara. 2019. Using Large Scale Social Media Experiments in Public Administration: Assessing Charitable Consequences of Government Funding of Nonprofits. *Journal of Public Administration Research and Theory* 29(4): 627–639.
 - John, Peter, and Toby Blume. 2018. How Best to Nudge Taxpayers? The Impact of Message Simplification and Descriptive Social Norms on Payment Rates in a Central London Local Authority. *Journal of Behavioral Public Administration* 1(1). <https://doi.org/10.30636/jbpa.11.10>.
 - Kisida, Brian, and Patrick J. Wolf. 2015. Customer Satisfaction and Educational Outcomes: Experimental Impacts of the Market-Based Delivery of Public Education. *International Public Management Journal* 18(2): 265–85.
 - Levitt, Steven D., and John A. List. 2007. What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World? *Journal of Economic Perspectives* 21(2): 153–74.
 - ———. 2009. Field Experiments in Economics: The Past, the Present, and the Future. *European Economic Review* 53(1): 1–18.
 - Linos, Elizabeth. 2018. More than Public Service: A Field Experiment on Job Advertisements and Diversity in the Police. *Journal of Public Administration Research and Theory* 28(1): 67–85.
 - Linos, Elizabeth, Lisa T. Quan, and Elspeth Kirkman. 2020. Nudging Early Reduces Administrative Burden: Three Field Experiments to Improve Code Enforcement. *Journal of Policy Analysis and Management* 39(1): 243–65.
 - Linos, Elizabeth, Joanne Reinhard, and Simon Ruda. 2017. Levelling the Playing Field in Police Recruitment: Evidence from a Field Experiment on Test Performance. *Public Administration* 95: 943–56.

- Linos, Elizabeth, and Nefara Riesch. 2020. Thick Red Tape and the Thin Blue Line: A Field Study on Reducing Administrative Burden in Police Recruitment. *Public Administration Review* 80(1): 92–103.
- List, John A. 2011. Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off. *Journal of Economic Perspectives* 25(3): 3–16.
- Lonati, Sirio, Bernardo F. Quiroga, Christian Zehnder, and John Antonakis. 2018. On Doing Relevant and Rigorous Experiments: Review and Recommendations. *Journal of Operations Management* 64(1): 19–40.
- Menger, Andrew, and Robert M. Stein. 2018. Enlisting the Public in Facilitating Election Administration: A Field Experiment. *Public Administration Review* 78(6): 892–903.
- Moher, David, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G. Altman. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine* 151(4): 264–9.
- Moynihan, Donald P., Pamela Herd, and Hope Harvey. 2014. Administrative Burden: Learning, Psychological, and Compliance Costs in Citizen-State Interactions. *Journal of Public Administration Research and Theory* 25(1): 43–69.
- Mutz, Diana. 2011. *Population-Based Survey Experiments*. Princeton: Princeton University Press.
- Nosek, Brian, Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. The Preregistration Revolution. *Proceedings of the National Academy of Sciences* 115(11): 2600–6.
- Open Science Collaboration. 2015. Estimating the Reproducibility of Psychological Science. *Science* 349(6251): aac4716.
- Pager, Devah. 2007. The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future. *Annals of the American Academy of Political and Social Science* 609: 104–33.
- Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.
- Peisakhin, Leonid, and Paul Pinto. 2010. Is Transparency an Effective Anti-corruption Strategy? Evidence from a Field Experiment in India. *Regulation & Governance* 4(3):



261–80.

- Pfaff, Steven, Charles Crabtree, Holger Kern, and John B. Holbein. 2019. Does Religious Bias Shape Access to Public Services? A Large-Scale Audit Experiment among Street-Level Bureaucrats. SocArXiv. <https://osf.io/preprints/socarxiv/9khds/>.
- Porter, Ethan, and John C. Rogowski. 2018. Partisanship, Bureaucratic Responsiveness, and Election Administration: Evidence from a Field Experiment. *Journal of Public Administration Research and Theory* 28(4): 602–17.
- Sanders, Michael, and Elspeth Kirkman. 2019. I've Booked You a Place, Good Luck: A Field Experiment Applying Behavioral Science to Improve Attendance at High Impact Recruitment Events. *Journal of Behavioral Public Administration* 2(1). <https://doi.org/10.30636/jbpa.21.24>.
- Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science* 22(11): 1359–66.
- Sinclair, Betsy, Margaret McConnell, and Donald P. Green. 2012. Detecting Spillover Effects: Design and Analysis of Multilevel Experiments. *American Journal of Political Science* 56(4): 1055–69.
- Skiera, Bernd, and Abou Nabouh. 2013. PROSAD: A Bidding Decision Support System for Profit Optimizing Search Engine Advertising. *Marketing Science* 32(2): 213–20.
- Thomsen, Mette Kjærgaard, and Morten Jakobsen. 2015. Influencing Citizen Co-production by Sending Encouragement and Advice: A Field Experiment. *International Public Management Journal* 18(2): 286–303.
- Vashdi, Dana R. 2013. Teams in Public Administration: A Field Study of Team Feedback and Effectiveness in the Israeli Public Healthcare System. *International Public Management Journal* 16(2): 275–306.
- Worthy, Ben, Peter John, and Matia Vannoni. 2017. Transparency at the Parish Pump: A Field Experiment to Measure the Effectiveness of Freedom of Information Requests in England. *Journal of Public Administration Research and Theory* 27(3): 485–500.