منظمة العمل الدولية

WHAT WORKS
In Youth Employment

دليل قياس الوظائف اللائقة للشباب

الرصد والتقييم والتعلّم في برامج سوق العمل النشطة

الملاحظة 5

طرق تقييم أثر التدخلات التي تستهدف عمالة الشباب

دليل قياس الوظائف اللائقة للشباب

الرصد والتقييم والتعلّم في برامج سوق العمل النشطة

الملاحظة 5
طرق تقييم أثر التدخلات التي تستهدف عمالة الشباب

حقوق النشر محفوظة لمنظمة العمل الدولية 2018

الطبعة الأولى بالإنكليزية 2018

تتمتع منشورات مكتب العمل الدولي بحماية حقوق المؤلف بموجب البروتوكول رقم 2 المرفق بالاتفاقية العالمية لحماية حقوق المؤلف، على انه يجوز نقل مقاطع قصيرة منها بدون إذن، شريطة أن يشار حسب الأصول إلى مصدرها. وأي طلب للحصول على إذن أو ترجمة يجب أن يوجه إلى منشورات مكتب العمل الدولي (الحقوق والتراخيص)، جنيف (العنوان أدناه) أو عبر البريد الالكتروني: rights@ilo.org. يرحب مكتب العمل الدولي بهذه الطلبات.

يجوز للمكتبات والمؤسسات والمستخدمين الآخرين المسجلين مع منظمات حقوق الاستنساخ النسخ وفقاً للتراخيص الممنوحة لهم لهذا الغرض. للعثور على منظمات حقوق الاستنساخ في بلدكم، يمكنكم زيارة www.ifrro.org.

منظمة العمل الدولية
دليل قياس الوظائف اللائقة للشباب. الرصد والتقييم والتعلّم في برامج سوق العمل النشطة - الملاحظة 5. طرق تقييم أثر التدخلات التي تستهدف عمالة الشباب، منظمة العمل الدولية، 2018

ISBN: 978-92-2-630795-7 (print)
78-92-2-630796-4 (web pdf)

Also available in English: Guide on Measuring Decent Jobs for Youth - Monitoring, evaluation and learning in labour market programmes - Note 5: Identifying an appropriate impact evaluation method. ISBN 978-92-2-131664-0 (print), 978-92-2-131665-7 (web pdf), Geneva, 2018.

لا تنطوي التسميات المستخدمة في منشورات مكتب العمل الدولي، التي تتفق مع تلك التي تستخدمها الأمم المتحدة، ولا العرض الوارد للمادة التي تتضمنها، على التعبير عن أي رأي من جانب مكتب العمل الدولي بشان المركز القانوني لأي بلد أو منطقة أو إقليم، أو لسلطات أي منها، أو بشان تعيين حدودها.

ومسؤولية الآراء المعبر عنها في المواد أو الدراسات أو المساهمات الأخرى التي تحمل توقيعا هي مسؤولية مؤلفيها وحدهم، ولا يمثل النشر مصادقة من جانب مكتب العمل الدولي على الآراء الواردة بها.

والإشارة إلى أسماء الشركات والمنتجات والعمليات التجارية لا تعني مصادقة مكتب العمل الدولي عليها، كما إن إغفال ذكر شركات ومنتجات أو عمليات تجارية ليس علامة على عدم إقرارها.

ويمكن الحصول على مطبوعات مكتب العمل الدولي عن طريق المكتبات الكبرى أو مكاتب منظمة العمل الدولية الموجودة في كثير من البلدان أو مباشرة من قسم المطبوعات على العنوان التالي:

ILO Publications
International Labour Office
CH – 1211, Geneva 22 – Switzerland

يمكن طلب مجاناً قائمة بالمنشورات والمطبوعات الجديدة من العنوان المذكور أعلاه أو عن طريق البريد الإلكتروني: pubvente@ilo.org ومن خلال شبكة الإنترنت: www.ilo.org/publns

---

طبع في سويسرا

التصميم والتخطيط للمركز الدولي للتدريب لمنظمة العمل الدولية، تورينو، إيطاليا.

المحتويات

تحدي الإسناد 2

التحديات المحددة لتقييم برامج سوق العمل النشطة التي تركز على الشباب 7

الأساليب الكمية لتقييم الأثر 10

التوزيع العشوائي – التصميم العشوائي 12

كيف يعمل 12

متى يمكن استخدام التصميم العشوائي؟ 15

المزايا 15

العيوب 15

تكييف التصميمات العشوائية مع السياقات المختلفة 18

التصميم العشوائي المرحلي 19

تصميم الترويج العشوائي 20

الفرق في الاختلافات (DID) 22

كيف يعمل 22

متى يمكن استخدام تصميم الفرق في الاختلافات؟ 24

المزايا 24

العيوب 24

مطابقة درجة الميل (PSM) 26

كيف تعمل 26

متى يمكن استخدام مطابقة درجة الميل؟ 28

المزايا 28

العيوب 28

تصميم انحدار الانقطاع (RDD) 31

كيف يعمل 31

متى يمكن استخدام تصميم انحدار الانقطاع؟ 32

المزايا 32

العيوب 32

مقارنات بسيطة: قبل وبعد 34

تحسين ملاءمة تقييمات الأثر الكمي 36

قياس مجموعة متنوعة من التأثيرات 36

الجمع بين المناهج الكمية والنوعية 37

النقاط الرئيسية 41

الموارد الرئيسية 41

المراجع 41

دراسة حالة: تقييم نمو المشاريع الصغيرة في المناطق الريفية عبر طرق التقييم المختلفة 43

الأشكال

5-1: بيان مرئي لأثر التدخل 3

5-2: مراعاة جميع أساليب التقييم الممكنة خلال مرحلة التخطيط 10

5-3: خطوات التصميم العشوائي 12

5-4: اختيار عينات للبرامج الصغرى والكبرى 14

5-5: تصميم التقييم 16

5-6: متابعة برنامج المشروع 21

5-7: مثال عن تحليل الفرق في الاختلافات 23

5-8: مقارنة بين المشاركين وغير المشاركين 27

5-9: تصميم تقييم الأثر (مبسط) 29

5-10: الآثار على نتائج المشاركة في سوق العمل 30

5-11: رسم بياني عن عينة للانقطاع 31

5-12: الانقطاع في احتمالية مشاركة المقاطعات في البرنامج 33

5-13: مقارنة النتائج قبل وبعد 35

الجداول

5-1: حجم العينة المطلوبة للكشف عن الآثار المهمة 9

5-2: نظرة عامة على طرق تقييم الأثر المختلفة 11

5-3: فئات أسئلة تقييم الأثر 38

طرق تقييم أثر التدخلات التي تستهدف عمالة الشباب

الشروط المسبقة:

يستحسن امتلاك فهم أساسي لأساليب البحث الكمي. توضح هذه الملاحظة عدداً من أساليب تقييم الأثر التي يشيع استخدامها وتشرح محاسن ومساوئ كل منها، بما في ذلك الاعتبارات النظرية والعملية.

أهداف التعلم:

في نهاية هذه الملاحظة، سيتمكن القارئ من:

- استخلاص الاعتبارات والتحديات الرئيسية التي يتعين مراعاتها عند السعي إلى إحداث تأثير من خلال طرح السؤال التالي: "ماذا كان سيحدث لنفس الأشخاص / الأسر / المجتمع إذا لم يتم تنفيذ التدخل؟"
- بناء وضع مغاير للواقع من أجل معرفة التغير في النتائج التي يمكن أن يُعزى سببه إلى التدخل، وتحديد الخصائص الأساسية التي يجب أن تشاركها مجموعات المعالجة والمقارنة لضمان الصلاحية الداخلية.
- مقارنة إيجابيات وسلبيات تقنيات التقييم المختلفة وكيف تهدف إلى القضاء على **التحيّز في الاختيار**
- فهم طرق البحث الكمي المختلفة فهماً شاملاً، من التصاميم العشوائية الكاملة إلى الأساليب شبه التجريبية، مثل الفرق في الاختلافات، ومطابقة درجة الميل، وتصميم انحدار الانقطاع
- استخدام أساليب نوعية، ليس فقط لمعرفة ما "حدث" - تحديد متوسط تأثير المعالجة الناشئ عن التدخل – وبل لمعرفة "لماذا".

الكلمات الرئيسية:

الإسناد، المقارنة قبل وبعد، مجموعة المقارنة، الوضع المغاير للواقع، الفرق في الاختلافات، الصلاحية الخارجية، الصلاحية الداخلية، التصميم العشوائي، تتبع العملية، مطابقة درجة الميل، التصميم العشوائي المرحلي، تصميم الترويج العشوائي، تصميم انحدار الانقطاع، مجموعة المعالجة.

توفر هذه الملاحظة[1] للممارسين لمحة عامة عن مختلف الأدوات المتاحة لتقييم الأثر وتقدم إرشادات حول تحديدها في ظل ظروف محددة، وكيفية تنفيذ هذه الأدوات لتقييم آثار التدخلات التي تستهدف عمالة الشباب. بينما يمكن أن تستند تقييمات الأثر إلى كل من الأساليب الكمية والنوعية، إلا أن هذه الملاحظة تركز في المقام الأول على الأساليب الكمية وتقدم أساليب نوعية من أجل توفير مكمل قيّم في سياق نُهج الأساليب المختلطة.

تحدي الإسناد

قبل الاستطراد في البحث، نحتاج أولاً إلى توضيح ما نعنيه بالأثر. في الملاحظات السابقة من هذا الدليل، استخدمنا هذا المصطلح للإشارة إلى الأهداف أو النتائج العليا للبرنامج التي ترتبط بالتغيرات في مستويات المعيشة عموماً، مثل الحد من الفقر أو زيادة رفاهية الأفراد والأسر. ومع ذلك، في سياق تقييمات الأثر، فإننا نفهم الأثر بشكل أضيق على أنه التغيير في النتائج (مثل حالة التوظيف، ووقت العمل، والأرباح) الذي يمكن أن يُعزى سببه إلى تدخلنا.

وكما تم مناقشته في الملاحظة 4، تحاول تقييمات الأثر الإجابة عن الأسئلة السببية؛ بمعنى هل يحسّن التدخل (السبب) النتائج بين المستفيدين (الأثر). فمثلا:

- هل يمكن إسناد سبب التغيّرات الملحوظة في احتمالية حصول المتدربين على عمل إلى تدخل التدريب المهني؟
- هل يؤدي التدخل القائم على تقديم التوجيه والمشورة إلى ارتفاع نسبة الرضا بين صفوف أصحاب العمل وارتفاع نسبة الاحتفاظ بالوظائف؟
- هل يعزز التدخل القائم على تقديم التوجيه والمشورة إلى الشركات الناشئة من فرص إقامة المشاريع والاستدامة؟

يتم تحديد نتائج سوق العمل التي نهتم بها بالعديد من العوامل المعقدة، مثل السياق العام للتنمية الاجتماعية والاقتصادية، والتغيرات في الظروف السياسية و/أو الشخصية، وما إلى ذلك. ومن ثم، ينطوي تحديد درجة التغيرات في هذه النتائج التي يمكن **إسناد** سببها إلى تدخل معين على تحدٍ. يتمثل الغرض من تقييم الأثر تحديداً في التغلب على تحدي الإسناد المذكور عن طريق قياس مدى مساهمة برنامج معين، وهذا البرنامج فقط، في التغيير بالنتائج ذات الاهتمام.

تعريف

الإسناد: إسناد ارتباط سببي بين التغيرات المرصودة (أو المتوقع حدوثها) والتدخل المحدد.

[1] تعتمد هذه الملاحظة على مواد تم تطويرها في الأصل من قبل دوفلو وآخرون (2016)، وخاندكر وآخرون (2010)، وجيرتلر وآخرون (2016) مع تكييف بعض المواد والرسوم التوضيحية في مجال عمالة الشباب وتقديم عرض أكثر إيجازاً لأساليب تقييم الأثر.

بعبارة أخرى، تحاول تقييمات الأثر تقييم إن كانت التغييرات الملحوظة في النتائج ذات الاهتمام يمكن أن تُعزى إلى تدخل أو مشروع معين مع تحديد نطاق ذلك الإسناد وأسبابه.

ينصب التركيز في هذه الملاحظة على ما يسمى **الوضع المغاير للواقع** الذي تستند إليه عادة تقييمات الأثر الكمي.[2] يحدد هذا النهج أثر التدخل على أنه الفرق بين النتائج الملحوظة في إطار التدخل وما يسمى بالسيناريو المغاير للواقع: "ما الذي كان سيحدث لنفس الأشخاص / الأسر / المجتمع إذا لم يتم تنفيذ التدخل؟". يصور الشكل 1-5 مفهومي الأثر والوضع المغاير للواقع.

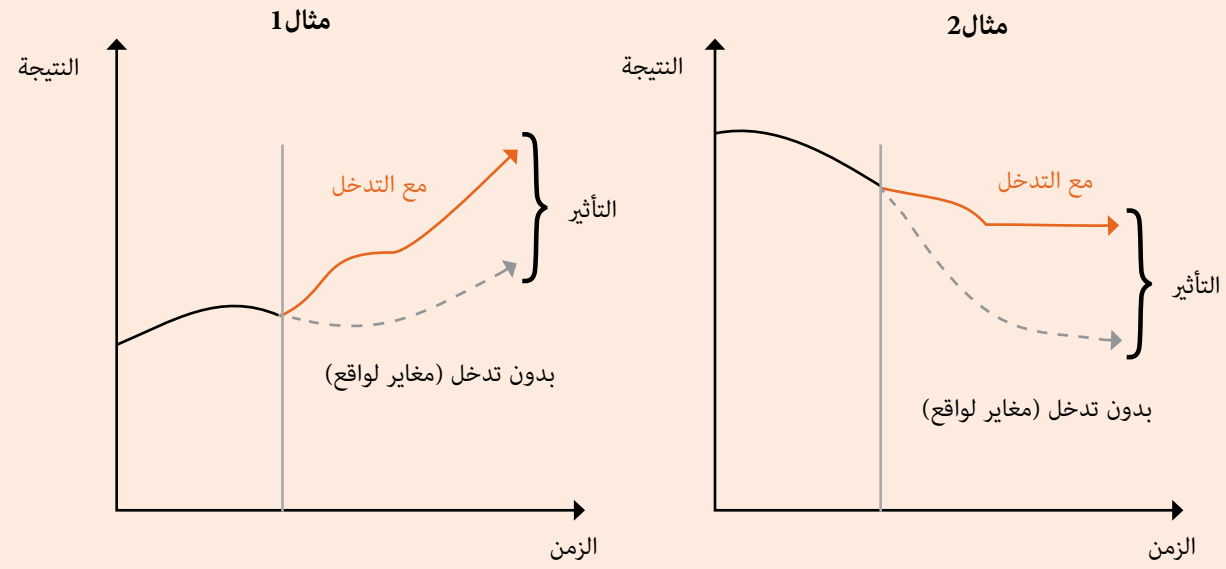
من المستحيل عملياً قياس الوضع المغاير للواقع بشكل حقيقي. تحاول طرق تقييم الأثر تحديد التأثيرات السببية من خلال تقدير أو بناء الوضع المغاير للواقع وذلك نموذجياً - ولكن ليس دائماً - مع مجموعات المقارنة، التي تعرف أحياناً باسم مجموعة التحكم. تُعرف مجموعة المشاركين باسم مجموعة المعالجة أو مجموعة المشاركين.

2 قد يكون تصميم تقييم الأثر الذي يمزج بين الأساليب النوعية والكمية مناسباً اعتماداً على التدخل الجاري تقييمه وسياقه، على النحو الموضح بمزيد من التفصيل أدناه.

تعريف

الوضع المغاير للواقع: يصف الوضع المغاير للواقع نتيجة معينة كان سيعيشها أي مشارك في البرنامج إذا لم يتم تنفيذ البرنامج. بحكم التعريف، لا يمكن ملاحظة الوضع المغاير للواقع بشكل مباشر. لذلك، يجب تقديره، على سبيل المثال، باستخدام مجموعات المقارنة.

الشكل 5-1: بيان مرئي لأثر التدخل



يجب أن تشترك **مجموعات المعالجة والمقارنة** في نفس الخصائص في ثلاث طرق على الأقل (غيرتلر وآخرون، 2016):

1. ينبغي أن تكون مشابهة من حيث الخصائص القابلة وغير القابلة للرصد: يمكن أن تشمل **الخصائص القابلة للرصد** العمر والجنس ومستوى التعليم، والوضع الاجتماعي والاقتصادي، وخصائص الأسرة، والوضع الوظيفي، وما شابه ذلك. ويمكن أن تشمل **الخصائص غير قابلة للرصد** من بين عوامل أخرى الدوافع والاهتمامات والقيم والعقائد ومستوى الدعم الأسري. لا يشترط تطابق جميع الأشخاص في مجموعة المعالجة مع جميع الأشخاص في مجموعة المقارنة، ولكن يجب أن تتشارك المجموعتان في نفس متوسط الخصائص.

2. يجب أن يكون من المتوقع تفاعل مجموعات المعالجة والمقارنة مع التدخل بطريقة مشابهة: على سبيل المثال، يجب أن يكون من المرجح زيادة النتائج، مثل المهارات أو الدخل، لصالح أعضاء مجموعة المعالجة كما هو الحال بالنسبة لمجموعة المقارنة.

3. يجب أن يكون لمجموعات المعالجة والمقارنة مستويات مماثلة من التعرض للتدخلات الأخرى: على سبيل المثال، يجب أن يكون لدى كلتا المجموعتين نفس فرص الحصول على خدمات الدعم الأخرى المقدمة من قبل الحكومة المحلية، والمنظمات غير الحكومية، وما إلى غير ذلك.

4. عندما تشترك مجموعات المعالجة والمقارنة في أوجه التشابه المذكورة أعلاه، نستطيع أن نستنتج بثقة أن أي اختلافات نراها في النتائج بين المجموعتين يمكن أن تُعزى إلى التدخل. من ناحية أخرى، إذا اختلفت مجموعة المقارنة عن مجموعة المعالجة بشكل كبير، فإن مقارنات النتائج بين مجموعات المعالجة والمقارنة لن تعكس أثر التدخل فحسب، بل وتبعات هذه الاختلافات كذلك. وهذا ما يسمى التحيّز في الاختيار.

تعريف

مجموعة المقارنة هي مجموعة مستخدمة لتقدير الوضع المغاير للواقع في تقييم الأثر. وعلى النقيض من أعضاء مجموعة المعالجة، لم يتعرض أعضاء مجموعة المقارنة إلى التدخل المراد تقييمه. وغالباً ما يتم استخدام المصطلحين "مجموعة المقارنة " و"مجموعة المراقبة" بالتبادل. ولأغراض هذه الوثيقة، سوف نستخدم المصطلح العام مجموعة المقارنة في جميع أجزاء الوثيقة.

تعريف

مجموعة المعالجة: تُعرف مجموعة الأشخاص الذين يشاركون بفاعلية في التدخل باسم مجموعة المعالجة أو مجموعة المشاركين

تعريف

الخصائص القابلة للرصد وغير القابلة للرصد: يمكن قياس الخصائص القابلة للرصد من خلال أساليب جمع البيانات المناسبة (مثل الدراسات الاستقصائية) وغالباً ما تتضمن العمر والجنس والمستوى التعليمي والحالة الاجتماعية والاقتصادية والخصائص العائلية والوضع الوظيفي وما إلى ذلك. أما الخصائص غير القابلة للرصد فهي تلك العوامل التي لا يمكن قياسها أو لا يتم قياسها في تقييم (الأثر) ويمكن أن تشمل التحفيز والاهتمام والقيم والإيديولوجيات ومستوى الدعم الأسري، علماً بأنه تم تطوير مقاييس مساعدة بالنسبة للعديد من تلك الخصائص غير القابلة للرصد (غير الكاملة)،

يحدث **التحيّز في الاختيار** في العادة عندما يختلف المشاركون وغير المشاركين في التدخل من حيث الخصائص التي لا يتم رصدها، والتي تؤثر على كل من احتمالية مشاركة الأفراد في التدخل (و/أو استكماله) والنتائج ذات الاهتمام.

من المرجح أن يختلف المتقدمون للمشاركة في معظم برامج عمالة الشباب اختلافاً كبيراً عن أولئك الذين لا يتقدمون، وأن هذه الاختلافات لا يمكن ملاحظتها بسهولة. فعلى سبيل المثال، قد يكون المشاركون في مشروع الإرشاد الوظيفي أكثر تحفزاً ويستطيعون الوصول إلى معلومات أفضل حول كيفية العثور على وظيفة مقارنة بغير المشاركين، حتى قبل بدء التدخل. وفي هذه الحالة، قد يكون المشاركون أكثر نجاحاً من حيث نتائج سوق العمل ولكن سيكون من غير الواضح إن كان سبب ذلك يعود إلى التدخل أو الميزة الأولية في ظروف البداية.

تتضمن أحد الأهداف الرئيسية لتقنيات التقييم المعروضة هنا القضاء على التحيّز في الاختيار، مما يمكّن من إسناد الاختلافات التي يتم ملاحظتها في النتائج بين مجموعات المعالجة والمقارنة إلى التدخل.

لا يمكن ملاحظة التحيّز في الاختيار شأنه في ذلك تماماً شأن الوضع المغاير للواقع. وما لم يتم الانتباه وأخذ الحيطة والحذر عند اختيار مجموعات المعالجة والمقارنة بشكل مناسب، سوف تشتمل مقارنة بسيطة لنتائج سوق العمل للمجموعات المعالجة والمقارنة على كل من أثر التدخل والتحيّز في الاختيار. ومن المستحيل عموماً معرفة المدى الدقيق الذي يمكن به إسناد أي عنصر إلى أي سبب معين أو آخر.

تحاول تقنيات التقييم الواردة أدناه اختيار مجموعات المعالجة والمقارنة للقضاء على التحيّز في الاختيار، بحيث لا تعكس المقارنات بين المجموعات المعالجة والمقارنة سوى أثر التدخل.

تعتبر مجموعة المقارنة الجيدة ركيزة أساسية تضمن **الصلاحية الداخلية** للتقييم التي تحدد موثوقية ومصداقية نتائج التقييم. تتفاعل الصحة الخارجية عندما نبدأ في التفكير في قابلية نقل هذه النتائج: هل ينبغي توسيع نطاق التدخل ليشمل مجتمعات أخرى، أم تنفيذه على مستوى إقليمي أو وطني؟ هل يمكننا توقع نتائج مماثلة إذا قمنا بتصميم هذا البرنامج في سياقات أخرى و/أو لشريحة مختلفة من السكان المستهدفين؟ وعادة ما تحظى هذه الأسئلة بالاهتمام الكبير لصناع السياسات.

يجب أن نضع في الاعتبار أن الظروف لن تكون متطابقة أبداً عند تكرار أو توسيع نطاق التدخل. وبالتالي، ولتحقيق **الصلاحية الخارجية**، لا بد من

تعريف

الصلاحية الخارجية: تعني الصلاحية الخارجية في تقييم الأثر أن الأثر السببي الملحوظ يمكن تعميمه على جميع الأفراد المؤهلين. لذلك، من الضروري أن تكون عينة التقييم عينة تمثيلية لجميع الأفراد المؤهلين حتى يكون التقييم ذو صلاحية خارجية.

تعريف

التحيّز في الاختيار: يحدث التحيّز في الاختيار عندما ترتبط أسباب مشاركة الفرد في البرنامج بالنتائج. يحدث هذا التحيّز غالباً عندما تختار مجموعة المقارنة بنفسها الانسحاب من البرنامج (على سبيل المثال، المتسربين).

تعريف

الصلاحية الداخلية: لتحقيق الصلاحية الداخلية، يجب أن يشتمل تقييم الأثر على مجموعة مقارنة تقدم تقديراً صحيحاً للوضع المغاير للواقع. ويمكن لتقييم الأثر الصحيح داخلياً أن يسند بوضوح التغيرات في النتائج إلى التدخل عن طريق التحكم عن الاختلافات المحتملة بين مجموعة المعالجة والمقارنة. ويمكن تحقيق ذلك من خلال تطبيق التقنيات التجريبية أو شبه التجريبية بشكل مناسب.

فهم الجوانب المعقدة المحيطة بالبرنامج في وقت ومكان وسياق تنفيذه، والأثر المحتمل على نتائج التقييم. فعلى سبيل المثال، قد تكون خدمات التوظيف لشباب الخريجين فعّالة في المنطقة بسبب وجود طلب مقابل من الاقتصاد المحلي على هذه المجموعة المستهدفة، والسمعة الحسنة للمؤسسات التعليمية، وتنفيذ البرنامج في موسم ارتفاع الطلب من أصحاب العمل على العمل. وقد يظهر تنفيذ نفس التدخل في مناطق أخرى وطوال العام، نتائج مختلفة للغاية.

من الضروري تحليل الآليات السببية التي تستند إليها النتائج المرصودة من أجل فهم، ليس فقط هل ينجح التدخل، بل ولماذا وفي أي سياق يمكن توقع نجاحه. تعتبر الأساليب النوعية، مثل تلك المطبقة في سياق التقييم القائم على النظرية، ذات أهمية أساسية في هذا العمل.

في سياق هذه الملاحظة، سيتم تقديم طرق مختلفة لتقييم الأثر الكمي، يتبعها مثال عن طريقة نوعية وملاحظات تتناول كيف يمكن لنهج الطرق المختلطة أن يساعد في تحقيق كل من نتائج تقييم الصلاحية الداخلية والخارجية.

المربع 5-1: دعم منظمة العمل الدولية لتقييم الأثر

قامت إدارة التقييم التابعة لمنظمة العمل الدولية (المعروفة أيضاً باسم "EVAL") بتطوير مجموعة متنوعة من الموارد لدعم تقييم الأثر (IE):

- إطار تقييم الأثر: وضعت إدارة التقييم ورقة بيان الموقف حول كيف ومتى ولماذا يجب أن يتم النظر في تقييم الأثر وتنفيذه، بناءً على مدخلات من موظفي منظمة العمل الدولية. وتغطي هذه الورقة القضايا الرئيسية، مثل الاستخدام المحدد والغرض من تقييم الأثر؛ والمطابقة بين أسئلة بحث التقييم والمنهجية المناسبة؛ واستخدام مجموعة من المنهجيات التكميلية والمتاحة؛ وجدوى وقيمة تقييم الأثر؛ والحاجة ليس فقط إلى تحديد الأثر (ماذا) ولكن أيضا كيف ولماذا.
- مرفق مراجعة تقييم الأثر (*IERF*): وضعت إدارة التقييم آلية مراجعة تمكن موظفي منظمة العمل الدولية من طرح الأسئلة وطلب مراجعة أوراق المفاهيم والمقترحات الكاملة والخطط والتقارير من أجل المساعدة في تخطيط تقييم الأثر أو تصميمه أو تنفيذه (EVAL_impact@ilo.org). وتُتاح ملاحظة موجزة عن تشغيل هذا المرفق.
- مخزون تقييمات الأثر المنفذة في منظمة العمل الدولي: يسمح المخزون بسهولة الوصول إلى المعارف المؤسسية ضمن مجموعة متنوعة من مجالات التدخل.
- تقييم نوعية تقييمات الأثر في منظمة العمل الدولية: من أجل الرصد والإبلاغ عن التقدم الذي تحققه منظمة العمل الدولية في استخدام تقييم الأثر وجودته، سوف تقوم إدارة التقييم بإجراء تقييم دوري لجودة تقييمات الأثر في جميع أنحاء المنظمة.
- شبكة غير رسمية لتقييم الأثر كمجتمع للممارسة: تجتمع هذه المجموعة غير الرسمية من الزملاء المعنيين والمهتمين بتقييم الأثر على أساس منتظم لتبادل الخبرات وتقديم مراجعة النظراء بشأن تقييم الأثر، حسب الحاجة.

تهدف هذه الموارد إلى دعم منظمة العمل الدولية من أجل زيادة تعزيز قدرتها من حيث استخدام تقييم الأثر، وتوثيق المعرفة بعوامل النجاح ولمن، وتقييم الأثر.

التحديات المحددة لتقييم برامج سوق العمل النشطة التي تركز على الشباب[3]

تؤثر طبيعة برامج سوق العمل النشطة (ALMPs)، وتحديداً تلك التي تركز على استهداف شريحة الشباب، على جوانب متعددة من تصميم التقييم الصحيح. وكخلفية لإجراء مناقشة أكثر تفصيلاً لقضايا التصميم في الملاحظة 6، يقدم هذا القسم لمحة عامة عن بعض السمات الأكثر شيوعاً لبرامج سوق العمل النشطة التي تركز على الشباب ويصف سمات معينة لتصميم التقييم المرتبطة تحديداً بهذه الأنواع من البرامج. إن تحديد الميزات التي يحتمل أن تكون موجودةً في أي بيئة معينة سيساعد في وضع تصميم التقييم المناسب.

برامج إلزامية أو اختيارية

من الخصائص الأساسية للتدخل الذي يستهدف الشباب هو تحديد إن كان البرنامج إلزامياً أم اختيارياً. وتُدمج البرامج الإلزامية في العديد من خدمات التوظيف العامة، بما في ذلك المتعلقة بتأمينات البطالة وبرامج التدريب. وفي هذه السياقات، يُطلب من الشباب المشاركة في برنامج سوق العمل النشط المرتبط بإعانات البطالة. ومع ذلك، وكما سيتم بيانه في الأقسام التالية، تخلق المشاركة الإلزامية في أي تدخل يستهدف عمالة الشباب تحديات أمام تقييم الأثر، حيث تحتاج المقاييس الصحيحة للأثر إلى مجموعة معالجة ومجموعة مقارنة مكافئة. ولا يمكن تطبيق بعض أساليب تقييم الأثر إلا على البرامج الاختيارية التي تستقطب المشاركين من مجموعة واسعة من مقدمي الطلبات الذين لهم حرية قرار المشاركة.

عدم الامتثال: عدم الحضور والتسرب

في كثير من التدخلات الاختيارية التي تستهدف عمالة الشباب، يخفق جزء كبير من الأشخاص المسندين إلى البرنامج في التسجيل في البرنامج (ما يسمى عدم الحضور) أو ينسحبون قبل الانتهاء من البرنامج (التسرب). ويرتبط هذا التحدي بوجه خاص بالشباب الذين يتمتعون بدرجة عالية من التنقل، ويميلون إلى تغيير العنوان ومكان العمل بشكل متكرر والتناوب بين العمل والدراسة.

يقول ديفيد كارد وآخرون (2011) ما يلي:

> من النادر تحقيق معدلات إتمام البرنامج بأكثر من 80 في المائة، ويشيع انخفاض المعدلات إلى 50 في المائة، ويعد الفشل في توقع المشكلات الناتج عن حالات عدم الحضور والتسرّب أحد الأسباب الرئيسية لضعف تصميم تقييمات برامج سوق العمل النشطة (2011، ص 13).

بينما لا يؤدي عدم امتثال أعضاء ينتمون لمجموعة البرنامج أو مجموعة المقارنة إلى إبطال تصميم التقييم في حد ذاته، إلا أنه يعقّد من تفسير النتائج، ويفيد حاجة التقييم إلى جمع بيانات حول معدلات المشاركة الفعلية في البرنامج من قبل مجموعة البرنامج ومجموعة المقارنة.

تعتمد صلاحية التصميم العشوائي اعتماداً رئيسياً على التكافؤ بين النتائج المرصودة من مجموعة المقارنة والنتائج المغايرة للواقع من مجموعة المعالجة. وفي معظم الحالات، يختلّ هذا التكافؤ

[3] هذا القسم مقتبس من كارد وآخرون (2011).

عند فقدان أو تسرب أي أعضاء من أي مجموعة محددة أو مجموعة أخرى. ولهذا السبب، يجب أن يستند تحليل التصميم العشوائي إلى مقارنة بين مجموعات المعالجة والمقارنة حسبما تم تحديدهما منذ البداية، باستخدام بيانات عن كل شخص تم تخصيصه في البداية لهذه المجموعات. ويعرف ذلك في المؤلفات المعنية بالتقييم التجريبي باسم "نية معالجة التحليل".

التعيين والفرز

نظراً لأنه لا يتم تخصيص سوى مجموعة محددة من جميع الشباب المسجلين في تقييم الأثر للحصول على البرنامج بالفعل، قد تؤدي نسبة الاستيعاب في التقييم إلى تعطيل التدفق الطبيعي للعملاء في أي برنامج مستمر[4]. ولا يعد ذلك مصدر قلق خاص في بيئة يزيد فيها عدد المتقدمين عن الأماكن المتاحة في البرنامج: وعندئذ، يعمل الاختيار العشوائي كأداة تقنين ملائمة وموضوعية. إلا أنه عند الحاجة إلى تدفق منتظم من المرشحين لشغل الأماكن المتاحة في البرنامج، قد يعترض مشغلو البرنامج على تخصيص بعض من عملائهم المحتملين إلى مجموعة المقارنة وقد يحاولون تجاوز عملية التخصيص. ويجب التعرف مسبقاً على احتمالية حدوث ذلك. فإذا كان الأمر كذلك، قد يتضمن تخطيط التقييم تخصيص ميزانية لجهود التعيينات الإضافية من أجل زيادة تدفق العملاء الجدد، وكذلك موارد إضافية من أجل رصد الامتثال لقواعد التعيين عن كثب. فعلى سبيل المثال، قد تقتصر برامج سوق العمل النشطة التي تستهدف الشباب على الرجال والنساء العاطلين عن العمل الذين تتراوح أعمارهم بين 16 و30 سنة. وعادة، يجب استخدام نفس إجراءات وقواعد الفرز الخاصة بالأهلية لاختيار المشاركين للتقييم.

أحجام العينات

تستند المبادئ التوجيهية لتحديد أحجام العينات الضرورية لتقييم برنامج سوق العمل النشطة إلى حساب القدرة القياسية. يتمثل المكون الرئيسي لهذا الحساب في تقدير حجم الأثر المعقول للبرنامج (على سبيل المثال، تأثير البرنامج على نتيجة الاهتمام، معبراً عنه كجزء من الانحراف المعياري لهذه النتيجة). وفي ضوء هذه القيمة والاختيارات المعيارية لمستوى الدلالة الإحصائية (على سبيل المثال، 5 في المائة) ومدى ملاءمة قوة التصميم (على سبيل المثال، 0.80)، يكون من السهل حساب أحجام العينات المناسبة لمجموعات المعالجة والمقارنة بتصميم عشوائي لمجموعات متساوية الحجم. وقد طوّر كل من كارد وآخرين (2011) إرشادات (كما هو موضح في الجدول 5-1) توضح تفاصيل حجم العينة المطلوبة لقياس نطاق التأثيرات. يعرض كل صف معدل العمالة بين صفوف مجموعة المقارنة، ويمثل كل عمود الاختلاف بين مجموعات المعالجة والتحكم. فعلى سبيل المثال، إذا كان معدل العمالة بنسبة 50٪ في المائة في مجموعة المقارنة، يكون حجم العينة المطلوبة 6354 مشارك ونفس العدد لغير المشاركين للكشف عن تأثير كبير من 2.5 نقطة مئوية في العمالة.

[4] على سبيل المثال، إذا حضر 100 عميل جديد إلى مواقع البرامج كل شهر، وكان هناك 80 مكاناً مفتوحاً للتسجيل في البرنامج كل شهر، فيمكن، على الأكثر، تسجيل 40 شخصاً شهرياً في التقييم: سيتم تخصيص 20 شخصاً للبرنامج (جنباً إلى جنب مع 60 عميلاً جديداً آخرين ممن ليسوا جزءاً من التقييم) و20 شخصاً لمجموعة المقارنة.

الجدول 5-1: حجم العينة المطلوبة للكشف عن الآثار المهمة

		أثر البرنامج					
		2.5%	5.0%	7.5%	10.0%	12.5%	15.0%
معدل العمالة بين صفوف مجموعة التحكم	30%	5475	1417	650	376	247	176
	35%	5883	1511	688	396	259	183
	40%	6166	1574	713	408	265	186
	45%	6323	1605	723	412	266	186
	50%	6354	1605	719	408	262	183
	55%	6260	1574	702	396	254	176
	60%	6040	1511	671	376	240	165
	65%	5695	1417	625	349	221	151

المصدر: كارد وآخرون (2011)

* حسب افتراضات قياسية (السلطة = 0.8، الأهمية = 0.5، مجموعات متساوية الحجم)، باستخدام أمر sampsi في Stata

عند التفكير في حجم الأثر ذي الاهتمام لبرنامج سوق العمل النشطة، يوصي كارد وآخرون و (2011) بوضع هذه البرامج في سياقها ويفيدون:

> لقد أظهرت مجموعة كبيرة جداً من الأبحاث بأن كل سنة إضافية من التعليم الرسمي في معظم البلدان حول العالم ترتبط بزيادة في الكسب تصل إلى حوالي 10 في المائة. ويمكن القول بأن البرنامج النموذجي لسوق العمل النشطة ينطوي على استثمار أصغر من الاستثمار في السنة النموذجية للتعليم المدرسي الرسمي. لذلك، يعتبر حجم التأثير الذي يقل عن 10 في المائة معقولاً، وبالنسبة للبرامج الأقل كثافة، قد تكون أحجام التأثيرات التي لا تزيد عن 5 في المائة مقبولة (2011، ص. 19).

توقيت استطلاعات المتابعة

يعد توقيت استطلاعات المتابعة قراراً مهماً من حيث ضمان آثار البرنامج، إذ تستخدم العديد من تقييمات برامج سوق العمل النشطة استبيان متابعة لمدة عام، ويرجع ذلك جزئياً إلى شروط عقد التقييم التي غالباً ما تتطلب صدور تقرير ختامي في غضون عامين أو ثلاثة أعوام. ومن ناحية أخرى، تشير المؤلفات الحالية حول برامج سوق العمل النشطة إلى أن تأثير البرامج المكثفة، مثل التدريب في الفصول الدراسية وبرامج التدريب على رأس العمل، لا يتجلى إلا بعد سنتين أو ثلاث سنوات من الدخول في البرنامج، وليس بعد عام واحد فقط (كارد وآخرون، 2011). وبناءً على هذه الدراسات، ومراعاة فترات الانقطاع في العديد من البرامج، من المستحسن تطويل أفق ما بعد البرنامج الذي لا تقل مدته عن عامين بالنسبة لبرامج سوق العمل النشطة.

ومع ذلك، توجد مفاضلة بين القدرة على رصد الآثار طويلة الأجل والتأكد من وجود تصميم صحيح لتقييم الأثر: وبينما يتمتع الشباب بدرجة عالية من التنقل وقد ينتقلون في جميع أنحاء البلد (أو ربما يهاجرون) بعد استكمال تعليمهم أو برامج التدريب، تتزايد صعوبة تعقب عددٍ كافٍ من المستفيدين من البرنامج مع تقدم الوقت. وبينما يتعذر الاتصال بعددٍ كبيرٍ من الشباب الذين شاركوا في المسح الأساسي لتقييم الأثر من أجل استطلاعات المتابعة (معدل استنزاف عالٍ)، يصبح من الصعب أو المستحيل بشكل متزايد الكشف عن الأثار وتحديد كميتها بشكل موثوق بسبب انخفاض القدرة الإحصائية والتحيّز المحتمل.

الأساليب الكمية لتقييم الأثر

تحقق الأساليب الكمية الموصى بها لتقييم الأثر الصلاحية الداخلية وتدرأ التحيّز في الاختيار، وذلك بمقارنة المجموعات مع المعالجة وبدونها، والتي لا تختلف على النحو الأمثل إلا من هذه الناحية. ويمكن تحقيق ذلك على أكمل وجه إذا كانت لدينا السيطرة على المستفيدين وغير المستفيدين من التدخل. وعندئذ، يمكن إجراء **تصاميم التقييم التجريبية**، علماً بأن التجارب الموجهة باستخدام العينات العشوائية هي أكثرها شيوعاً، فإن كان التخصيص إلى مجموعتي المعالجة **والمقارنة عشوائياً** تماماً، ستكون المجموعتان متشابهتين جداً، في المتوسط، قبل بداية البرنامج وسنكون قد قطعنا شوطاً طويلاً نحو ضمان الصلاحية الداخلية.

لا يكون التوزيع العشوائي ممكناً أو مستحسناً دائماً لعدة أسباب، موردها أدناه. وعندئذ، يمكن استخدام طرق أخرى تسعى إلى إجراء مقارنات صحيحة داخلياً عن طريق إنشاء وضع صحيح مغاير للواقع. ويطلق على هذه الأساليب اسم الطرق **شبة التجريبية لتقييم** الأثر، علماً بأن الأساليب الأكثر استخداماً هي الفرق في الاختلافات (DID)، ومطابقة درجة الميل (PSM) وتصميم انحدار الانقطاع (RDD)، والتي سيتم تناولها جميعاً باختصار في هذا القسم. ومن المستحسن عموماً دراسة جميع جهود تقييم الأثر الممكنة والموازنة بين محاسنها ومساوئها بعناية قبل الشروع في التقييم. يقدم الجدول 5-2 نظرة عامة على طرق تقييم الأثر الكمي المختلفة.

الشكل 5-2: مراعاة جميع أساليب التقييم الممكنة خلال مرحلة التخطيط

المصدر: www.freshspectrum.com

تعريف

التصميم التجريبي: تعتمد التصاميم التجريبية على بعض عناصر التوزيع العشوائي في تخصيص المشاركين إلى مجموعات معالجة ومقارنة، ويمكن أن تنتج تقديرات تأثير عالية المصداقية ولكنها غالباً ما تكون مكلفة وصعبة التنفيذ، لأغراض تدخلات معينة.

التجربة المنضبطة العشوائية هي دراسة يتم فيها توزيع الأشخاص عشوائياً (بالصدفة وحدها) لتلقي المعالجة، مثل المشاركة في تدخل محدد.

التصميم شبه التجريبي: تُستخدم نُهج التصميم شبه التجريبي لإنشاء مجموعة مقارنة صحيحة باستخدام وسائل إحصائية للتحكم في الاختلافات بين الأفراد المشمولين بالبرنامج الجاري تقييمه والأفراد غير المشمولين.

الجدول 5-2: نظرة عامة على طرق تقييم الأثر المختلفة

المنهجية	الوصف	من في مجموعة المقارنة؟	الافتراضات المطلوبة	البيانات المطلوبة
قبل وبعد	يقيس كيفية تحسّن (أو تغير) المشاركين في البرنامج بمرور الوقت	المشاركون في البرنامج أنفسهم - قبل المشاركة في البرنامج	كان البرنامج هو العامل الوحيد الذي يؤثر على أي تغييرات في النتائج المقاسة بمرور الوقت	بيانات قبل وبعد للمشاركين في البرنامج
مقارنة المشاركين بغير المشاركين	يقيس الفرق بين المشاركين في البرنامج وغير المشاركين بعد اكتمال البرنامج	الأفراد الذين لم يشاركوا في البرنامج (لأي سبب)، وتم جمع بياناتهم بعد انتهاء البرنامج	غير المشاركين المتطابقين مع المشاركين إلا من حيث المشاركة في البرنامج	بيانات ما بعد البرنامج للمشاركين وغير المشاركين
الفرق في الاختلافات	يقيس تحسّن (أو تغير) المشاركين في البرنامج بمرور الوقت مقارنةً بتحسن (تغير) غير المشاركين	الأفراد الذين لم يشاركوا في البرنامج (لأي سبب)، وتم جمع بياناتهم قبل وبعد البرنامج	إذا لم يتم تنفيذ البرنامج، فستكون للمجموعتين مساراً متماثلاً خلال هذه الفترة (سيشتركان في نفس الاتجاه الزمني "المشترك")	بيانات قبل وبعد لكل من المشاركين وغير المشاركين
مطابقة درجة الميل	يتم مطابقة الأفراد في مجموعة المعالجة مع غير المشاركين الذين لديهم خصائص مشابهة قابلة للرصد. إن متوسط الفرق في النتائج بين الأفراد المتطابقين هو التأثير المقدر	غير المشاركين الذين لديهم مجموعة من الخصائص التي تتنبأ باحتمال دخولهم في البرنامج كمشاركين	العوامل التي تم استبعادها (لأنها غير قابلة للرصد و/أو لم يتم قياسها) لا تسبب تحيّز في النتائج لأنها إما غير مرتبطة بالنتيجة أو لا تختلف بين المشاركين وغير المشاركين	النتائج وكذلك "المتغيرات للمطابقة" لكل من المشاركين وغير المشاركين
تصميم انحدار الانقطاع	يتم تصنيف الأفراد بناءً على معايير محددة وقابلة للقياس. هناك نقطة انقطاع لتحديد المؤهلين للمشاركة. يتم قياس التأثير بمقارنة نتائج المشاركين وغير المشاركين بالقرب من خط الانقطاع	الأفراد الذين يقتربون من خط الانقطاع، ولكنهم يقعون من على جانب هذا الخط الذي لا يؤهلهم (فقط) للبرنامج	بعد التحكم في المعايير (وغيرها من المقاييس الاختيارية)، فإن الاختلافات الباقية بين الأفراد الذين يقعون مباشرة تحت وفوق درجة الانقطاع ليست ذات دلالة إحصائية ولن تسبب تحيّزاً في النتائج. من المتطلبات الضرورية لهذا الأمر أن يتم الالتزام بمعايير الانقطاع بشكل صارم	النتائج وكذلك بيانات معايير التصنيف (مثل السن، والمؤشر، وما إلى ذلك). متغيرات الخلفية الاجتماعية والاقتصادية مستحسنة للغاية.
التقييم العشوائي	يتم تخصيص عينة من الأفراد المؤهلين عشوائياً في مجموعتين: أولئك الذين يستفيدون من التدخل وأولئك الذين لا يستفيدون منه. إن الأثر هو الفرق في النتائج بين المجموعتين. وهناك طرق مختلفة لتنفيذ التوزيع العشوائي	يتم تخصيص المشاركين عشوائياً إلى مجموعات المعالجة والمقارنة	نجاح التوزيع العشوائي والالتزام به؛ بمعنى أن المجموعتين متطابقتان إحصائياً (من حيث العوامل المرصودة وغير المرصودة)	بيانات النتيجة لمجموعات المعالجة والمقارنة. البيانات الأساسية ومتغيرات الخلفية مستحسنة.

التوزيع العشوائي – التصميم العشوائي

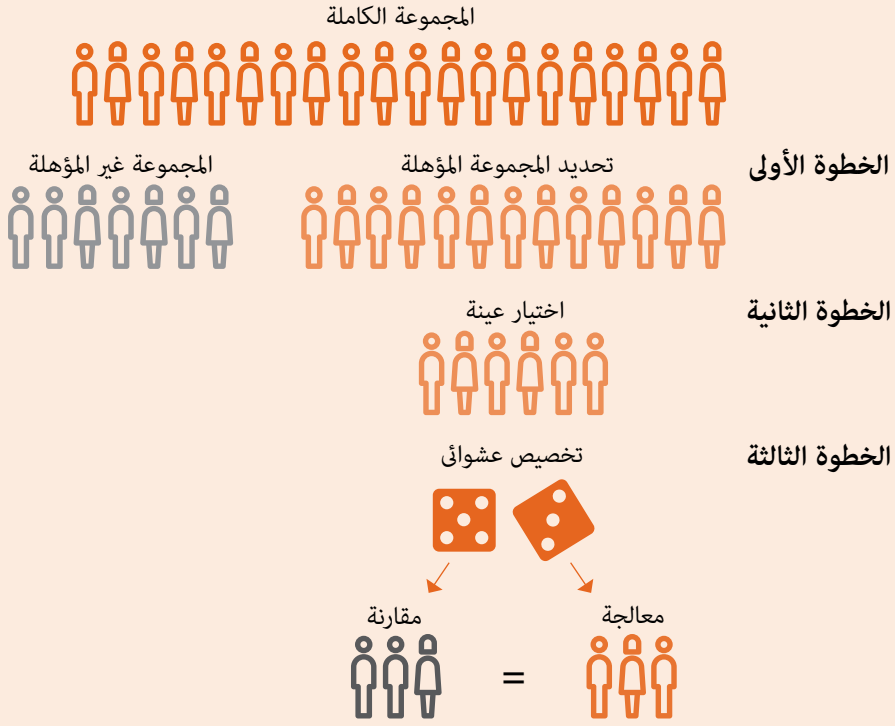
التصميم العشوائي هو وسيلة بسيطة وشفافة لتخصيص الشباب إلى المجموعات التي ستتفيد من خدماتنا (مجموعة المعالجة) وتلك التي لن تستفيد منها (مجموعة المقارنة)، وهي الطريقة المستخدمة لتصميم تجارب عشوائية منضبطة. إذا تم تخصيص عينة كبيرة بالقدر الكافي من الأشخاص من نفس الفئة السكانية بشكل عشوائي إلى إحدى المجموعتين، فستكون للمجموعتين خصائص مشابهة قابلة للرصد في المتوسط (العمر والجنس والطول ومستوى التعليم، وما إلى ذلك). وبنفس القدر من الأهمية، سوف يتقاسمون، في المتوسط، نفس الخصائص التي لا يمكن رصدها (مثل الدوافع والحالة الذهنية).

من خلال التوزيع العشوائي، يمكن أن يعزى الفرق في النتائج التي نلاحظها بين المجموعتين في نهاية تدخلنا إلى التدخل، لأن جميع العوامل الأخرى التي يمكن أن تؤثر على النتائج متساوية بشكل عام.

كيف يعمل

هناك ثلاث خطوات للتصميم العشوائي (انظر الشكل 5-3).

الشكل 5-3: خطوات التصميم العشوائي



المصدر: بيان ذاتي.

الخطوة 1: تحديد السكان المؤهلين

تتمثل الخطوة الأولى في أي تجربة عشوائية منضبطة في العثور على مجموعة من الشباب المؤهلين للتدخل. إذا كان عالم طبي يدرس تأثير عقار معين على مرض بعينه في مرحلة الطفولة، فسيبحث عن مجموعة محددة من الأطفال ولن يسجل البالغين أو المسنين في التدخل. وبالمثل، قد يستهدف برنامج التدريب على رأس العمل الشباب في المناطق الحضرية من فئة عمرية محددة، وبالتالي لن يشمل البالغين أو شباب الريف. والمهم هنا وجود معايير واضحة وشفافة (العمر والجنس ومستوى الدخل والحالة الوظيفية وما إلى ذلك) والقدرة على الاتصال بمن سيكون مؤهلاً للانضمام إلى التدخل وبمن لن يستطيع ذلك.

الخطوة 2: انتقاء عينة للتقييم

لتقييم التدخل، لا نحتاج إلى اختبار كل من سيشارك في التدخل، بل نحتاج فقط لاختيار مجموعة تمثيلية من الناس بأعداد كافية لأغراض تقييمنا؛ وهي ما نطلق عليه اسم عينتنا، والتي ستكون من الشباب الذين سنجمع البيانات عنهم. بينما تقدم الملاحظة 6 المزيد من التفاصيل حول كيفية تحديد **العينة** وحجمها، يتراوح حجم العينة النموذجية للتدخل الذي يستهدف عمالة الشباب المراد تقييمه من خلال التصميم العشوائي ما بين 500 و2000 مشارك في الدراسة (مع تقسيم متساو تقريباً في العادة على مجموعتي المعالجة والمقارنة).

يمكن اختيار العينة للتقييم بطريقتين، ويتحدد هذا بحسب إن كان التدخل كبيراً أو صغيراً. قد يتبين في التدخل الصغير أن هناك 10000 مستفيد مؤهل، مثل الشباب في المناطق الحضرية الذين تتراوح أعمارهم بين 16-24 سنة. وقد يكون للتدخل ميزانية كافية لمساعدة 500 منهم. ففي الأحوال المثلى، يكون حجم مجموعة المقارنة مماثلاً لمجموعة المعالجة، لذلك يتعين اختيار 1000 من 10000 شاب للتدخل والتقييم (انظر الشكل 4-5، الصورة اليسرى).

وقد تكون البرامج الكبيرة أكبر من حجم العينة المطلوب للتقييم. إذا كان التدريب الوظيفي قادراً على خدمة 4.000 شاب، فليس من الضروري إيجاد 4.000 شاب إضافي للمقارنة. بدلاً من ذلك، قد تكون هناك حاجة إلى 1.000 شاب فقط. يمكن للتدخل عندئذ تحديد عينة من 5.000 شاب من إجمالي عدد المشاركين البالغ عددهم 10.000 شاب. ومن بينهم، يمكن ضمان انضمام 3.000 شاب إلى التدخل. بعدئذ، يتم تقسيم الشباب الباقين وعددهم 2.000 شاب عشوائياً بين التدخل ومجموعة المقارنة (الشكل 4-5، الصورة اليمنى).

من أجل اختيار عينة تمثيلية من مجموع السكان المؤهلين البالغ عددهم 10.000 شاب، ينبغي انتقاء عينة (سواء كان عددها 1.000 شاب كما في الحالة الأولى أو 5.000 كما في الحالة الثانية) عشوائياً من السكان المؤهلين. فمن خلال الاختيار العشوائي، سيكون لدى المشاركين، في المتوسط، خصائص متشابهة مع مجموع السكان المؤهلين. وعلى الرغم من أننا لا ندرج سوى عدد محدود من الشباب في الدراسة، فإن الأثر المحتمل للتدخل يمكن استقراءه لتغطية جميع السكان المؤهلين، في هذه الحالة، 10.000 شاب.

تعريف

العينة هي مجموعة فرعية من السكان. ونظراً لأنه من المستحيل أو غير العملي عادةً جمع معلومات حول جميع السكان موضوع الاهتمام، يمكننا بدلاً من ذلك جمع المعلومات حول مجموعة فرعية ذات حجم يمكن التحكم فيه. إذا تم اختيار المجموعة الفرعية بشكل جيد، فمن الممكن استنتاج النتائج عن جميع السكان.

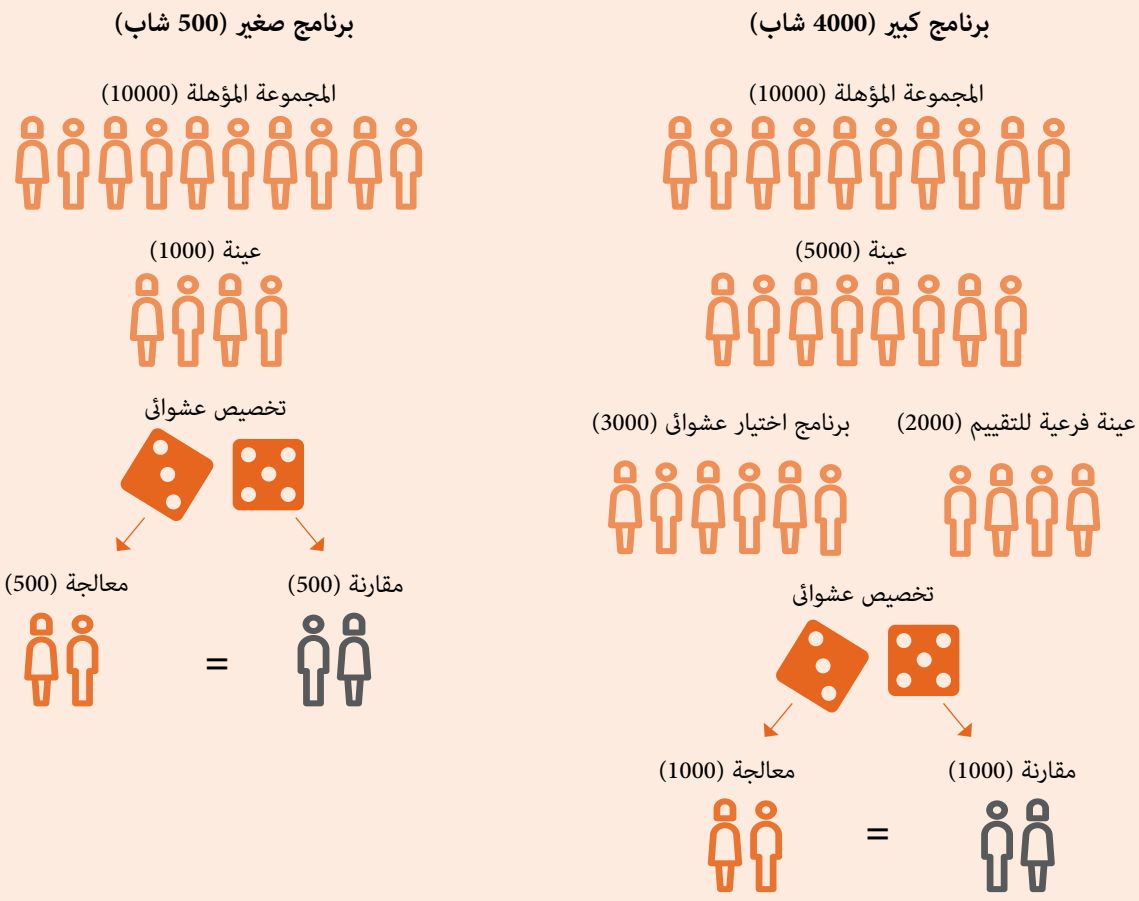
الخطوة 3: عشوائية التوزيع

تتمثل الخطوة التالية في تعيين عينة مختارة من الشباب إلى مجموعتي المعالجة والمقارنة المتساويتين في الحجم تقريباً. ففي التجارب العشوائية المنضبطة، يكون لكل شاب فرصة متساوية في الاستفادة من التدخل. ويمكن إجراء التوزيع العشوائي من خلال التقنيات التقليدية، مثل القرعة عن طريق رمي العملة أو النرد أو سحب الأسماء من القبعة، ويمكن إجراؤه بشكل علني عند الرغبة إن كانت العينة صغيرة نسبياً (على سبيل المثال، لن يكون سحب 2000 اسم من القبعة عملياً). وبدلاً من ذلك، إذا كان عدد الأشخاص كبيراً، يفضل الاختيار عشوائياً باستخدام برامج الكمبيوتر، مثل مايكروسوفت إكسل. ويمكن

معلومة استرشادية

تتمثل إحدى طرق الحصول على عينة عشوائية من الشباب في إعداد قائمة بمجموع السكان من الشباب من التعداد السكاني أو سجلات تسجيل الناخبين أو بعض قواعد البيانات الأخرى، ثم الاختيار عشوائياً من تلك القائمة. إذا لم يكن هذا النهج ممكناً، سيؤدي استهداف المناطق حيث يتفاعل الشباب بشكل عشوائي، مثل المراكز الحضرية، إلى إعداد عينة عشوائية. إذا كان من المعلوم أن الشباب يقضون وقتاً في 50 مركزاً مختلفاً حول المدينة أو البلد، فمن المرجح أن يؤدي اختيار المراكز بشكل عشوائي ثم اختيار جزء من الشباب في هذه المراكز للمشاركة في الدراسة إلى اختيار مجموعة من الشباب مع القليل من التحيّز. ستناقش الملاحظة 6 أخذ العينات بمزيد من التفصيل.

الشكل 5-4: اختيار عينات للبرامج الصغرى والكبرى



إجراء التوزيع العشوائي على عدة مستويات. يعمل تخصيص العينة إلى مجموعتي المعالجة أو المقارنة بشكل عشوائي على انتقاء المشاركين بشكل عادل، وتطوير وضع جيد مغاير للواقع: إذا كان حجم العينة كبيراً بما فيه الكفاية، سيكون لدى الشباب في مجموعة المعالجة، في المتوسط، نفس الخصائص القابلة للرصد وغير القابلة للرصد مثل تلك التي لدى مجموعة المقارنة.

متى يمكن استخدام التصميم العشوائي؟

يمكن استخدام التقييم العشوائي عندما يتم تخطيط التقييم قبل التنفيذ (المحتمل)، وعندما لا يمكن للتدخل سوى خدمة جزء صغير من الشباب المؤهلين. وطالما لا تتوفر الموارد اللازمة التي تمكّن التدخل من خدمة كامل السكان المؤهلين، فلا توجد مخاوف أخلاقية في وجود مجموعة مقارنة، لأنه سيتم استبعاد مجموعة فرعية من السكان بالضرورة من التدخل. وعندئذ، يمكن الحفاظ على مجموعات المقارنة لقياس أثر التدخل على المدى القصير والمتوسط والطويل (جيرتلر وآخرون، 2016). الأهم من ذلك، لن يتم الحفاظ على الميزة المركزية للتوزيع العشوائي – وهي تشارك مجموعتي المعالجة والمقارنة، في المتوسط، في نفس الخصائص – إلا إذا استطعنا المتابعة مع جميع أعضاء مجموعتي المعالجة والمقارنة (تقريباً). تشكل معدلات الاستنزاف العالية تهديداً خطيراً على الصلاحية الداخلية لنتائجنا لكل طريقة لتقييم الأثر، والأساليب التي تستخدم تقنيات التوزيع العشوائي دون استثناء.

المزايا

- التصميم العشوائي هو الأسلوب الأكثر فعّالية لتطوير الوضع المغاير للواقع لأنه يؤدي إلى تشكيل مجموعة مقارنة متطابقة بشكل جيد (بالاعتماد على افتراضات أقل من الأساليب الأخرى). لذلك يعتبر التصميم الأكثر مصداقية لقياس الأثر.
- يعد أبسط الطرق التقييمية من حيث التحليل، إذ أن أثر التدخل في تجربة عشوائية هو ببساطة الفرق في النتائج بين مجموعتي المعالجة والمقارنة.
- يتيح للمجتمعات المشاركة بشكل مباشر في عملية الاختيار من أجل توزيع المنافع بشكل عادل وشفاف.
- سهولة التنفيذ وإبلاغ موظفي البرنامج.

العيوب

- يمكن أن ينطوي إجراء تجربة عشوائية على تكلفة كبيرة للغاية من حيث الموارد والوقت.
- لا يمكن تطبيق أي تنفيذ لاحق لهذه الطريقة. يجب أن يكون تخطيط التقييم جزءاً من تخطيط التدخل (وهو ممارسة جيدة على أي حال، لكنه لا يمثل دائماً الواقع في عمل المشروع).
- يتطلب استبعاد مجموعة مقارنة من التدخل طوال مدة تقييم الأثر. قد تنشأ مخاوف سياسية و/أو أخلاقية على الرغم من معيار التوزيع العشوائي بشكل شفاف (انظر المزيد في قسم "تكييف التصميمات العشوائية للسياقات المختلفة" أدناه).
- يجب على المنظمات التأكد من موافقة الشركاء وأصحاب المصلحة المحليين على هذه الطريقة.
- تعتمد الصلاحية الداخلية للتصميم العشوائي على نجاح التوزيع العشوائي والمحافظة عليه طوال فترة الدراسة، وهو الأمر الذي قد لا يمكن تحقيقه بسهولة. وقد يتقوّض هذا

الشرط في حالة تنفيذ التوزيع العشوائي بشكل غير صحيح، أو إذا لم تلتزم مجموعتا المعالجة أو المقارنة بالوضع المعين لهما (بمعنى إذا لم يشارك أفراد المعالجة من التدخل أو استفاد أفراد المقارنة من البرنامج)، أو إذا تسرب المشاركون من الدراسة قبل الانتهاء أو إذا كانت هناك **آثار غير مباشرة**: فمثلاً، قد يؤدي قيام الأشخاص الذين تلقوا التدريب على الوظيفة بنقل المهارات والمعارف المكتسبة إلى أقرانهم إلى التشويش على الفصل الواضح بين مجموعتي المعالجة والمقارنة. وتمثل هذه الحالات مشكلة كبيرة، لأنها قد تجعل النتائج متحيّزة بشكل كبير وبالتالي تهدد الصلاحية الإجمالية للتقييم.

تعريف

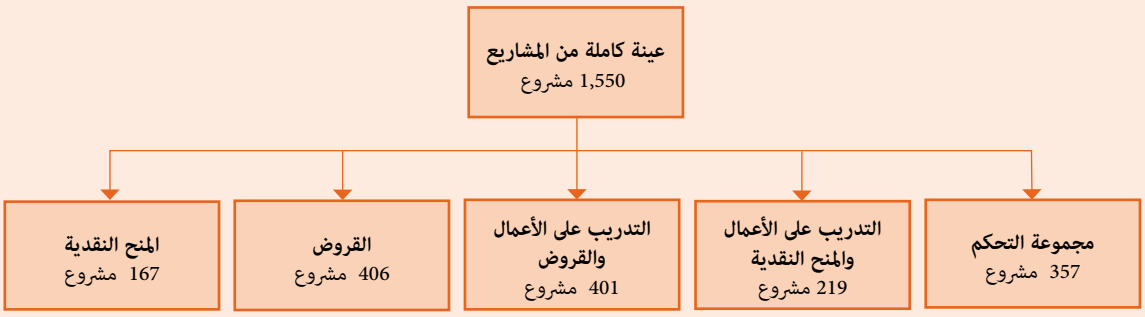
آثار غير مباشرة: الآثار الجانبية هي آثار التدخل على غير المشاركين؛ على سبيل المثال، في حالة انتشار المعارف المكتسبة من التدريب على المهارات داخل إحدى القرى، حتى بين صفوف أولئك الذين لم يحضروا الدورة التدريبية.

المربع 5-2: تقييم تدخل منظمة العمل الدولية ابدأ وحسّن مشروعك (SIYB)

لقد تم تصميم التقييم لاختبار إن كان توسيع فرص الحصول على رأس المال عن طريق المنح أو القروض سيؤدي إلى زيادة أرباح المشاريع متناهية الصغر المملوكة للرجال أو النساء، وإن كان يمكن لبرنامج التدريب على ريادة الأعمال التابع لمنظمة العمل الدولية "ابدأ وحسّن مشروعك" سيعزز تلك الآثار.

قام فريق البحث بمسح 4.637 مؤسسة متناهية الصغر من إحصاء الأعمال وقاموا باختيار 1550 من أصحاب المشاريع وإدراجهم في عينة التقييم بناءً، من بين معايير أخرى، على التعبير عن الاهتمام بتلقي تدريب منظمة العمل الدولية والمشاركة في تدخل القرض. وشملت العينة أصحاب المشاريع الصغيرة المهتمين بتحسين أعمالهم (على سبيل المثال، صالونات تصفيف الشعر، ومحلات البيع بالتجزئة، والخياطون). تم تقسيم العينة عشوائياً إلى خمسة مجموعات معالجة، والتي حصلت على التدخلات التالية: (1) قروض؛ و(2) منح نقدية و(3) تدريب على الأعمال وقروض؛ (4) تدريب على الأعمال ومنح نقدية؛ و (5) لا تدخل (مجموعة المقارنة) (الشكل 5-5).

الشكل 5-5: تصميم التقييم



فيما يلي بيان التدخلات الرئيسية التي تم تقديمها لأصحاب المشاريع:

- التدريب على الأعمال: يستهدف البرنامج التدريبي ابدأ وحسّن مشروعك (SYB) رواد الأعمال المبتدئين (أو الناشئين) ويتكون من دورة تدريبية مدتها خمسة أيام، يليها نشاط ميداني وجلسات مشورة فردية وجماعية. ويقوم المتدربون بإعداد خطة عمل ودراسة جدوى قابلة للتمويل تفصيلية (انظر www.ilo.org/siyb).
- تم تقديم منح نقدية غير مشروطة بقيمة 200 دولار أمريكي عبر حسابات مصرفية مجانية في مؤسسة محلية للتمويل الأصغر (MFI). ومنح أصحاب المشاريع حرية الاختيار في استخدام القرض.
- تم عرض قروض شبه مشروطة، تتراوح قيمتها بين 180 و220 دولاراً، بسعر فائدة سنوي مخفض نسبته 20 في المائة من جانب مؤسسة التمويل الأصغر، على أن يتم سداد القروض إلى مؤسسة التمويل الأصغر، ولكن بدون أي تبعات في حالة إساءة استخدام المال.

يساوي حجم المنح والقروض حوالي 1.5 مرة من الأرباح الشهرية للمشاريع المتوسطة.

للتحقق من مدى "نجاح" التوزيع العشوائي"، عمل التقييم على مقارنة أصحاب المشاريع في مجموعة المعالجة بأولئك الموجودين في مجموعة المقارنة بشأن 26 متغيراً مختلفاً وتبين أن المشاريع المعالجة وغير المعالجة كانت متطابقة، في المتوسط، قبل التدخل بالنسبة لكل المميزة تقريباً.

تم مسح أصحاب الأعمال قبل التدخل (مسح خط الأساس) وبعد ستة أشهر وتسعة أشهر وسنتين بعد التدخل (ثلاثة استطلاعات متابعة).

كان الربح التجاري هو المتغير الرئيسي للنتائج محل الاهتمام، وقد كشف التقييم عن زيادة كبيرة في الأرباح بعد التدخل، ولكن فقط بالنسبة لأصحاب المشاريع الذكور. لم تؤد أي من التدخلات إلى زيادات مستدامة في الأرباح لأصحاب المشاريع من الإناث. كما شهدت النساء ذوات الأرباح المبدئية المرتفعة تأثيرات سلبية خلال جميع التدخلات. وبينما كانت الاستجابة الأولية للمنح إيجابية، إلا أن هذه الزيادة اختفت تماماً حتى أنها أصبحت سلبية بمرور الوقت. فالنساء اللاتي حصلن على المنحة حققن ربحا أقل بنسبة 35 في المائة من أقرانهن اللواتي لم تحصلن على أي تدخل. وبعد تسعة أشهر، كانت النساء إما لسن أفضل حالاً أو أسوأ حالاً من نظرائهن في مجموعة المقارنة.

وجد التقييم أيضا أن القرب من أفراد الأسرة يمثل في العادة قوة إيجابية على الأعمال التجارية للرجال وسلبية للنساء. وقد تعرضت النساء المتزوجات واللواتي يعيشن مع أسرهن في نفس المقاطعة انخفاضاً كبيرا في الأرباح.

المصدر: فيالا، 2015.

تكييف التصميمات العشوائية مع السياقات المختلفة

يتردد بعض أصحاب البرامج في تعيين المستفيدين المحتملين عشوائياً في مجموعتي المعالجة والمقارنة، ويتمثل الشاغل العام في أن التقييم يؤدي إلى حجب الفوائد الواضحة ظاهرياً (مثل فرص التدريب) عن الأفراد المحتاجين، وهو أمر غير أخلاقي. ومع ذلك، يتجاوز الطلب بشكل كبير ما يمكن توفيره بالنسبة للعديد من التدخلات، وكما هو موضح بالتفصيل في المربع 5-3، قد يكون التوزيع العشوائي في الواقع أكثر أخلاقية من طرق الاختيار الأخرى.

المربع 5-3: هل التوزيع العشوائي أخلاقي؟

يعدّ تعيين المستفيدين المحتملين في مجموعتي المعالجة والمقارنة بشكل عشوائي غير أخلاقي في بعض الأحيان. وقد تكون هذه المخاوف صحيحة في حالات معينة، على سبيل المثال عندما يمكن توسيع نطاق السياسة أو التدخل التي من المرجح أن تنجح أو التي أثبتت نجاحها بتكلفة منخفضة إلى عدد كبير من السكان. ومع ذلك، تنشأ إحدى الحالات التالية في أكثر الأحيان:

- **عدم اليقين من تأثير المشروع.** لن يكون من الواضح بالنسبة لمعظم البرامج إن كان للتدخل أثر إيجابي وكبير على الفرد والمجتمع بما يسوّغ الموارد التي يتم إنفاقها. فعلى سبيل المثال، قد تؤدي البرامج الموجهة نحو الفتيات مع استبعاد الأولاد إلى زيادة العنف ضد المرأة. وقد يؤدي التدخل الذي يستهدف التمويل الأصغر للشباب إلى تفاقم أوضاع المشاركين إن لم يتمكنوا من سداد قروضهم. وقد يقلل برنامج التدريب المصمم بشكل سيئ من فرص العمل. وقد تكون تكلفة الزيادة في الدخل (على سبيل المثال، 100 دولار أمريكي لكل مشارك) مرتفعة جداً (على سبيل المثال، 1000 دولار أمريكي للشخص الواحد). ومن ثم، يفضل بشكل كبير تقييم التدخل على أساس مجموعتي المعالجة والمقارنة بالنسبة للتدخلات التي لم تثبت بعد كفاية تأثيرها وهيكليتها من حيث التكلفة.
- **قيود الميزانية. في الواقع**، يستحيل خدمة جميع المحتاجين بسبب محدودية الموارد. بمعنى أن معظم البرامج تقدم الميزات والخدمات إلى عدد محدود فقط من المستفيدين، وبالتالي تستثني الآخرين، سواء كان ذلك بشكل صريح أو غير صريح. فعلى سبيل المثال، إذا لم يتوفر في التدخل الذي يستهدف تدريب الشباب إلا عدد محدود من الأماكن، فسيحصل بعض الشباب على التدريب دون غيرهم. وبالمثل، إذا تم تنفيذ تدخل في منطقة معينة، يتم استبعاد الشباب المؤهلين في المقاطعات الأخرى. ويسمح التوزيع العشوائي لمديري البرامج بتخصيص الأماكن النادرة في تدخلاتهم بطريقة عادلة تمنح الجميع فرصة متكافئة للمشاركة. إذا تم التوزيع العشوائي بشكل علني (مثل اليانصيب أثناء فعّالية عامة)، يمكن أن يعزز ذلك أيضاً الشفافية في عملية الاختيار ويقلل من مخاوف السكان من أن الاختيار كان قائماً على التفضيلات الشخصية أو السياسية.

ومع ذلك، يستحيل في بعض الأحيان تشكيل مجموعة مقارنة نقية عشوائياً بحيث لا يتعرض الشباب في تلك المجموعة للتدخل. يتجنب كلا من التصميم العشوائي المرحلي والترويج العشوائي الفصل الصارم بين مجموعتي المعالجة والمقارنة ويمكن أن يكونا بديلاً عملياً لتصميم تقييم الأثر التجريبي عندما لا تكون التصاميم العشوائية ممكنة أو مستحسنة.

معلومة استرشادية

باتباع النهج المرحلي، يجب تخصيص وقتٍ كافٍ بين كل مرحلة من مراحل التدخل لبيان التأثيرات. على سبيل المثال، إذا اعتقد أحد مسؤولي التدخل بأن أثر التدخل لن يتجلى إلا بعد عامين من التدخل، يجب أن يكون الوقت بين المرحلتين الأولى والأخيرة عامين على الأقل. ولهذا، قد لا تكون البرامج الصغيرة أو قصيرة المدى مناسبة لهذا النهج.

التصميم العشوائي المرحلي

نظراً لأن العديد من البرامج تنشط في المجتمع لعدة سنوات، قد يكون من الصعب سياسياً وبرمجياً عدم إشراك مجموعة من الأفراد المحتاجين في التدخل. وهناك وجه مختلف للتصميم العشوائي يتمثل في التصميم المرحلي، ويتمثل الاختلاف الرئيسي في طريقة تعيين الأشخاص في مجموعتي المعالجة والمقارنة. ففي الممارسة العملية، يتم تقسيم المستفيدين المحتملين بشكل عشوائي إلى مجموعتين أو أكثر ثم إطلاق التدخل بمرور الوقت، بحيث يشارك الأفراد من المجموعة الأولى في التدخل أولاً، ثم المجموعة الثانية فالثالثة، وهكذا. وعندما تكون المجموعات في قائمة الانتظار، يمكن أن تكون بمثابة مجموعة المقارنة حتى تستفيد من التدخّل.

وعلى سبيل المثال، قد تمتلك منظمة غير حكومية معينة ميزانية كافية لتدريب 1500 شاب، ولكن قد لا يكون لديها القدرة على إجراء كامل التدريب في وقت واحد. بدلاً من ذلك، يمكنها اختيار تدريب 500 شخص في السنة على مدى ثلاث سنوات. إذا أمكن تحديد جميع المشاركين الـ 1500 في البداية، قد يكون التوزيع العشوائي المرحلي أفضل طريقة تقييم يمكن اعتمادها. يتم تقسيم الـ 1500 شاب بشكل عشوائي على ثلاث مجموعات. وفي السنة الأولى، بينما تتلقى المجموعة الأولى التدريب، تظل المجموعتان الثانية والثالثة في قائمة الانتظار ويمكن أن يكونا بمثابة مجموعتي مقارنة. وفي العام الثاني، تبقى المجموعة الثالثة فقط لأغراض المقارنة. وبحلول السنة الثالثة، سوف تحصل جميع المجموعات الثلاث على التدريب.

وحيث إن الأفراد يتم اختيارهم عشوائياً للمجموعات المختلفة، يمكن مقارنة المستفيدين من التدخل أولاً بالمستفيدين من التدخل فيما بعد. وتناسب هذه الطريقة في كثير من الأحيان الطبيعة الخاصة بإطلاق العديد من البرامج.

ومع ذلك، ونظراً لاستفادة الجميع في نهاية المطاف من البرنامج، لا يعد التصميم المرحلي النهج الأمثل لبيان الأثر الطويل الأجل للتدخل بسبب عدم وجود مجموعة مقارنة في نهاية المطاف. وستواجه حتى البرامج الكبيرة طويلة الأمد صعوبة في مطالبة المشاركين بالانتظار لمدة ثلاث أو أربع سنوات قبل أن يأتي دورهم في المشاركة. لذا، غالباً ما تكون الفترة الزمنية للنتائج محدودة بين عام أو عامين. علاوة على ذلك، هناك خطر من أن يغير المشاركون سلوكياتهم أثناء فترة الانتظار إلى الانضمام إلى التدخل، مما يقوض قدرتهم على العمل كمجموعة مقارنة جيدة. على سبيل المثال، قد يتوقفون عن البحث عن عمل تحسباً للانضمام إلى التدخل الذي يستهدف التدريب على المهارات.

تصميم الترويج العشوائي

قد لا يكون من الممكن أو المستحسن في بعض الحالات استبعاد أي مستفيدين محتملين وعندما لا يتم إطلاق التدخل على مراحل. عندئذ، قد تكون طريقة الترويج العشوائي (تسمى أيضاً تصميم التشجيع) مناسبة. عندما لا يكون من الممكن تعيين الشباب عشوائياً إلى مجموعتي تستفيد الأولى من التدخل والأخرى لا تستفيد، قد يمكن بدلاً من ذلك الترويج للتدخل عشوائياً. بمعنى أنه بدلاً من اختيار المستفيدين وغير المستفيدين من الميزات والخدمات عشوائياً، فإننا نشجع المستفيدين من هذه الميزات عشوائياً.

قد يأخذ التشجيع العشوائي أشكالاً مختلفة. ففي حالة حسابات الادخار الخاصة بالشباب، قد نقوم بالإعلان عن المبادرة بشكل عشوائي في مدارس مختارة. وبالنسبة للبرامج التدريبية، يمكننا تعيين أخصائي اجتماعي لزيارة منازل الشباب العاطلين عن العمل بشكل عشوائي، ووصف البرنامج وعرض تسجيل الشباب على الفور. وفي حالة وجود حملة لمحو الأمية المالية، نستطيع إرسال رسائل نصية إلى جزء من الجمهور المستهدف بشكل عشوائي، دون الآخر. وفي جميع الحالات، لن يشارك بعض الأشخاص من المجموعة التي تم الترويج إليهم في التدخل، وسيدخل البعض من المجموعة التي لم يتم الترويج إليها في التدخل بالفعل. ولكن إذا كان التشجيع فعّالاً، فيجب أن يكون معدل التسجيل من صفوف المجموعة التي تم الترويج إليها أعلى من معدل بين التسجيل من صفوف المجموعة التي لم يتم الترويج إليها.

لتقييم أثر التدخل، لا يمكننا، للأسف، ببساطة مقارنة نتائج المشاركين في التدخل بنتائج غير المشاركين. يختلف الأشخاص الذين يختارون المشاركة في التدخل دائماً تقريباً عن أولئك الذين لم يختاروا ذلك، وقد لا يكون العديد من هذه الاختلافات قابلاً للرصد أو للقياس. حتى لو كان الترويج عشوائياً، لا تكون المشاركة في التدخل عشوائية، لذلك تكون المقارنة بين المشاركين وغير المشاركين أشبه بالمقارنة بين التفاح والبرتقال.

ومع ذلك، يمكننا مقارنة النتائج بين جميع من حصلوا على التشجيع وجميع الشباب في مجموعة المقارنة. ونظراً لإجراء الترويج عشوائياً، تمتلك المجموعات التي تم الترويج ولم يتم الترويج إليها، في المتوسط، خصائص متساوية. وبالتالي، يمكن إسناد الاختلاف المرصود في متوسط النتائج بين المجموعتين إلى حقيقة تسجيل هؤلاء الأشخاص في التدخل فقط بسبب حصولهم على الترويج.

تتمثل الميزة الرئيسية لهذا التصميم في أن حملات الترويج العشوائية لا تحرم أبداً أي شخص من المشاركة في البرنامج، ولكنها بدلاً من ذلك تمكن الناس من اتخاذ قرارهم بشأن المشاركة في التدخل بمحض اختيارهم. ومع ذلك، تحتاج هذه الدراسات في كثير من الأحيان إلى عينات كبيرة الحجم لتوفير تقديرات موثوقة عن الأثر، مما يزيد التكاليف.

معلومة استرشادية

قد يناسب تقييم الترويج العشوائي ما يلي:

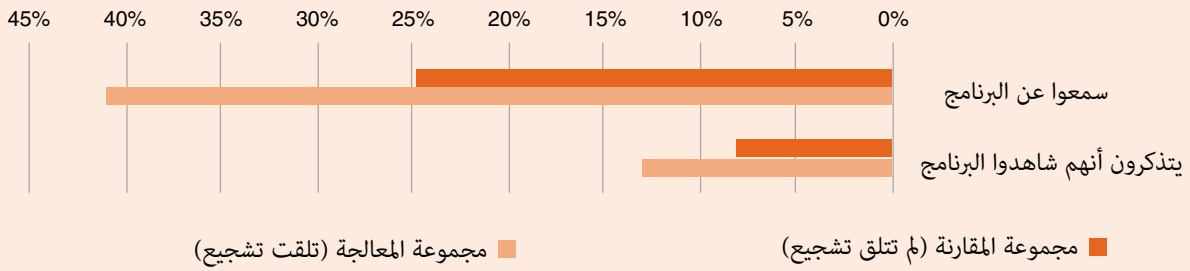
- التدخلات التي توزع قسائم التدريب
- التدخلات التي تشجع الشباب على فتح حسابات الادخار
- التدخلات التي تستخدم الحملات القائمة على وسائل الإعلام الجماهيري.

المربع 5-4: تقييم التدخل التعليمي الترفيهي في مصر

المشروع هو برنامج تلفزيوني واقعي تم عرضه على إحدى القنوات التليفزيونية المصرية الأكثر شعبية بهدف تعزيز ريادة الأعمال بين صفوف المشاهدين الشباب. لتقييم أثر البرنامج، اختار فريق بحثي، بدعم من منظمة العمل الدولية، تصميم الترويج العشوائي. من عينة الدراسة التي تضم 9.277 شخصاً، تلقت مجموعة معالجة تم اختيارها عشوائياً رسائل تذكير نصية حول البرنامج لتشجيعهم على المشاهدة.

أظهر مسح المتابعة الذي تم إجراؤه بعد حوالي عام ونصف من عرض البرنامج بوضوح ارتفاع نسب معرفة الشباب في مجموعة المعالجة (الذين تلقوا رسائل التذكير) بالبرنامج ومشاهدتهم لحلقة واحدة على الأقل مقارنةً بالشباب في مجموعة المقارنة (الذين لم يستلموا رسائل التذكير)، انظر أيضاً الشكل 5-6.

الشكل 5-6: متابعة برنامج المشروع



يمكن استغلال هذه الاختلافات ذات الدلالة الإحصائية لتقدير التأثيرات الفعلية للبرنامج على توجهات المشاهدين ونتائج سوق العمل. ويمكن ذلك نظراً لإمكانية إجراء افتراضين:

1. عدم اختلاف مجموعتي المعالجة والمقارنة بشكل منهجي في أي خصائص مرصودة أو غير قابلة للرصد والتي يمكن أن ترتبط بمتغيرات النتائج بسبب التوزيع العشوائي.
2. إمكانية إسناد أي اختلاف بين مجموعتي المعالجة والمقارنة إلى الاختلاف في احتمالية مشاهدة البرنامج لأن استلام الرسائل وحدها لا يؤثر على التوجهات أو نتائج سوق العمل.

وتخلص الدراسة إلى أن مشاهدة البرنامج لم يؤثر على ميل الشباب نحو إقامة مشروع تجاري، ولكنه أدى إلى تقليل التوجهات التمييزية ضد المرأة.

المصدر: برسوم وآخرون، 2017.

الفرق في الاختلافات (DID)

للأسباب التي سبق توضيحها، من غير الممكن أو المستحسن في بعض الأحيان استخدام أساليب التقييم التجريبي. في هذه الحالة، هناك مجموعة من الطرق شبه التجريبية لتقييم الأثر، والتي يمكنها أيضا تقديم نتائج ذات صلاحية داخلية قوية، ومن أكثرها استخداماً نهج الفرق في الخلافات (DID) الذي يقارن بين التغيير في النتائج التي تحققها مجموعة المعالجة والتغيير في النتائج التي تحققها مجموعة المقارنة.

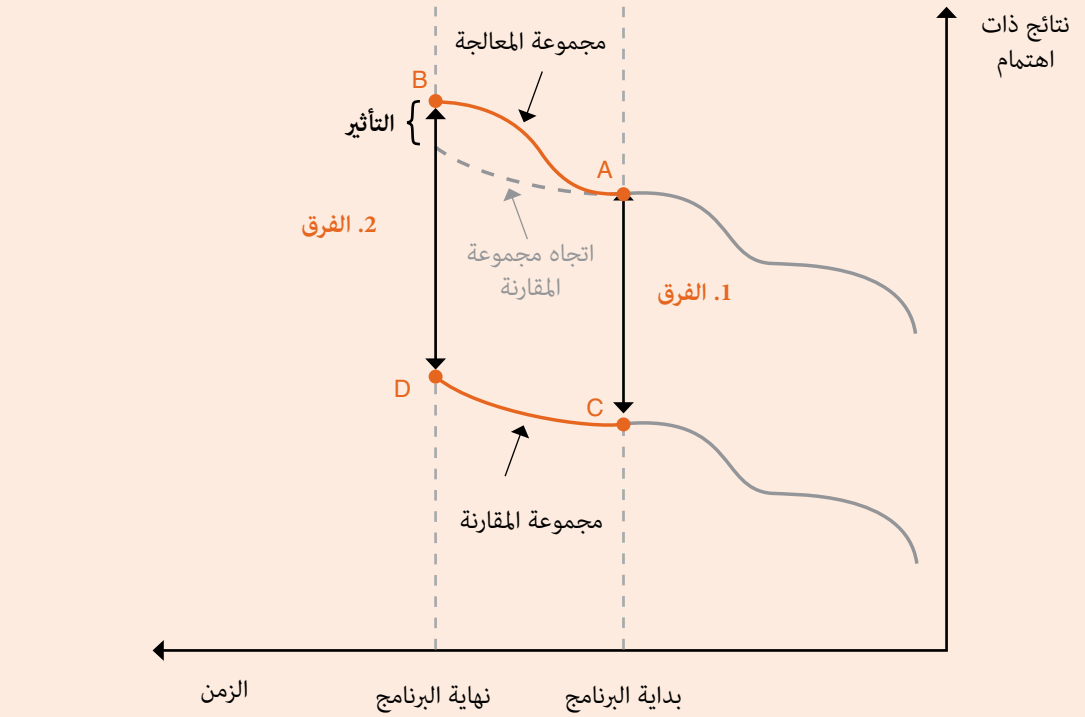
كيف يعمل

تحديد مجموعة المقارنة: تعتمد تصاميم الفرق في الاختلافات على وجود مجموعة مقارنة التي يمكن أن نفترض بشكل معقول بأن تطورها في نتائجنا الرئيسية سيكون نفس تطور مجموعة المعالجة خلال الفترة الزمنية للتدخل. وتحقيقاً لهذه الغاية، من المستحسن اختيار المجموعات ذات الخصائص المتشابهة.

لنتخيل أن هناك تدخلاً يهدف لتدريب الشباب على العمل مدته ستة أشهر، وأننا نريد تقييم آثاره على نتائج سوق العمل، ولا يمكن إجراء التوزيع العشوائي لأماكن التدريب. بدلاً من ذلك، سنأخذ عينة من الشباب من نفس السن والمستوى التعليمي والخلفية الاجتماعية الاقتصادية وحالة سوق العمل من مجتمع آخر كمجموعة المقارنة.

تقدير الأثر: لتطبيق تقنية تقييم الفرق في الاختلافات، نحتاج إلى (أ) قياس النتائج محل الاهتمام (على سبيل المثال، وضع سوق العمل، انظر الملاحظة 2) لكل معالجة ومجموعات المقارنة قبل بداية التدخل الذي يستهدف التدريب على العمل و(ب) قياس نتائج كلا المجموعتين في وقت معين بعد حدوث التدخل. وعلى الرغم من أننا حاولنا تحديد مجموعة مقارنة من الشباب بحيث تبدو مشابهة لمجموعة شباب المعالجة، من المرجح وجود اختلافات بين المجموعتين قبل بداية التدريب، وأن هذه الاختلافات سوف تستمر بعد ذلك. يبين الشكل 5-7 انخفاض مؤشر نتائج مجموعة المقارنة بصورة كبيرة بالنسبة لحالة معينة (على سبيل المثال، حالة العمالة) عند خط الأساس. ومع ذلك، لا يؤثر ذلك على الطريقة. يقارن أسلوب الفرق في الاختلافات الفرق في النتائج بين المجموعتين في نهاية التدخل (النقطة "ب" ناقص "د") ويعدلها للاختلاف في النتائج بين المجموعتين في البداية ("أ" ناقص "ج"). إن طرح هذه الاختلافات من بعضها بعضاً (بمعنى التفريق بين الاختلافين، ومن ذلك اشتق اسم الأسلوب) يعطي فكرة عن تأثير البرنامج؛ ويوضح إن كان التدخل الذي يستهدف التدريب قد زاد من حالة عمالة المشاركين مقارنة بغير المشاركين ومدى تلك الزيادة. يوضح السيناريو الوارد في الشكل 5-7 متوسط التأثير الإيجابي للتدخل الذي يستهدف التدريب على العمل.

الشكل 5-7: مثال عن تحليل الفرق في الاختلافات



المصدر: مقتبس من جيرتلر وآخرون، 2016.

افتراض "الاتجاه العام": يتمثل الافتراض الأساسي الذي يستند إليه هذا الأسلوب في أنه على الرغم من إمكانية الاختلاف النسبي في الخصائص المرصودة وغير المرصودة بين مجموعتي المعالجة والمقارنة (والتي تنعكس في مستويات الدخل المختلفة في بداية التدخل)، فإن اختلافاتهما ثابتة بمرور الوقت، أو غير متغيرة بزمن. ويمكننا ذلك من استخدام اتجاه مجموعة المقارنة لتقدير ما كان سيحدث لمجموعة المعالجة لولا تنفيذ التدخل. ولذلك، ليس علينا أن نفترض أن النتائج كانت ستظل ثابتة بدون التدخلات، بل أن مجموعتي المعالجة المقارنة تشتركان في نفس الاتجاه مع مرور الوقت. وهذا ما نقصده من افتراض "الاتجاه العام".

بالرجوع إلى مثال التدريب على العمل المذكور أعلاه، لكي نتمكن من تطبيق أسلوب الفرق في الاختلافات، علينا التأكد أنه لن يكون هناك أي عوامل تؤثر بشكل منهجي على نتائج الشباب من "المجتمع المعالج" بشكل مختلف عن نتائج "مجتمع المقارنة" على مدى الأشهر الستة المقبلة، بصرف النظر عن مهمة التدريب. على سبيل المثال، قد يؤدي النمو الاقتصادي الأسرع أو السياسة المحلية الجديدة التي تقدم حوافز للشركات من أجل تشغيل الشباب أو إغلاق أي مؤسسة كبرى لأصحاب العمل الكبار في إحدى المجتمعين، إلى تقويض هذا الافتراض وبالتالي وجود تحيّز في نتائج التقييم الخاصة بنا.

يتمثل إحدى الاختبارات الجيدة لتحديد إن كان من الواقعي افتراض تساوي التوجهات بين

الملاحظة 5. طرق تقييم أثر التدخلات التي تستهدف عمالة الشباب

المشاركين وغير المشاركين في مقارنة التغييرات في نتائجهم قبل تنفيذ التدخل. ويتطلب هذا النهج نقاط بيانات متعددة قبل التدخل. وحيث إن العديد من الدراسات الاستقصائية الأساسية يمكن أن تصبح باهظة التكلفة بسرعة كبيرة، يمكن إجراء هذا الاختبار بسهولة إذا كانت البيانات الإدارية حول مؤشرات نتائجنا الرئيسية متاحة بتكلفة منخفضة (على سبيل المثال، الوضع الوظيفي من وكالات التوظيف في القطاع العام أو درجات الاختبار من السنوات الدراسية السابقة). إذا تحركت نتائج المجتمعين بالترادف قبل بداية التدخل، يمكننا أن نكون أكثر ثقة بأن نتائجها ستستمر في هذا الاتجاه أثناء التدخل. ومع ذلك، إذا كانت اتجاهات ما قبل التدخل مختلفة، قد لا يكون افتراض الاتجاه المتساوي صحيحاً.

متى يمكن استخدام تصميم الفرق في الاختلافات؟

يتم تطبيق هذه الطريقة بشكل مفيد عندما تتوفر بيانات جيدة عند نقاط متعددة قبل بداية التدخل نظراً لأنها تفترض ثبات الاختلافات بين المشاركين وغير المشاركين بمرور الوقت. ولتحسين مصداقية تقديرات الأثر، من المفضل إجراء ثلاث جولات على الأقل لجمع البيانات: اثنتان قبل المعالجة، وواحدة على الأقل في نهاية التدخل (انظر أعلاه). وهذا يعني أنه ما لم تتوافر بيانات حول المشاركين وغير المشاركين من خلال قنوات أخرى، مثل المسح الأسري الحالي، يمكن أن تزداد تكاليف هذا التقييم بكثير عن تكاليف تقنيات تقييم الأثر الأخرى.

المزايا

- توفر هذه الطريقة سبيلاً لمراعاة الاختلافات المرصودة وغير المرصودة بين المشاركين وغير المشاركين. وبتعبير أدق، تتحكم في جميع التأثيرات الفردية التي تظل ثابتة بمرور الوقت، أو التي تشترك في نفس مسار التغيير بمرور الوقت (بمعنى أن مجموعتي المعالجة والمقارنة تظهران اتجاهات مماثلة في النتائج محل الاهتمام).
- وحتى إن كانت الطريقة غير تجريبية، إلا انها تسمح بإجراء فحص (جزئي) للفرضية التي تجعلها صالحة داخلياً، ما يعني أننا نستطيع معرفة إن كانت آثارنا المقدرة صحيحة أو غير ذلك. وفي حالة توفر بيانات إدارية جيدة، يمكن تطبيق هذه الطريقة بسهولة نسبية وحتى بعد تنفيذ التدخل، استناداً إلى البيانات السابقة واللاحقة من البرنامج.

العيوب

- تنتج هذه الطريقة نتائج أقل موثوقية من طرق الاختيار العشوائية.
- من أجل اختبار الافتراض الرئيسي "للاتجاهات المشتركة"، هناك حاجة إلى ثلاث مجموعات من البيانات على الأقل، وبالتالي يمكن أن يكون التنفيذ مكلفاً إن لم تكن البيانات متوفرة في البداية.

المربع 5-5: تقييم مكون تنشيط سوق العمل للمشاركين في برنامج للتحويل النقدي المشروط في الأرجنتين

برنامج خطة جيفيس (Plan Jefes) هو برنامج للتحويل النقدي المشروط تم إطلاقه خلال الأزمة الاقتصادية الأرجنتينية 2001-2002. وقد شملت إصلاحات هذا البرنامج بعد التعافي من الأزمة تنفيذ التدريب والتأمين الوظيفي (Seguro de Capacitacion y Empleo) في عام 2006، من أجل تقديم الدعم في رفع مستوى المهارات والتدريب المهني، والبحث عن عمل والالتحاق الوظيفي للمشاركين المؤهلين في البرنامج.

ويحصل المشاركون في برنامج التدريب والتأمين الوظيفي على راتب شهري ويتم موافاتهم بتدابير التنشيط التالية:

- المساعدة لإتمام التعليم الابتدائي والثانوي
- التدريب المهني والتلمذة الصناعية
- خدمات الوساطة العمالية
- تدابير إيجاد فرص العمل غير المباشرة (مثل دعم التوظيف)
- تشجيع العمل الحر وإنشاء المشاريع الصغيرة.

درست منظمة العمل الدولية أثر تطبيق أدوات سوق العمل النشطة للمستفيدين من برنامج خطة جيفيس على وضعهم في سوق العمل وجودة العمل بمقاييس الفرق في الاختلافات. ومن أجل عزل تأثير هذه الأدوات، كان لابد من تحديد مجموعة مقارنة ذات سمات مشابهة للمشاركين في برنامج التدريب والتأمين الوظيفي.

وحيث إن الانتقال من برنامج خطة جيفيس إلى البرنامج الجديد كان تدريجياً، تمكن الباحثون من اختيار المشاركين في برنامج الخطة الذين استوفوا متطلبات الاستفادة من برنامج التدريب والتأمين الوظيفي، ولكن لم يتم نقلهم بعد إلى البرنامج الجديد. وتمثلت إحدى الافتراضات الرئيسية المهمة لاستراتيجية تحديد الهوية في أن الانتقال بين البرامج لم يتأثر بأي عوامل قد تكون السبب في الفروق في النتائج محل الاهتمام. وتم اختيار ما مجموعه 1.149 من غير المشاركين، استناداً إلى بيانات من المسح الأسري الدائم للأرجنتين - وهو مسح يجريه كل ثلاثة أشهر المعهد الوطني الأرجنتيني للإحصاء (INDEC) الذي يتضمن أسئلة عن الخصائص الشخصية للأفراد والتعليم وأداء سوق العمل. كان المشاركون المختارون وغير المشاركين متشابهين في الجنس والعمر ومستوى التحصيل العلمي.[5]

قارن المقيِّمون مجموعة من النتائج بين المجموعتين في فترتين زمنيتين مختلفتين (خط الأساس والمتابعة). وقد سمح هذا النهج بتحديد التأثيرات السببية لبرنامج التدريب والتأمين الوظيفي أثناء التحكم بالتحيّز فيما يخص الاختيار، وذلك بسبب الخصائص المرصودة وغير المرصودة للمشاركين.

وأظهرت نتائج الدراسة أن البرنامج أثر إيجابياً على جودة عمل المشاركين، بمعنى احتمال وجود وظيفة في القطاع المنظم وارتفاع معدل الأجر في الساعة، وانخفاض احتمالية الحصول على وظيفة منخفضة الأجر والعمل لساعات مفرطة. ولم يؤثر على وضعهم الوظيفي (أي احتمال توظيفهم). كما أظهر التقييم آثاراً غير متجانسة، وكشف عن أن البرنامج كان له تأثير كبير على المستفيدين من الشباب، ولكن لم يكن له تأثير على النساء.

5 من أجل تصحيح الاختلافات الملاحظة، قام الباحثون أيضاً بتطبيق طريقة مطابقة درجة الميل (PSM). انظر القسم التالي للحصول على وصف أكثر تفصيلاً لهذه المنهجية.

مطابقة درجة الميل (PSM)

تعتبر مطابقة درجة الميل نهجاً شائع الاستخدام بين طرق التقييم شبه التجريبية، ويتمثل مبدأها الأساسي في تشكيل مجموعة مقارنة من خلال مطابقة المشاركين مع غير المشاركين المماثلين، بناءً على احتمالية توقعهم للمشاركة في التدخل. وتسمى هذه درجة الميل، والتي تحسب على أساس مجموعة من الخصائص تمت ملاحظتها.

كيف تعمل

يجب تحديد مجموعة من المتغيرات المشتركة ذات الصلة من أجل حساب درجة الميل لغير المشاركين، بناءً على احتمال خضوعهم للمعالجة، بهدف إدراج حساب جميع المتغيرات المشتركة التي تؤثر على المشاركة في البرنامج والنتائج في حساب درجة الميل. بعدئذ، يتم مطابقة غير المشاركين مع المشاركين بناءً على نتائج كل منهما. هناك طرق مختلفة لمطابقة الإجراءات، غير أن الطريقة الأكثر شيوعاً تتمثل في مطابقة أقرب الجيران، بحيث يتم مطابقة كل مشارك مع غير المشارك صاحب أقرب درجة ميل. وكلما كانت الدرجة متقاربة، كانت جودة المطابقة أفضل. يمكن إجراء اختبارات التوازن لتقييم مدى نجاح المطابقة. وبالتالي، فإن متوسط الفرق في النتائج محل الاهتمام بين المجموعتين يعادل أثر التدخل.

وكمثال على ذلك، لنفترض أن هنالك برنامج تدريب على المهارات يستهدف الشباب في المناطق الريفية بمشاركة 1000 مشارك. وتتوفر بيانات ما قبل البرنامج حول الخصائص الأساسية للمشاركين، مثل: الجنس والعمر والتعليم والجوانب الرئيسية لتاريخ مشاركتهم في سوق العمل. يمكن استخدام بيانات المسح الثانوية القائمة لتشكيل مجموعة المقارنة على أساس درجة ميلها، مع تقدير احتمالية المعالجة لعدد كبير من الأفراد على أساس الخصائص المذكورة أعلاه.. سيتم اختيار ما مجموعه 1000 شخص من ذوي أفضل درجات الميل المطابقة كمجموعة مقارنة للتدخل. ويمكن جمع بيانات ما بعد التدخل لمجموعة المقارنة من خلال نفس مصدر البيانات الثانوية (إذا كانت مجموعة بيانات، فقد يتم جمع موجات منتظمة).

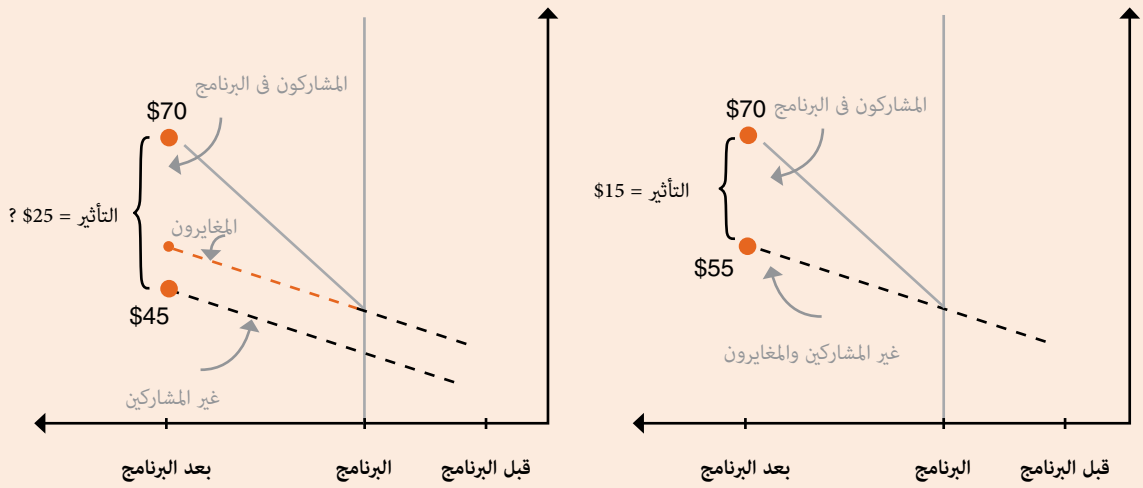
المربع 5-6: مقارنة المشاركين وغير المشاركين

على الرغم من أننا قد نستطيع تحديد مجموعة المقارنة في بعض الأحيان، إلا أنه لن تتوفر لدينا بيانات إلا عن عدد قليل من متغيرات النتائج الرئيسية، ولن توجد معلومات حول المتغيرات المشتركة مثل الخلفية الاجتماعية والاقتصادية والمعارف والمهارات وما إلى ذلك. عندئذ يمكننا استخدام منهجية تقييم التأثير البسيطة ومقارنة نتائج المشاركين وغير المشاركين. وبالتالي، يتم تقدير الوضع المغاير للواقع من خلال نتائج الأشخاص الذين لم يشاركوا في البرنامج. ومع ذلك، من غير المحتمل أن تؤدي هذه الطريقة إلى نتائج موثوقة أو معلومات مفيدة حول الأثر الحقيقي لبرنامجنا.

وعلى وجه الخصوص، إذا اختلف غير المشاركين (مجموعة المقارنة) عن المشاركين (مجموعة المعالجة) بطرق مرتبطة بالنتائج، فلن يكون هذا النوع من المقارنة صالحاً وسيتضمن تحيّزاً في الاختيار. وبتعبير أدق، تعتمد هذه الطريقة على فرضيتين قويتين. أولاً، يجب أن نفترض أن المشاركين في البرنامج وغير المشاركين يتمتعون، في المتوسط، بنتائج متشابهة في بداية البرنامج.

يصور الجانب الأيمن من الشكل 5-8 وضعاً يفترض زيادة دخل المشاركون بالفعل في بداية التدخل عن غير المشاركين. تؤدي هذه الحالة إلى المبالغة في تقدير التأثير الحقيقي لتدخلنا.

الشكل 5-8: مقارنة بين المشاركين وغير المشاركين



ثانياً، يجب الافتراض بأن تطور المجموعتين كان سيكون مماثلاً بمرور الوقت لولا حدوث التدخل، مما يتطلب الافتراض بأن طريقة تفاعل المشاركين وغير المشاركين كانت ستكون متطابقة، في المتوسط، مع جميع العوامل الخارجية. يرجى ملاحظة انطباق هذا الافتراض على الحالة المبينة في الجانب الأيمن من الشكل 5-8. يتحرك الخط المنقط باللون الأسود، الذي يصف تطور غير المشاركين بمرور الوقت، مع الخط الأحمر المنقط، الذي يصف تطور المشاركين في حالة عدم تنفيذ التدخل، بالتوازي مع الوقت. للحصول على تقديرات دقيقة للأثر من خلال هذه الطريقة، يجب الإبقاء على هذين الافتراضين.

ويتمثل العيب الرئيسي لهذه الطريقة في أننا إذا لم نرصد سوى النتائج بعد التدخل، فلن نستطيع اختبار أي من الافتراضين مباشرة، الأمر الذي قد لا يكون صحيحاً في العديد من الحالات. وبالنظر إلى معيارنا لاختيار الشباب للتدخل، قد لا يكون على أساس الأولوية من يأتي أولاً يُخدم أولاً. عندئذ، من المرجح أن يشارك في التدخل الشباب الذين يمكنهم الوصول بشكل أفضل إلى المعلومات المتعلقة بوجود البرنامج، وأولئك الذين يعيشون بالقرب من مكانه، وأولئك الذين يشجعهم آباؤهم أو ببساطة المتحمسين أكثر للمشاركة.

للإيجاز، لا يُنصح بإجراء المقارنة بين المشاركين وغير المشاركين في نهاية أي تدخل دون وجود معرفة واسعة بمتغيرات الخلفية التي تمكن من استخدام تقنيات أكثر تطوراً، مثل مطابقة درجة الميل، لتقييم الأثر.

متى يمكن استخدام مطابقة درجة الميل؟

تكون مطابقة درجة الميل طريقة مفيدة بشكل خاص عند توفر قدر كبير وغني من البيانات الثانوية، نظراً لأهميتها في تحديد درجة ميل جيدة ومطابقة عدد كافٍ من المشاركين وغير المشاركين الذين لديهم نتائج مماثلة، بمعنى إيجاد منطقة كبيرة كافية من الدعم المشترك. علاوة على ذلك، تعتمد مطابقة درجة الميل على افتراض أن العوامل المرصودة وحدها هي ما تؤثر على كلا من المشاركة والنتائج (افتراض الاستقلالية المشروطة). وبالتالي، لا ينبغي تطبيق مطابقة درجة الميل إلا إذا كان هناك فهم جيد لدوافع المشاركة في البرنامج والنتائج محل الاهتمام، ويجب تجنبها إذا كان من المتوقع أن تؤثر الخصائص غير القابلة للرصد على تلك المتغيرات. وفي جميع الأحوال، هناك حاجة إلى إجراء دراسة دقيقة قبل اتخاذ أي قرار بشأن عدد المتغيرات المحددة للاختيار من أجل تقدير درجة الميل.

المزايا

مطابقة درجة الميل هي منهجية قوية لتقييم الأثر ويمكنها، إذا تم استيفاء افتراضاتها، أن تساعد في إزالة التحيّز في الاختيار وتقديم نتائج ذات صلاحية داخلية. وكما هو الحال بالنسبة للطرق شبه التجريبية الأخرى، يمكن تطبيقها على أساس مصادر البيانات الموجودة ولا يلزم أي تخصيص عشوائي للتدخل.

من خلال مطابقة درجة الميل من أجل الحصول على المعالجة، تقلل هذه الطريقة من عدد الأبعاد التي يجب أن تتطابق فيها مع المشاركين ووحدات المقارنة إلى بعد واحد، وبالتالي تجعل المطابقة بسيطة نسبياً.

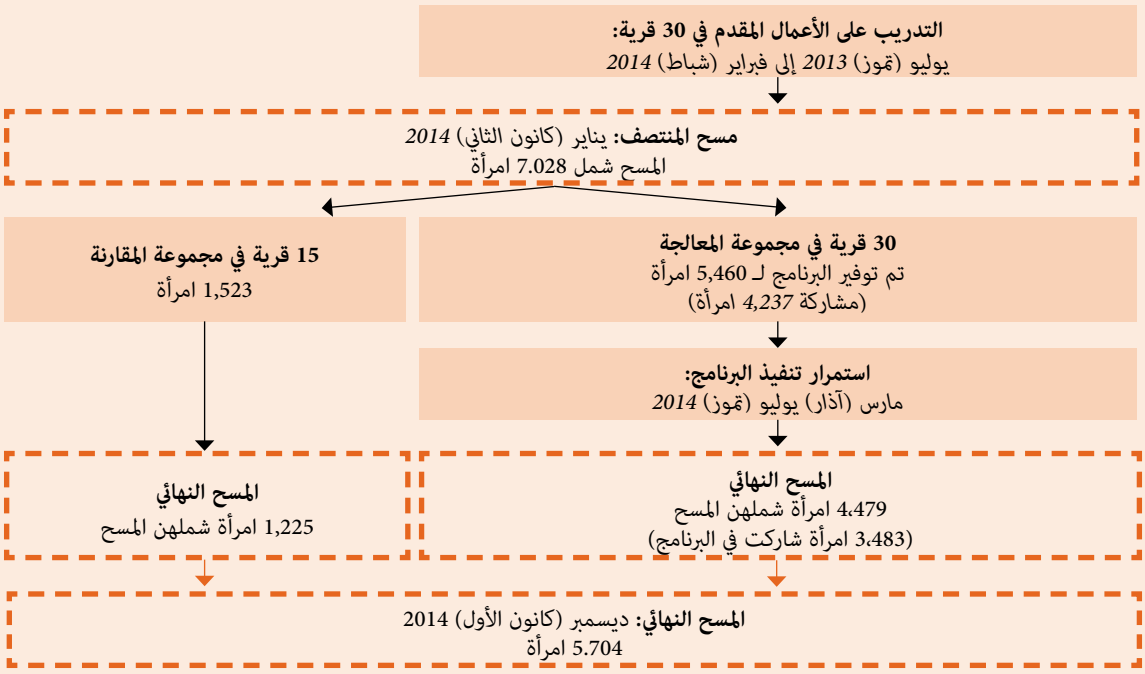
العيوب

- يتطلب تطبيق مطابقة درجة الميل إلى مجموعات كبيرة من البيانات في العادة.
- لا يمكن إجراء المطابقة إلا على خصائص قابلة للرصد. وبالتالي، يظل الخطر قائماً في إمكانية أن يؤدي التحيّز في الاختيار بسبب الخصائص غير المرصودة التي تحرك عملية المشاركة في البرنامج إلى التأثير على نتائج التقييم.
- ينطوي تطبيق مطابقة درجة الميل على تعقيدات إحصائية وتتطلب مستوىً مناسباً من الخبرة.

المربع 5-7: تمكين الشابات من خلال التدريب المهني والتدريب على الأعمال في المناطق الريفية في صعيد مصر - برنامج نقدر نشارك

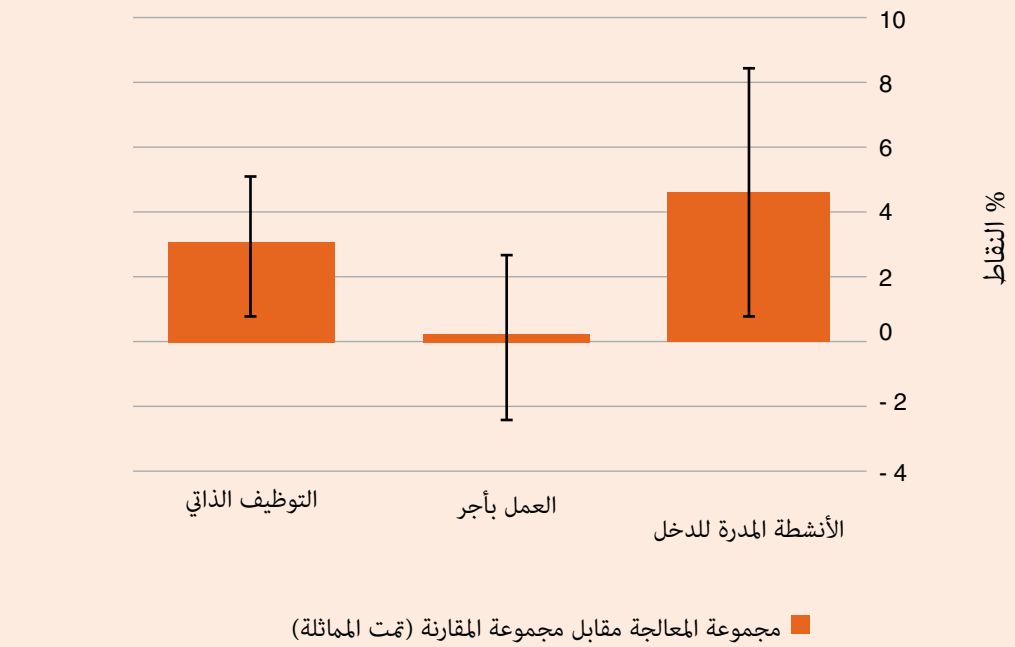
نفذ مجلس السكان برنامج "نقدر نشارك" في المناطق الريفية في صعيد مصر في الفترة من عام 2013 إلى 2014. واستهدف البرنامج 4,500 شابة تتراوح أعمارهن بين 16 و29 سنة، باعتماد نهج "سبل العيش الآمن" من خلال تلبية الاحتياجات الخاصة بالمجتمعات المحلية للنساء المستضعفات. وقد هدف التدخل إلى تمكين الشابات من خلال تزويدهن بالتدريب على المهارات المهنية والتجارية ودعمهن في إقامة المشاريع عمل أو البحث عن عمل. تألف برنامج التدريب من ثلاثة مكونات رئيسية: (1) التدريب على مهارات العمل، (2) التدريب المهني، (3) المهارات الحياتية والحقوق القانونية والتربية المدنية.

الشكل 5-9: تصميم تقييم الأثر (مبسط)



صاحب هذا التدخل تقييم لأثر برنامج نقدر نشارك على نتائج مشاركة الشابات في سوق العمل وتدابير التمكين الاجتماعي. استخدم التقييم تصميم مطابقة درجة الميل. وتم حساب التأثيرات من خلال مطابقة النساء اللواتي شاركن في البرنامج مع نساء لهن خصائص اجتماعية واقتصادية مماثلة من القرى في مجموعة المقارنة ومقارنة النتائج الرئيسية للبرنامج فيما بينهما (انظر الشكل 5-9).

الشكل 5-10: الآثار على نتائج المشاركة في سوق العمل



وجد التقييم أثراً كبيراً للبرنامج على التمكين الاقتصادي للمشاركات في البرنامج، تم قياسه من خلال مشاركتهن في الأنشطة المدرة للدخل. وكانت المشاركات في البرنامج أكثر عرضة بنسبة 4.5 نقطة مئوية للمشاركة في نشاط مدر للدخل من النساء في مجموعة المقارنة. وعلى النحو المبين في الشكل 5-10، كان معظم الأثر الإيجابي مدفوعاً بزيادة مشاركة المشاركات في أنشطة التوظيف الذاتي. في المقابل، لم يتغير مستوى المشاركة في العمل بأجر بشكل كبير بالنسبة للنساء في مجموعة المعالجة.

المصدر: منظمة العمل الدولية، 2017.

تصميم انحدار الانقطاع (RDD)

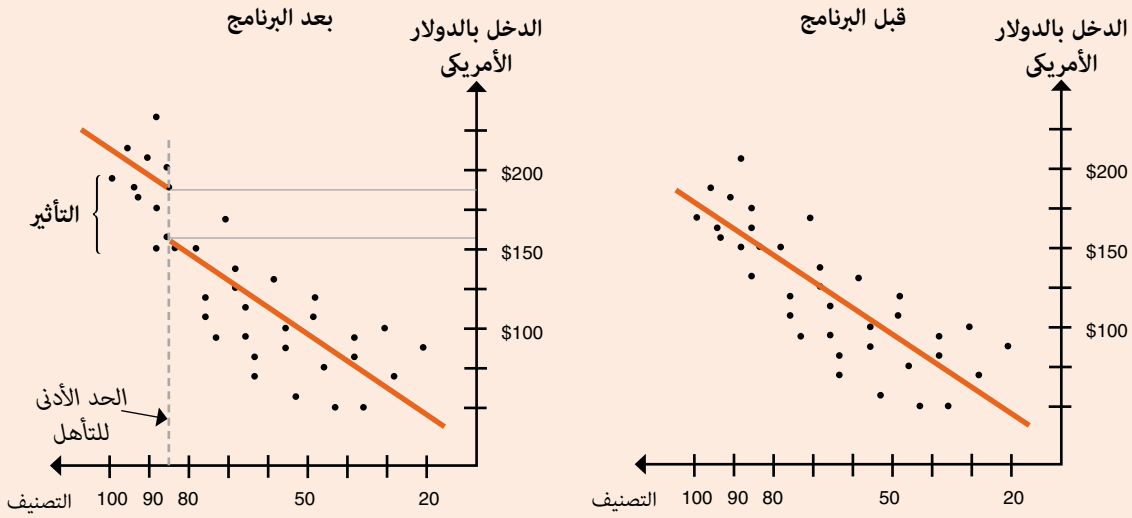
غالباً ما تستخدم تصميمات انحدار الانقطاع (RDDs) عندما تعتمد أهلية التدخل في سوق العمل على شكل من أشكال الترتيب المستمر للمستفيدين المحتملين أو المتقدمين، على سبيل المثال السقف العمري للمشاركة.

كيف يعمل

تتمثل فرضية تصميم الانقطاع (أو مؤشر الأهلية) في أن الأشخاص الذين يسجلون نقاط أعلى وأقل بقليل من السقف المحدد لا يختلفون كثيراً عن بعضهم بعضاً، أو قد يكون الفرق مستمراً على الأقل عبر الدرجات. فعلى سبيل المثال، من غير المرجح أن يكون الشباب الذين يبلغون 25 عاماً والذين قد يكونون مؤهلين لبرنامج تدريب على مهارات الشباب، مختلفين تماماً عن أقرانهم الذين يبلغون من العمر 26 عاماً والذين قد لا يعتبرون مؤهلين بعد الآن. إذا كان لدينا وضع يكون فيه بعض هؤلاء الشباب الذين يتلقون البرنامج (أولئك الذين هم أعلى بقليل من السقف المحدد) وبعض أولئك الذين لا يحصلون عليه (أولئك الذين أقل بقليل من السقف المحدد) لا يختلفون بشكل أساسي عن بعضهم بعضاً، عندئذ تمكننا مقارنة نتائج هاتين المجموعتين، بدورها، من تحليل أثر البرنامج.

يوضح الشكل 5-11 ما قد نجده عند تحليل أثر مبادرة القروض الصغرى للشباب. يشير الرسم البياني الأيسر إلى الميل في ارتفاع دخل أولئك الذين حققوا درجات أفضل بالفعل في وقت تقديم طلب المشاركة في البرنامج. وقد توجد أسباب كثيرة لذلك، على سبيل المثال، يرتفع مستوى دخل الذين يتمتعون بمستوى أعلى قليلاً من التعليم بالفعل، ويساعدهم تعليمهم أيضاً في الحصول على درجات أفضل. أو يتمتع الذين هم أكثر تحمساً

الشكل 5-11: رسم بياني عن عينة للانقطاع



لإقامة مشروع تجاري بالفعل بمهارات ريادة الأعمال بشكل أكبر، وينعكس ذلك في دخولهم الأعلى، ويساعدهم هذا الدافع أيضاً في إقناع لجنة التحكيم بدعمهم. وهناك العديد من التفسيرات الأخرى الممكنة، والتي لا نحتاج بالضرورة إلى فهمها لتطبيق هذه الطريقة.

عند بدء البرنامج، حدد بنك التمويل الأصغر المحلي سقف استلام القرض عند 85 درجة، وتم قبول أو رفض دعم جميع المتقدمين وفقاً لترتيبهم بالنسبة لهذا السقف. والآن نود أن نقرر إن كان لبرنامج القروض الصغرى أي أثر على الدخول. وكما هو موضح في الرسم البياني 5-11 (الرسم البياني الأيمن)، نفترض حصول الذين حققوا درجة أقل من 85 على نفس النتائج السابقة، في حين أن دخل الذين حققوا درجة 85 فما فوق قد ازداد في جميع المجالات. من هذه المعلومات، يمكن تحديد أثر البرنامج، الذي سيتم تمثيله بالفرق في النتائج (بمعنى انقطاع العلاقة الخطية) بالقرب من نقطة الانقطاع.

متى يمكن استخدام تصميم انحدار الانقطاع؟

في العديد من الحالات، لا نستطيع التخطيط للتقييم أثناء تصميم البرنامج. ومع ذلك، قد نتمكن في بعض الأحيان من استخدام قواعد الاستهداف الخاصة بالبرنامج للحصول على معلومات لاحقة جيدة لمجموعة المقارنة. تستخدم بعض البرامج ترتيباً مستمراً للمستفيدين المحتملين، مثل درجات الاختبار، أو درجات الائتمان، أو مؤشر الفقر، ولديها نقطة انقطاع لقبول المشاركة في البرنامج. وبالنسبة للتدخلات التي تستهدف عمالة الشباب، يوجد في الغالب سقف ملزم من حيث العمر، بحيث لا يتأهل للمشاركة في البرنامج سوى الشباب الذين تقل أعمارهم عن سن محدد. يمكن استخدام قاعدة الأهلية هذه لإجراء تقييم الأثر على أساس تصميم انحدار الانقطاع.

المزايا

- يمكن تطبيق تصميم انحدار الانقطاع بعد البرنامج، في حالة توفر بيانات إدارية كافية.
- يمكن أن يستفيد تصميم انحدار الانقطاع من قاعدة قائمة لإنشاء مجموعة مقارنة صحيحة، وبالتالي لا تتطلب استبعاد مجموعة مؤهلة من التدخل.

العيوب

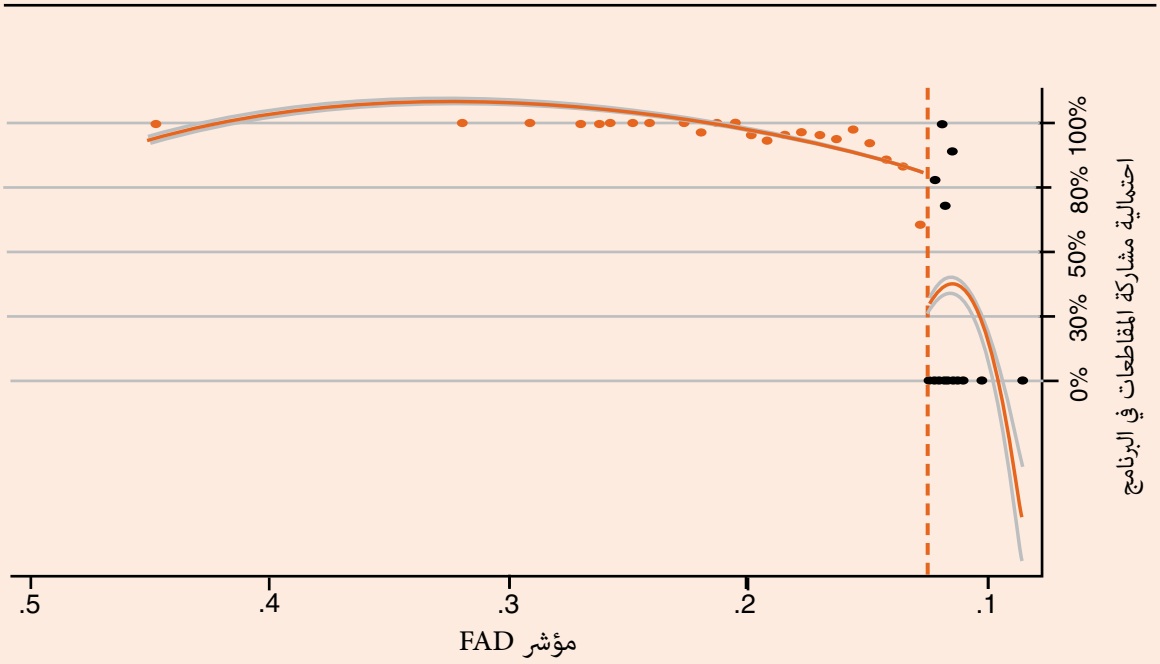
- يتمثل الشرط الرئيسي لاستخدام تصميمات الانقطاع في تحديد معايير المشاركة في البرنامج من خلال قاعدة استهداف محددة بوضوح؛ وبعبارة أخرى، من خلال مقياس أو درجة مستمرة. ولنجاح هذه الطريقة، نحتاج إلى العديد من عمليات الرصد في المنطقة التي تقع أعلى وأقل بقليل من نقطة الانقطاع من أجل الحصول على عدد كافٍ من الشباب للمقارنة فيما بينهم. وما لم يتم التقييم بدون بيانات أساسية أو بالاستفادة من سجلات البرامج الموجودة، يتطلب تصميم الانقطاع جمع بيانات مماثلة للتصميم العشوائي، وبالتالي يتحمل تكلفة مماثلة.
- تقتصر القيمة المعلوماتية للنتائج على العينة حول نقطة الانقطاع. وقد يكون ذلك مناسباً، على سبيل المثال، في المناقشات المتعلقة بتوسيع نطاق البرنامج ليشمل المجموعات أو المناطق العمرية الأخرى.

المربع 5-8: فعّالية برنامج العمل مقابل خدمات الرعاية الاجتماعية في نتائج سوق العمل: Construyendo Perú

تتزايد شعبية برامج الأشغال العامة كأداة سياسية في البلدان النامية. ففي الفترة من 2007 إلى 2011، نفذت حكومة بيرو برنامج Construyendo Perú بهدف أساسي وهو دعم الأفراد العاطلين عن العمل في حالات الفقر. وقد أتاح لهم البرنامج الحصول على تدريب مؤقت على التوظيف وتطوير المهارات من خلال تمويل مشاريع الاستثمار العام باستخدام مكثف للعمالة غير الماهرة.

قامت منظمة العمل الدولية بتقييم الآثار المتوسطة والطويلة الأجل للبرنامج باستخدام نهج انحدار الانقطاع. يستغل التقييم قاعدة رئيسية لمعايير المشاركة في البرنامج على مستوى المقاطعة تتمثل في اختيار مقاطعات المستفيدين من خلال تصنيفها وفقاً لمؤشر (Factor de Asignación Distrital)، وهو مؤشر مركب يجمع بين المعلومات الديموغرافية مع مؤشر لأوجه القصور في التنمية البشرية ومؤشر شدة الفقر. وعلى هذا النحو، سُمح للمقاطعات التي حققت درجة معينة أعلى من السقف المحدد في المؤشر (بمعنى المقاطعات التي تعاني من قصور أعلى في الفقر ونقص التنمية) بالمشاركة في البرنامج، ولم تشترك المقاطعات التي هي أقل من السقف في البرنامج. ويعد ذلك مثالاً على تصميم انحدار الانقطاع. وعلى النحو المبين في الشكل 5-12، كان من المرجح بدرجة كبيرة أن تشارك المقاطعات الأعلى بقليل من نقطة الانقطاع في البرنامج أكثر من تلك التي كانت أقل بقليل من السقف المحدد.

الشكل 5-12: الانقطاع في احتمالية مشاركة المقاطعات في البرنامج



لذلك، يتمثل السبب وراء التقييم في تقدير التأثير السببي للبرنامج من خلال مقارنة نتائج الأفراد حول نقطة الانقطاع في مؤشر FAD. ووجد التقييم أنه على المدى المتوسط (من ثلاث إلى خمس سنوات)، ساعد التدخل في زيادة العمالة وخفض عدم النشاط بالنسبة للنساء والمشاركين الأقل تعليماً في البرنامج. ومع ذلك، لم يتمكن البرنامج من تحسين آفاق المشاركين الأقل تعليماً من حيث جودة العمل، وفي الواقع، كان له أثر ضار من حيث منظور جودة العمل لدى النساء والأفراد الأكثر تعليماً (على سبيل المثال، بزيادة احتمال التوظيف في القطاع غير المنظم).

المصدر: اسكودرو، 2016.

مقارنات بسيطة: قبل وبعد

لا يمكن في بعض الأحيان إجراء التوزيع العشوائي، وعلاوة على ذلك، قد لا تتوفر الظروف المواتية لإجراء تقييم شبه تجريبي صحيح، وذلك مثلاً، عند عدم إمكانية العثور على مجموعة مقارنة مناسبة يتوفر لها معلومات خط الأساس و/أو عند عدم إمكانية تأكيد افتراض الاتجاه المشترك. عندئذ، من المستحسن النظر في جدوى إجراء تقييم كمي للأثر من الأساس.

إذا كان من غير الممكن تضمين مجموعة مقارنة في تقييم الأثر، يعتمد النهج الأكثر بساطة على مقارنة نتائج المشاركين في البرنامج قبل وبعد التدخل. ويمكن لهذا النهج البسيط تقديم فكرة عن التغيير الذي يحدث خلال مسار التدخل، ولكن يجب اعتباره جزءاً من نظام رصد وليس دليلاً على الأثر السببي للتدخل نظراً لعدم وجود طريقة للتأكد من أن التغيير المرصود يعزى إلى التدخل المعني أو إلى ظروف أخرى.

وبالنظر إلى مثال برنامج التدريب، قد نلاحظ أن الدخل الشهري للمشاركين قد ارتفع من 50 دولار قبل التدخل إلى 60 دولار بعد التدخل، وبالتالي يمكن الاستنتاج بأن أثر التدخل كان 10 دولار شهرياً للشخص الواحد (انظر الشكل 5-13، الرسم البياني على الجانب الأيسر). ومع ذلك، إذا لم يتم تنفيذ التدخل، كان مستوى الدخل سيزداد على أي حال بسبب تغير الظروف (أي أن الوضع الذي نواجهه يتوافق مع أحد السيناريوهات المبينة في الرسم البياني الأيمن من الشكل رقم 5-13)، وبذلك لن نستطيع على الحصول على تقدير دقيق للتدخل.

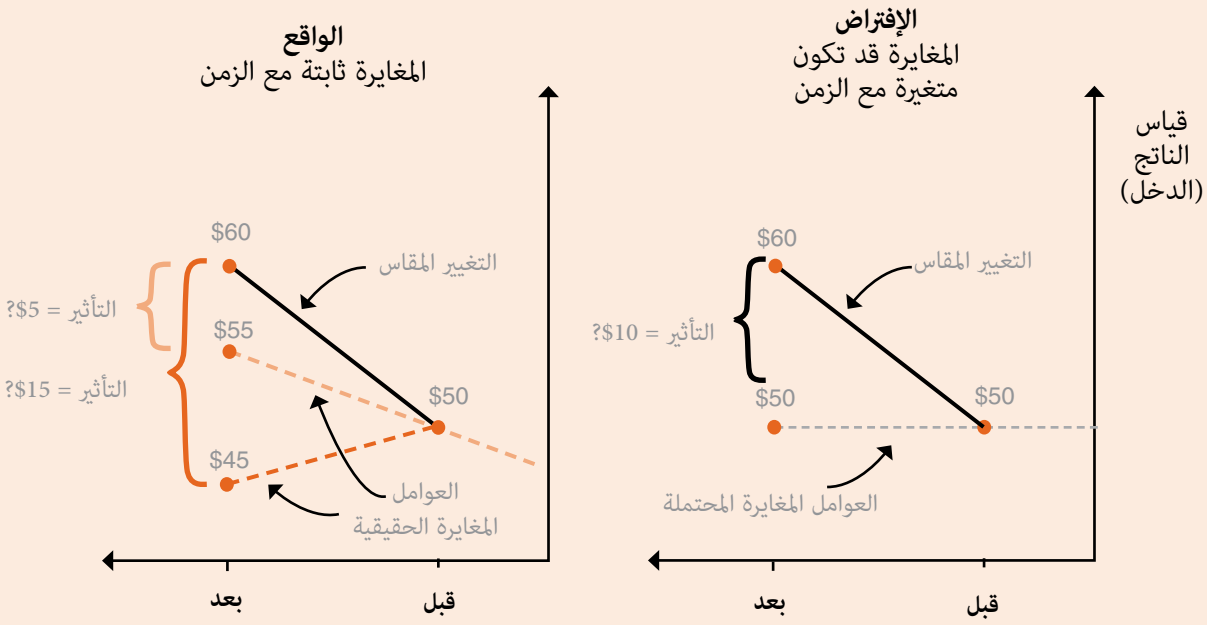
ونظراً لعدم إمكانية رصد السيناريوهات المغايرة للواقع (الخطوط المنقطة في الشكل)، لا توجد طريقة لمعرفة إن كانت الحالة التي تنطبق في تقييم معين هي الحالة الظاهرة على الجانب الأيسر من الشكل 5-13 أو تلك الموجودة على الجانب الأيمن. لذلك من المستحيل التأكد من أن الأثر الذي نقدره بهذه الطريقة هو الأثر الحقيقي لتدخلنا أو أثر "فاسد".

معلومة استرشادية

يُشار إلى مقارنات قبل وبعد أحياناً بدراسات التتبع، لا سيما في سياق المسوحات القياسية التي يتم إجراؤها للخريجين من التعليم الثانوي أو العالي أو برامج التعليم والتدريب المهني (والفني). لا تتوفر لدراسات التتبع إلا وسائل محدودة لتقييم الأثر، ولكنها تعد أدوات قوية لقياس قابلية توظيف الخريجين وجمع التعليقات لتحسين برنامج الدراسة. للحصول على دليل تفصيلي لتنفيذ دراسات التتبع، انظر شومبورج (2016).

قد يكون إجراء المقارنات قبل وبعد منطقياً في حالة وجود أسباب تدعو إلى الاعتقاد بأنه عند عدم تنفيذ المعالجة، ستبقى النتائج في المتوسط دون تغيير. ويمكن أن ينطبق ذلك على التدخلات التي (أ) يتم تنفيذها خلال فترة زمنية قصيرة (على سبيل المثال، التدخلات التي تستهدف التدريب القصير على المهارات، أو خدمات الإرشاد الوظيفي أو الأحداث التي تهدف إلى تغيير مواقف المشاركين) و(ب) من المتوقع أن يكون لها تأثيرات ذات اهتمام على المدى القصير. ومع ذلك، تظل القيود المذكورة أعلاه كما هي، وعلى عكس الأساليب التجريبية وشبه التجريبية التي يتم تنفيذها بشكل جيد، لا يمكن اعتبار المقارنات البسيطة قبل وبعد تقييماً قوياً للأثر. ومع هذا، يمكن تحسين مستوى قوتها أولاً، بالتحكم في عوامل الالتباس المحتملة في نموذج الانحدار (بدلاً من مجرد مقارنة النتائج) و/أو، ثانياً، من أجل التطبيق التكميلي للأساليب الكمية من أجل التوصل إلى الآليات السببية الكامنة للتغيير المرصود في النتائج.

الشكل 5-13: مقارنة النتائج قبل وبعد



الملاحظة 5. طرق تقييم أثر التدخلات التي تستهدف عمالة الشباب

تحسين ملاءمة تقييمات الأثر الكمي

على النحو الموضح في القسم السابق، توجد مجموعة من الأساليب الكمية الجيدة لتوفير إجابة ذات صلاحية داخلية لمسألة التقييم الأساسية، "هل ينجح المشروع؟"، بمعنى "هل أثر ذلك على النتائج ذات الاهتمام حسبما تم تحديدها في أهدافنا للتدخل والتعلم؟". وبينما يعد السؤال حول إن كان للتدخل ككل تأثير مهماً، لا يعد بجميع الأحوال السؤال الوحيد الذي نسعى الإجابة عنه.

من أجل الحصول على فهم تفصيلي وشامل لكيفية وأسباب تكشف آثار البرنامج الذي يستهدف عمالة الشباب، نحتاج إلى "التعمق بشكل أكبر". إن وجود فهم واضح لتغاير التأثيرات والآليات السببية المؤدية إلى التأثيرات المرصودة يساعدنا أيضاً على استخلاص دروس قيمة من تقييمنا والحصول على فكرة أفضل حول احتمالية نجاح البرنامج في بيئات أخرى.

وهذا يعيدنا إلى القضية المهمة حول الصلاحية الداخلية والخارجية (انظر مقدمة هذه الملاحظة). من أجل تحقيق الصلاحية الداخلية، يجب أن تكون أساليبنا قوية ومطبقة بشكل صحيح. من أجل تحقيق الصلاحية الخارجية، نحتاج إلى فهم العوامل السياقية ذات الصلة ببرنامجنا وتأثيرها المحتمل على نتائج التقييم. فمن الصعب الوصول إلى الأهداف باستخدام طريقة واحدة.

على سبيل المثال، يمكن للتقييمات التجريبية، إذا تم تنفيذها بشكل صحيح، أن تقدم لنا معلومات موثوقة حول التأثيرات التي يمكن أن تعزى بشكل فريد إلى مشروعنا، ولكنها لا تستطيع أن تخبرنا إلا القليل جداً عن قابلية تكرارها في بيئات أخرى. بشكل حاسم، تخبرنا تقييمات الأثر الكمي "ما حدث" - متوسط تأثير المعالجة - ولكنها لا تخبرنا "لماذا". لهذا الغرض، يلزم إجراء التطبيق التكميلي للأساليب النوعية.

قياس مجموعة متنوعة من التأثيرات

أولاً، قد يكون من المفيد الحصول على صورة أكثر دقة عن الأثر الفعلي للبرنامج، الأمر الذي يمكن تحقيقه جزئياً ضمن التصميمات الكمية المذكورة أعلاه. وفيما يلي بيان بالأسئلة الممكنة ذات الصلة:

- هل تختلف النتائج عبر مجموعات مختلفة من المستفيدين (على سبيل المثال، المزايا للفتيان دون الفتيات)؟
- ما أثر التدخل على المدى القصير مقابل المدى الطويل؟
- هل للتدخل آثار جانبية إيجابية أو سلبية؟ هل هناك أي نتائج مقصودة أو غير مقصودة تتجاوز المجموعة المستهدفة الفعلية؟

ثانياً، قد نكون مهتمين أيضاً باختبار التصميمات الشاملة، واختبار كيفية تغير فعّالية تدخلنا أثناء تعديل التصميم. وتمكننا هذه التصميمات من التحقيق في الأسئلة التالية:

- هل تصميم تدخل واحد أكثر فعّالية من الآخر؟ قد نرغب في مقارنة التدخلات البديلة (تقديم منح لإقامة المشاريع مقابل قروض إقامة المشاريع إلى أصحاب المشاريع الشباب، مثلاً)، أو اختبار المجموعة الأكثر فاعلية من مكونات البرنامج (التدريب وحده، التدريب بالإضافة إلى التمرين الداخلي، أو التدريب بالإضافة إلى التمرين الداخلي والتوجيه).

- ما الجرعة الأكثر فاعلية للتدخل؟ على سبيل المثال، هل يجب أن نقدم 20 أو 50 أو 100 ساعة من التدريب؟ (انظر الجدول 3-5 للحصول على مزيد من أسئلة تقييم الأثر.)

تساعد التصميمات الشاملة على تحديد أكثر من مجرد التأثير الكلي للمشروع، فهي تقيم أيضاً ميزات محددة من التدخل وأسباب نجاحها أو عدم نجاحها. فعلى سبيل المثال، قد يوفر البرنامج تدريباً مهنياً والتدريب على المهارات في مجال ريادة الأعمال، مثل النجارة أو الخياطة، إلى جانب قدر ضئيل من رأس المال الأولي لإقامة المشروع التجاري. قد ينطوي توفير المنح النقدية على تكاليف أو صعوبات من الناحية السياسية، وبالتالي قد يتساءل مدير البرنامج إن كان رأس المال الأولي ضرورياً، أو إن كان المشاركون قادرين على تنفيذ تدريبهم بدون رأس المال. يمكن أن يساعد التصميم المتقاطع في تحديد أفضل تصميم للمشروع في هذه الحالة، الأمر الذي يتطلب من الناحية العملية مقارنة نتائج مجموعات المعالجة المختلفة بمجموعة مقارنة ومع بعضهم بعضاً.

الجمع بين المناهج الكمية والنوعية

علاوة على ذلك، قد نهتم بتركيز الضوء على القنوات التي يعمل الأثر من خلالها – بمعنى فهم لماذا وكيف يتجلى الأثر. فعلى سبيل المثال، قد نرغب في الإجابة عن الأسئلة التالية:

- كيف تحدث الأشياء على النحو المرصود؟ ولماذا؟
- لماذا لم يسر المشروع (أو جزء منه) وفقاً لتوقعاتنا؟
- ما الذي يمكن أن نتعلمه من الفشل؟

إذا تم تصور طرق التحقيق المحتملة هذه في مرحلة التصميم، يمكن اختبار النظريات جزئياً ضمن الأساليب الكمية المذكورة أعلاه. ولتحقيق هذه الغاية، يجب أن تتضمن المسوحات أسئلة مصممة لاستنباط العوامل المختلفة (النتائج الوسيطة) التي يُفترض أن يعمل التأثير من خلالها من أجل التحقق إن كان التدخل يؤثر على هذه النتائج الوسيطة.

ومع ذلك، يعتبر الكثير من نتائج التدخلات التي تستهدف عمالة الشباب (مثل الصحة العقلية أو التمكين أو العلاقات الأسرية) معقدة ومتعددة الأبعاد وقد لا يتم استنباطها بوسائل كمية. تسمح الطرق المختلطة بتتبع المؤشرات النوعية وتوفير تحليل لدراسة حالة مختارة من أجل المساعدة في تحقيق فهم أفضل للديناميكيات ونتائج التدخل. فعلى سبيل المثال، يمكن أن تساعد المقابلات النوعية المنظمة وشبه المنظمة، التي يتمتع المشاركون خلالها بحرية التعبير عن قصص واقعية تندرج خارج فئات المعلومات القابلة للقياس، في إكمال الفهم حول أثر البرنامج (بامبيرجر وآخرون، ص 7-6؛ ليوو وفايسين، 2009).

قد تكون طرق جمع البيانات النوعية مفيدة بشكل خاص لجمع المعلومات حول حسن تنفيذ التدخل (انظر الملاحظة 4 بشأن تقييم الأداء). ولا شك أن فهم عملية التنفيذ يمثل ركيزة أساسية لاكتشاف كيفية تأثير عملية تنفيذ التدخل على النتائج وتفسير النتائج بشكل صحيح من أجل تحديد إن كان سبب النتائج المخيبة للآمال يعزى إلى نقاط الضعف في تصميم التدخل أو في التنفيذ. علاوة على ذلك، قد تسلط التقنيات النوعية الضوء على سبب حدوث نتائج محددة، وعلى وجه الخصوص أسباب اختلاف التأثيرات بين السكان المستهدفين (على سبيل المثال، بين الشباب في المناطق الريفية والحضرية أو بين الشابات والشباب).

الجدول 5-3: فئات أسئلة تقييم الأثر

السؤال	الوصف	متطلبات البيانات الإضافية	نتيجة تقييم العينة وتفسيرها
ما أثر التدخل الكلي على نتائج "أ" و "ب" و "ج" في المجموعة "س"؟ ... في السياق "ص"؟	هذا هو السؤال القياسي لتقييم الأثر	• غير منطبق (جمع البيانات القياسية بناءً على الطريقة المختارة)	متوسط تأثير التدخل الذي يستهدف التدريب على دخل الشباب هو أكثر من 20 دولار في الشهر. التدخل له تأثير إيجابي على الدخل
هل تختلف النتائج عبر مجموعات السكان؟	غالباً ما تؤثر التدخلات على المجموعات بشكل مختلف (عدم التجانس بين التأثيرات). قد يؤدي قياس متوسط التأثير فقط إلى إخفاء هذه الاختلافات، لذلك نحتاج إلى تقسيم التأثيرات حسب المجموعة السكانية	• المعلومات الاجتماعية الديموغرافية للمشاركين ومجموعة المقارنة (العمر، الجنس، مستوى الدخل، وما إلى ذلك) • حتى يتسنى تصنيف النتائج، يجب زيادة عدد الأشخاص المشمولين بالتقييم (حجم العينة) مع كل فئة من فئات المعلومات المراد تحليلها.	متوسط الزيادة في الدخل هو 40 دولارا للبنين و 0 دولار للفتيات. يستفيد الشباب الأكبر سناً أكثر من الشباب الأصغر سناً (30 دولاراً أمريكياً مقابل 10 دولارات في المتوسط). لذلك، لا يعد التدخل بنفس القدر من الفعّالية لجميع المشاركين. يجب أن نفهم لماذا تختلف استفادة المجموعات، مع تكيّف أهداف وتصميم البرنامج لاستيعاب مجموعات معينة
ما أثر التدخل على المدى القصير مقابل المدى الطويل؟	قد لا يكون التغيير في النتائج ثابتاً بمرور الوقت. قد تختفي الآثار على المدى القصير، في حين لا تتجلى الآثار طويلة الأجل إلا بعد مرور سنوات على انتهاء التدخل	• البيانات على مدى فترة زمنية طويلة (وهذا يعني في كثير من الأحيان متابعة مجموعتي المعالجة والمقارنة لعدة سنوات)	في نهاية البرنامج، نلاحظ متوسط دخل شهري للمشاركين – 5 دولار أمريكي (خسارة) مقارنة مع الضوابط. بعد عامين من البرنامج، يبلغ متوسط الزيادة في الدخل الشهري لمجموعة المعالجة 20 دولاراً. لم يستطيع الذين شاركوا في التدريب العمل بقدر نظرائهم خلال فترة التدريب، لذلك فقدوا الدخل. إلا أنه ومع مرور الوقت، تم تنفيذ التدريب وتمكن المشاركون من الحصول على دخل أعلى من نظرائهم الذين لم يشاركوا. قد يؤدي النظر إلى النتائج على المدى القصير إلى نتائج مضللة.
هل للتدخل آثار جانبية؟	قد يكون للتدخل آثار جانبية على غير المشاركين (إيجابية وسلبية)	• بيانات تتجاوز مجموعتي المعالجة والمقارنة، لتشمل أفراد الأسرة أو المجتمع	لا يحصل المشاركون فقط على متوسط دخل يبلغ 20 دولاراً، بل شهد جيرانهم أيضاً زيادة قدرها 5 دولارات. يبدو أن المشاركين نقلوا المعارف إلى الآخرين
هل تصميم التدخل "أ" أو تصميم التدخل "ب" أكثر فعّالية؟	غالباً ما يكون هناك غموض حول أفضل تصميم ممكن للتدخل. يمكن أن ترتبط الأسئلة بمقارنة التدخلات البديلة أو مزيج من مكونات البرنامج	• العديد من مجموعات المعالجة (تحصل الأولى على التصميم "أ"، والثانية على التصميم "ب"، وهكذا) • يجب أن يكون عدد الأشخاص المشمولين بالتقييم كبيراً بما يكفي لإنشاء أكثر من مجموعة معالجة واحدة بالإضافة إلى مجموعة مقارنة	متوسط الزيادة في الدخل 5 دولارات لأولئك الذين تلقوا التدريب و30 دولار لأولئك الذين حصلوا على التدريب والتمرين الداخلي. وهكذا، يبدو أن توفير خبرة عملية بالإضافة إلى التدريب يؤدي إلى تحسن كبير في الأثر
ما الجرعة الأكثر فعّالية للتدخل؟	الكثير ليس دائما جيدا؛ إن إيجاد التوازن الصحيح بين مقدار الخدمة التي سيتم توفيرها مهم لتحقيق أقصى تأثير من جهة وتقليل التكاليف من ناحية أخرى.	• العديد من مجموعات المعالجة (تحصل الأولى على التصميم "أ"، والثانية على التصميم "ب"، وهكذا) • يجب أن يكون عدد الأشخاص المشمولين بالتقييم كبيراً بما يكفي لإنشاء أكثر من مجموعة معالجة واحدة بالإضافة إلى مجموعة مقارنة	متوسط الزيادة في الدخل هو 0 دولار لأولئك الذين تلقوا تدريباً لمدة شهر واحد، و 20 دولاراً لأولئك الذين تدريباً لمدة 3 أشهر، و 20 دولاراً لأولئك الذين تدريباً لمدة 6 أشهر. على الرغم من أن شهر واحد من التدريب لم يكن كافيا، إلا أن 6 أشهر من التدريب لم تكن لها فائدة إضافية مقارنة مع 3 أشهر من التدريب. يبدو أن الطول المثالي للتدريب حوالي 3 أشهر
لماذا (لا) ينجح التدخل؟ لماذا نجح فقط لجزء من السكان المستهدفين / عندما كان التدخل لمدة محددة؟	جنبا إلى جنب مع تقييم الأثر نفسه، يتعين فهم كيف ولماذا حدثت النتيجة على النحو المرصود	• يتم تثليث (الإقامة على ثلاثة أركان) كل من البيانات الكمية والنوعية مثالياً لإقامة علاقات سببية معقولة. على سبيل المثال، المقابلات المتعمقة مع المشاركين في التدريب والمدربين وأصحاب العمل	يقدر أصحاب العمل مهارات محددة يستطيع المشاركون اكتسابها في غضون ثلاثة أشهر. البقاء لفترة أطول في التدريب لم يكن له قيمة إضافية على مهارات الموظفين. لم يرغب أصحاب العمل في التحفيز على بقاء مدة تدريب لفترة أطول لأنهم سيفقدون وقت عمل الموظف أثناء التدريب. لم يؤد التدريب سوى إلى زيادة دخل الذكور لأن أصحاب العمل يميلون إلى إسناد مهام مختلفة للفتيات، حيث تكون المهارات المنقولة في التدريب أقل طلباً

المربع 5-9: تتبع العملية

يتضمن تتبع العملية التحليل المتعمق للأحداث المختلفة التي تربط التدخل بواحد أو أكثر من النتائج الوسيطة أو النهائية، وعلاقاتها السببية. تهدف أساليب التتبع في كثير من الأحيان (ولكن ليس دائمًا) إلى تطوير واختبار الآليات النظرية، التي يمكن تعميمها لتغطية التدخلات والسياقات الأخرى. باختصار، يتم تطبيق تتبع العملية على النحو التالي:

1. تطوير آلية سببية افتراضية لكيفية حدوث التغيير

كخطوة أولى، علينا إنشاء سرد للعملية المراد تقييمها. والذي يمكن أن يتمثل في نظرية التغيير للمشروع، بما في ذلك الأشخاص والأنشطة المرتبطة به. ومن المهم بيان العملية في عناصرها الصغرى الفردية، التي ينبغي أن تكون ضرورية لنجاحها وأن تكون قابلة للقياس.

على سبيل المثال: "يقوم المعلمون بتدريب الشباب العاطلين عن العمل على المهارات"؛ "يحضر الطلاب التدريب على المهارات"؛ "يكتسب الطلاب معارف ومهارات جديدة حول كيفية وأوقات البحث عن عمل"؛ "يبحث الطلاب عن العمل بشكل أكثر وأكثر كفاءة"؛ "تزداد احتمالية عثور الطلاب على عمل".

من أجل تقديم ادعاءات معقولة بشأن الروابط السببية بين الأجزاء المختلفة لهذه الآلية، من الضروري تحديد التفسيرات البديلة الممكنة لحدوث كل عنصر فردي والبحث عن أدلة لتأكيد أو استبعاد تلك التفسيرات. على سبيل المثال: "اكتسب الطلاب المعارف حول كيفية وأوقات البحث عن عمل بشكل مستقل"؛ "يجد الطلاب وظائف بسبب التحسن في وضع سوق العمل المحلي".

2. تحديد الأدلة المطلوبة وجمعها

بعد تحديد الآلية، أو نظريتنا، نحتاج إلى تحديد الأدلة التجريبية المطلوبة لتحليل كل رابط في السلسلة السببية. وينطبق ذلك على كل من آليتنا وعلى الفرضيات البديلة المنافسة. وبالتالي، سيتم جمع الأدلة المحددة مسبقاً من خلال جمع البيانات الأولية و/أو الثانوية. يمكن أن تتمثل مصادر هذا الدليل في المقابلات مع أصحاب المصلحة، ووثائق البرامج، وبيانات المسوح، ومحاضر الاجتماعات، والإحصاءات، من جملة أمور أخرى. وينبغي جمع الأدلة بطريقة تمكّن إما من تأكيد أو دحض الفرضيات المتنافسة المختلفة. ومن الممارسات الجيدة تثليث الأساليب، بمعنى استخدام أساليب مختلفة لتقييم نفس العنصر من زوايا مختلفة.

3. تقييم الأدلة واستخلاص النتائج

بعدئذ، يتم فحص الأدلة المجمعة في إجراء مماثل للإجراء المستخدم في المحاكمة الجنائية. ونهدف من استخدام تتبع العملية إلى إنشاء دعوى تقدم دليلاً كافياً للافتراض بشكل معقول بأن حدوث كل عنصر من عناصر الآلية يعزى إلى عنصر آخر، وأنهم تسببوا معاً في حدوث نتائج معينة.

هناك اختبارات مختلفة لتقييم قوة الأدلة لكل فرضية. فعلى سبيل المثال، يشير اختبار "الدليل الدامغ" إلى دليل مقنع يشير بشكل مباشر إلى الآلية محل الاختبار. لذلك، على سبيل المثال، يمكن لإفادة من أحد المشاركين في التدريب على المهارات مثل: "بفضل الأشياء التي تعلمتها في التدريب، تعززت ثقتي أكبر بكثير للتقدم إلى الوظائف وإرسال المزيد من طلبات العمل أكثر من ذي قبل" أن تمحنا الثقة إلى حد كبير أن هذا المشارك لم يزد من سلوكيات من البحث عن عمل - والذي يمكن أن يكون متغيراً لنتيجة متوسطة يتم قياسه كماً - لأسباب أخرى.[6]

عند تقييم الأدلة الخاصة بالفرضيات المتنافسة المختلفة، يتعين أن نضع في الاعتبار أن قوة الدليل الشامل لأي آلية معينة دائماً ما تكون مماثلة فقط لأضعف الأدلة عن ارتباط فردي واحد. وأخيراً، واستناداً إلى الاستنتاجات من هذا التمرين، سيتم تأكيد أو استبعاد الآلية المفترضة، بالإضافة إلى الفرضيات البديلة.

[6] للمزيد من الاختبارات وكيفية تطبيقها، انظر بينيت، 2010.

يمكن للعديد من استراتيجيات التقييم المذكورة أعلاه أن تسهم في تقييم أي تدخل محدد بدلاً من كونها بدائل لتقييم الأثر الكمي. لذلك، يمكننا استخدام نهج الطرق المختلطة من الجمع بين نقاط القوة وتعويض نقاط الضعف في كل من أدوات التقييم النوعي والكمي، ويسمح لنا بتعزيز قوة تصميم التقييم بشكل عام.

إن استخدام تصميم الطرق المختلطة لتقييم الأثر في الممارسة العملية يعني جمع كل من البيانات النوعية، على سبيل المثال من خلال الزيارات الميدانية أو المقابلات الرئيسية أو مناقشات مجموعات التركيز، والبيانات الكمية، مع الاعتماد على سبيل المثال على البيانات الإدارية أو الاستبيانات أو مصادر البيانات الثانوية، مثل الاستبيانات الأسرية (انظر أيضا الجدول 3-5 في الملاحظة 3).

ترتبط تصاميم تقييم الطرق المختلطة أيضاً ارتباطاً وثيقاً بتقييمات الأثر المستندة إلى النظرية وتوجهها. ووفقاً لملاحظات وايت وفيليبس (2012)، تهدف تقييمات الأثر المستندة إلى النظرية إلى إنشاء روابط سببية "من خلال جمع الأدلة للتحقق من صحة، أو إبطال، أو مراجعة التفسيرات المفترضة، بهدف توثيق الروابط في السلسلة السببية الفعلية". وغالباً ما يسعون إلى الجمع بين كل الأدلة الكمية والنوعية المتاحة لإثبات تأثير التدخل على المشاركين بما لا يدع مجالاً للشك. يقدم المربع 5-9 منهجية تتبع العملية كمنهجية تقييم التأثير مستندة إلى النظرية.

النقاط الرئيسية

1. **تجيب تقييمات الأثر عن الأسئلة السببية لتحديد إن كان التدخل قد تسبب في حدوث التغيير المرصود ونطاقه.** يتطلب فهم التأثير منّا عزل تأثيرات التدخل من العوامل الأخرى التي تؤثر على النتائج المستفيدة.

2. **يتطلب تقييم أثار التدخلات تقدير الوضع المغاير للواقع، وهو، ماذا كان سيحدث للمستفيدين في حالة عدم تنفيذ التدخل.** ولهذه الغاية، تعتمد معظم تصاميم تقييم الأثر الكمية على وجود مجموعة مقارنة تشترك في أكبر قدر ممكن من الخصائص مع المستفيدين.

3. **تشمل تصاميم تقييم الأثر القائمة على الملاحظة الفرق في الاختلافات وأساليب المطابقة. ويمكن تطبيقها في نطاق واسع من السياقات وعلى أساس مصادر البيانات الثانوية، ومع هذا لا يمكن لهذه الطرق تقييم أثر بعض التدخلات بشكل موثوق.** يمكن أن تنتج التصاميم التجريبية التي تعتمد على درجة معينة من التوزيع العشوائي تقديرات ذات مصداقية عالية للأثر ولكنها قد تنطوي على تكلفة عالية وصعوبات في التنفيذ لبعض التدخلات.

4. **لتحقيق أقصى قدر من المعرفة عن "أسباب" نجاح التدخلات، أو عدم نجاحها، يُوصى بشدة باستخدام نُهج الطرق المختلطة،** التي تعتمد على البيانات الكمية والنوعية والاستفادة من منهجيات التحليل المختلفة.

الموارد الرئيسية

- Gertler, P.J.; Martinez, S.; Premand, P.; Rawlings, L.B.; Vermeersch, C.M., 2016. ***Impact evaluation in practice***, Second Edition (Washington DC, Inter-American Development Bank and World Bank).

- Card, D.; Ibarrarán, P.; Villa, J.M. 2011. ***Building in an evaluation component for active labor market programs: A practitioner's guide***, IZA Discussion Paper No. 6085 (Bonn, Institute for the Study of Labour (IZA)).

- Khandker, S.R.; Koolwal, G.B.; Samad, H.A. 2010. ***Handbook on impact evaluation: Quantitative methods and practices*** (Washington, DC, International Bank for Reconstruction and Development (IBRD) and World Bank).

- Duflo, E.; Glennerster, R.; Kremer, M. 2006. ***Using randomization in development economics research: A toolkit,*** BREAD Working Paper No. 136 (Bureau for Research and Economic Analysis of Development).

المراجع

- Bamberger, M.; Rao, V.; Woolcock, M. 2010. *Using mixed methods in monitoring and evaluation: Experiences from international development*, Policy Research Working Paper No. 5245 (Washington, DC, World Bank).
- Bennett, A. 2010. "Process tracing and causal inference", in H. Brady and D. Collier (eds.) *Rethinking social inquiry: Diverse tools, shared standards* (2nd edn) (Lanham, MD, Rowman & Littlefield), pp. 207–220.
- Card, D.; Kluve, J.; Weber, A. 2009. *Active labour market policy evaluations: A meta-analysis*, IZA Discussion Paper No. 4002 (Bonn, Institute for the Study of Labour (IZA)).
- Card, D.; Ibarrarán, P.; Villa, J.M. 2011. *Building in an evaluation component for active labor market programs: A practitioner's guide*, IZA Discussion Paper No. 6085 (Bonn, Institute for the Study of Labour (IZA)).
- Duflo, E.; Glennerster, R.; Kremer, M. 2006. *Using randomization in development economics research: A toolkit*, BREAD Working Paper No. 136 (Bureau for Research and Economic Analysis of Development).
- Escudero, V. 2016. *Workfare programmes and their impact on the labour market: Effectiveness of Construyendo Perú*, Research Department Working Paper No. 12 (Geneva, ILO).
- Fiala, N. 2015. *Access to finance and enterprise growth: Evidence from an experiment in Uganda*, Employment Working Paper No. 190 (Geneva, ILO).
- Gertler, P.J.; Martinez, S.; Premand, P.; Rawlings, L.B.; Vermeersch, C.M., 2016. *Impact evaluation in practice*, Second Edition (Washington DC, Inter-American Development Bank and World Bank).
- International Labour Organization (ILO). 2017. *Empowering young women through business and vocational training: Evidence from rural Upper Egypt*, Taqeem Impact Brief Series, Issue 10 (Geneva).
- Khandker, S.R.; Koolwal, G.B.; Samad, H.A. 2010. *Handbook on impact evaluation: Quantitative methods and practices* (Washington, DC, International Bank for Reconstruction and Development (IBRD) and World Bank).
- Leeuw, F.; Vaessen, J. 2009. *Impact evaluations and development: NONIE guidance on impact evaluation* (Washington, DC, Network of Networks on Impact Evaluation).
- Schomburg, H. 2016. *Carrying out tracer studies: Guide to anticipating and matching skills and jobs*, Vol. 6 (Luxembourg, ETF, CEDEFOP, ILO).
- White, H.; Phillips, D. 2012. *Addressing attribution of cause and effect in small n impact evaluations: Towards an integrated framework*, Working Paper 15 (New Delhi, International Initiative for Impact Evaluation).

دراسة حالة:

تقييم نمو المشاريع الصغيرة في المناطق الريفية عبر طرق التقييم المختلفة

إخلاء مسؤولية: دراسة الحالة الماثلة تخيلية، وجاءت جميع المعلومات الواردة بها لأغراض تعليمية فقط."

أهداف التعلم المرجوة

بنهاية دراسة الحالة هذه، سيتمكن القارئ من إثبات القدرة على ما يلي:

- تحديد طرق تقييم التأثير دون إخباره بطريقة استخدام محددة
- استكشاف مشكلة إنتاج تقديرات للتأثيرات السببية للبرامج الإنمائية، والطرق المختلفة لتقدير التأثيرات باستخدام تصميمات مجموعة المقارنة
- تطوير فهم بديهي لموعد وكيفية تأثير أساليب تقييم الأثر على نتائج متحيّزة من خلال التعرف على مفهوم التحيّز في الاختيار ومدى كفاءة تصاميم مجموعة المقارنة بقدر قدرتها على إزالة التحيّز في الاختيار

مقدمة وسياق دراسة الحالة

تمثل المشاريع متناهية الصغر ركيزة حيوية في المناطق الريفية حيث لا تتوفر سوى خيارات محدودة للعمل في القطاع المنظم، سواء من أجل توفير فرص العمل في القطاع غير المنظم وضمان الأمن الاقتصادي الأسري لأصحاب المشاريع. ومع ذلك، بمجرد إقامة المشروع، هناك عدد من التحديات التي تواجه النمو.

ما الذي يمكن فعله للمساعدة في تطوير الأعمال الريفية؟ يختبر برنامج التدريب من أجل التمكين الاقتصادي الريفي (TREE) بعض هذه القيود لتحديد نوع الخدمات المالية والتدريب الذي يؤثر على تنمية المشاريع، والمستفيدين وأسباب ذلك.

أجرت منظمة العمل الدولية (ILO) التدريب موضوع الاختبار هنا باستخدام منهجية التدريب من أجل التمكين الاقتصادي الريفي، وهو نهج إنمائي يكفل اكتساب النساء والرجال الذين يعيشون في دائرة الفقر للمهارات والمعارف التي يحتاجونها لتحسين دخولهم والقيام بدور أكثر نشاطاً في تشكيل مجتمعاتهم. وعلاوة على ذلك، أصدرت منظمة التمويل الأصغر المحلية القروض إلى الأفراد بداية من 200 دولار أمريكي بأسعار فائدة سنوية مخفضة بنسبة 20 في المائة (أقل من معيار الـ 25 في المائة).

تركز دراسة الحالة الماثلة على 400 من أصحاب المشاريع متناهية الصغر في المناطق الريفية الذين عُرضت عليهم فرصة المشاركة في برنامج التدريب على المهارات والحصول على القروض. وفي المجمل، شارك 144 من إجمالي أصحاب المشاريع وعددهم 400 في التدريب وحصلوا على القروض.

مقارنة طرق تقييم الأثر المختلفة: هل نجح برنامج التدريب من أجل التمكين الاقتصادي الريفي؟

هل نجح برنامج التدريب من أجل التمكين الاقتصادي الريفي؟ هل حسّن البرنامج من أرباح المشاريع؟ ما هي المتطلبات التي تمكننا من قياس إن كان البرنامج ناجحاً أو إن كان له تأثير؟

بصفة عامة، يرتبط السؤال عن نجاح البرنامج بالسؤال عن تمكن البرنامج من تحقيق هدفه المتمثل في تغيير نتائج معينة للمشاركين، والتأكد من أن هذه التغييرات لا تُعزى إلى بعض العوامل الأخرى. علينا أن نثبت بالتزامن أنه في حالة عدم تنفيذ البرنامج، لم تكن التغيرات الملحوظة لتحدث (أو كانت ستكون مختلفة). لكن كيف نعرف ما الذي كان سيحدث؟ يتطلب قياس ما كان سيحدث في حالة عدم تنفيذ البرنامج الدخول إلى عالم وهمي لم يتم فيه عرض البرنامج على هؤلاء المشاركين. يشار إلى نتائج نفس المشاركين في هذا العالم الوهمي على أنه الوضع المغاير للواقع. ونظراً لأننا لا نستطيع أن نلاحظ الوضع المغاير للواقع الحقيقي، فإن أفضل ما يمكننا القيام به هو تقديره عن طريق تقليده.

يتمثل التحدي الرئيسي لتقييم أثر البرنامج في بناء أو محاكاة لوضع مغاير للواقع، وذلك في العادة عن طريق اختيار مجموعة من الأشخاص يتشابهون مع المشاركين في البرنامج بأكبر قدر ممكن، ولكنهم لم يشاركوا في البرنامج. تسمى هذه المجموعة مجموعة المقارنة، التي لا تختلف بشكل مثالي عن مجموعة المستفيدين إلا من حيث عدم المشاركة في البرنامج. ثم نقوم بتقدير "الأثر" على أنه الفرق الملحوظ في نهاية البرنامج بين نتائج مجموعة المقارنة ونتائج المشاركين في البرنامج.

وجدير بالذكر أن دقة تقدير الأثر ترتبط ارتباطاً وثيقاً بنجاح مجموعة المقارنة في محاكاة الوضع المغاير للواقع. لذلك، تمثل الطريقة المستخدمة لتحديد مجموعة المقارنة قراراً رئيسياً في تصميم أي تقييم للأثر.

هذا يعيدنا إلى أسئلتنا: هل نجح المشروع؟ ما تأثيره على النتيجة التي يجري تقييمها، وهي أرباح المشاريع؟

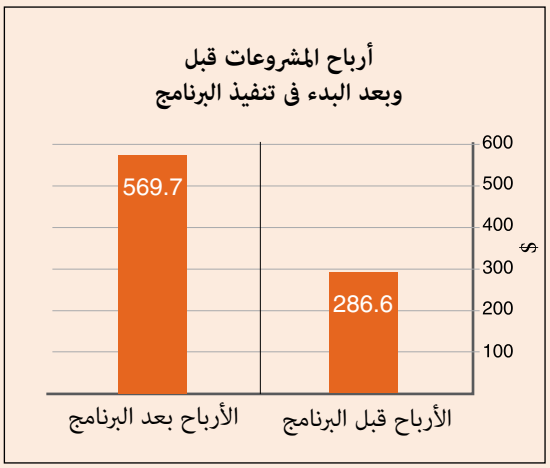
في حالتنا، يهدف هذا البرنامج في المقام الأول إلى "تحسين نمو المشاريع"، وتعد الأرباح (المقاسة بالدولار الأمريكي) مؤشراً رئيسياً على النتائج. لذا، عندما نسأل هل نجح هذا المشروع، فإننا نتساءل عما إذا حسّن أرباح المشاريع. الأثر هو الفرق بين الأرباح بعد تعرض المشاريع للتدخل مع تحديد أرباحها في حالة عدم تنفيذ التدخل.

ما مجموعات المقارنة وطرق تقييم الأثر التي يمكننا استخدامها؟ يوضح الخبراء (الوهميون) التالون طرقاً مختلفة لتقييم الأثر. يعرض الجدول 2-5، في نهاية دراسة الحالة، لمحة عامة عن أساليب التقييم المختلفة للإحاطة بها.

الجزء الأول: خبر صحفي: يساعد برنامج "التدريب من أجل التمكين الاقتصادي الريفي" المشاريع على النمو

يحتفل برنامج التدريب من أجل التمكين الاقتصادي الريفي بنجاح البرنامج، إذ حقق تقدماً كبيراً في سبيل هدفه المتمثل في مساعدة الشركات على النمو من خلال تقديم القروض والتدريب على المهارات. ويثبت البرنامج أن توفير التدريب على المهارات لأصحاب الأعمال، بالإضافة إلى القروض لتخفيف أعباء رأس المال، يمكن أن يحقق مكاسب كبيرة.

قبيل بدء البرنامج، بلغ متوسط أرباح المشاريع 286 دولار أمريكي. ولكن بعد قضاء بضعة أشهر فقط في البرنامج، تضاعفت أرباح هذه المشاريع!



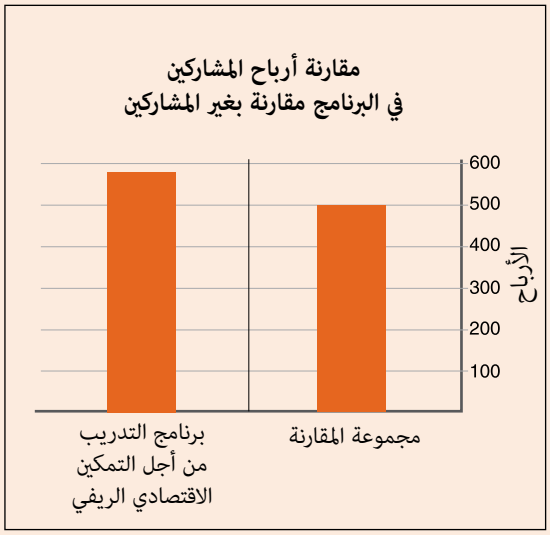
مواضيع للنقاش

1. ما نوع التقييم الذي يتضمنه هذا الخبر؟
2. ما الذي يمثل الوضع المغاير للواقع؟
3. ما التحديات التي تنطوي على هذا النوع من التقييم؟

الجزء الثاني: الرأي: مشروع "التدريب من أجل التمكين الاقتصادي الريفي" لا يصل إلى المستوى المطلوب

يعد برنامج التدريب من أجل التمكين الاقتصادي الريفي واحداً من أكبر أنظمة إدارة التدريب العالمية المستخدمة لدعم المشاريع متناهية الصغر والصغرى الموجودة حالياً في السوق بفضل نطاق وصوله المقدر إلى 6 مليون متدرب وشبكته المتنامية التي تضم أكثر من 17.000 مدرب و200 مدرب أستاذ في 2.500 مؤسسة شريكة. ولكن هل تتضاعف أرباح المشاريع في الحجم، على النحو المبين في المثال الأول؟ تشير الأدلة الأخيرة إلى خلاف ذلك.

◀ تم الاستعانة بفريق مستقل من المقيّمين للتحقق من صحة هذه النتائج. قارن الفريق أرباح المؤسسات التي خضعت للتدريب بأرباح المشاريع الأخرى في القرى المجاورة. ووجد الباحثون أن الشركات التي خضعت للتدريب من أجل التمكين الاقتصادي الريفي تنمو أرباحها بهامش بسيط قدره 64 دولار أمريكي، وليس 286 دولار أمريكي وفقاً للتقديرات الأصلية، بما يمثل زيادة في الأرباح نسبتها 12 في المائة فقط بعد 6 أشهر من برنامج التدريب من أجل التمكين الاقتصادي الريفي المقترن بالقروض. ويبدو أنه قد بولغ في تقدير حجم الإيرادات إلى حد كبير، وأن تأكيدات منظمة العمل الدولية بشأن نجاحات البرنامج غير صحيحة.



مواضيع للنقاش

1. ما نوع التقييم الذي يتضمنه هذا الرأي؟
2. ما الذي يمثل الوضع المغاير للواقع؟
3. ما التحديات التي تنطوي على هذا النوع من التقييم؟

الجزء الثالث: رسالة إلى المحرر: ينبغي على المقيّمين المستقلين النظر في التقييم بشكل عادل ودقيق

تداولت الجريدة عدة تقارير غير نزيهة بشأن البرامج التي تنفذها منظمة العمل الدولية، إذ تضمن مقال حديث أعدّه مقيّم مستقل ادعاءات تفيد بأن برنامج التدريب من أجل التمكين الاقتصادي الريفي لا يساعد الشركات في الواقع على النمو. ومع ذلك، يستخدم التحليل الوارد في المقال مقياساً غير صحيح لقياس الأثر، إذ يقارن بين أرباح المشاريع التي خضعت للتدريب وأرباح المشاريع الأخرى في نفس القرية، دون مراعاة حقيقة استهداف البرنامج للمشاريع ذات الأرباح المنخفضة بشكل خاص منذ بداية البرنامج. إذا قام البرنامج بمجرد استهداف المشاريع الكبرى، ومقارنتها بنظيراتها الصغرى، يمكن عندئذ ادعاء نجاح البرنامج دون إجراء أي جلسة تدريب واحدة أو تقديم قرض واحد. لكن لم يفعل البرنامج ذلك. وفي واقع الأمر، لا يتوقع البرنامج أن تتجاوز المشاريع الصغرى المشاريع الكبرى في القرية، إذ يقتصر هدف البرنامج على تحسين الوضع الحالي.

لذلك، يجب أن يكون المؤشر التحسن في الأرباح - وليس مستوى الربح النهائي. عندما كررنا التحليل باستخدام مقياس النتيجة الأكثر ملاءمة، تحسنت المشاريع التي شاركت في البرنامج بمعدل مضاعف عن المشاريع غير المشاركة (زيادة في الأرباح بمبلغ 283 دولار أمريكي مقارنة بـ 162 دولار أمريكي). لو اعتقد المقيّمون المستقلون أنهم ينظرون إلى النتيجة الأكثر ملاءمة، لكانوا سيقرون بالنجاح المذهل للبرنامج. ربما يجب عليهم التسجيل بأنفسهم في بعض دورات التدريب التي يوفرها البرنامج..

مواضيع للنقاش

1. ما نوع التقييم الذي يتضمنه هذا الخطاب؟
2. ما الذي يمثل الوضع المغاير للواقع؟
3. ما التحديات التي تنطوي على هذا النوع من التقييم؟

الجزء الرابع: تصميم التقييم الخاص بك لتقييم أثر برنامج التدريب من أجل التمكين الاقتصادي الريفي

كما ذكرنا سالفاً في دراسة الحالة الماثلة، هناك تحديات وتحفظات بشأن جميع طرق التقييم الثلاثة المذكورة أعلاه. لقد حان دورك الآن لتصميم تقييم لأثر برنامج التدريب من أجل التمكين الاقتصادي الريفي، بافتراض أن البرنامج لم يتم تنفيذه بعد.

بادئ ذي بدء، افترض أن فريقك البحثي قد قام بإجراء مسح لآلاف المشاريع الصغيرة من خلال تعداد الأعمال التجارية، واختار 1600 من أصحاب المشاريع لإدراجهم في التقييم. وقد أعرب جميع أصحاب المشاريع هؤلاء عن اهتمامهم بتلقي تدريب منظمة العمل الدولية والمشاركة في برنامج القروض. ومع ذلك، وبسبب محدودية الموارد، يخبرك مدير المشروع أن الحد الأقصى للمشاركين في برنامج التدريب والقروض يبلغ 800 شركة كحد أقصى.

نفترض أن النتيجة الرئيسية تظل "أرباح المشاريع".

1. كيف تصمم التقييم؟ وتحديداً، كيف تختار مجموعة المقارنة؟
2. متى يقوم فريق البحث الخاص بك بجمع البيانات ومن أي المشاريع (كافة المشاريع الـ 1600 أم من مجموعة فرعية فقط)؟
3. لماذا تعتقد أن هذه الطريقة لتقييم الأثر موثوق بها وتتغلب على بعض أو كافة أوجه القصور في الطرق الثلاث التي تم مناقشتها؟

978-92-2-630795-7

International Labour Organization
Youth Employment Programme
4 route des Morillons
CH-1211 Genève 22
Switzerland

youth@ilo.org
www.ilo.org/yep
www.ilo.org/taqeem