International Initiative for Impact Evaluation
Working Paper 4

المبادرة الدولية لتقييم الأثر

ورقة عمل 4

تصميم دراسات تقييم الأثر:
وجهات نظر

روبرت شامبرس ، دين كارلان ، مارتن رافاليون ، و باتريشيا روجرس

يونيو 2009

المبادرة الدولية لتقييم الأثر

ورقة عمل 4

تصميم دراسات تقييم الأثر:
وجهات نظر

روبرت شامبرس ، دين كارلان ، مارتن رافاليون ، و باتريشيا روجرس
يونيو 2009

مبادرة 3ie :

المبادرة الدولية لتقييم الأثر (المردود) منظمة دولية غير حكومية توفر الدعم المادي لتشجيع السياسات والبرامج التنموية. وهى تعتبر المنظمة الدولية الرائدة في توفير الدعم وإنتاج الأدلة ذات الجودة العالية على : ماذا يعمل وكيف يعمل ولما يعمل وبأي تكلفة بالنسبة الى أثر ومردود التنمية. ونحن نعتقد أن الأدلة الأفضل والمبينة على السياسات الجيدة سوف تجعل التنمية أكثر فعالية، وتحسن من حياة المجتمعات.

أوراق عمل المبادرة :

تركز هذه الأوراق على القضايا الحالية والمناقشات والتحديات التي تواجه مُعدي السياسات والعاملون في مجال تقييم الأثر، والقائمون على المراجعة المنهجية في هذا المجال. تعتمد أوراق العمل المرتبطة بالسياسات على النتائج ذات الصلة الناتجة عن عمليات تقييم الأثر والمراجعات المنهجية الممولة من المبادرة، ونتائج عن تقييمات دقيقة موثوق بها، ومراجعات تُقدم رؤى، أو تحليل جديد، أو نتائج وتوصيات. وتعتمد الأوراق التي تُركز على الطرق على مصادر مشابهة مساعدة للفهم، والتصميم، والإستخدام المتقدم للتقييم، والمراجعة الدقيقة والملائمة.

حول هذه الورقة :

نشرت هذه الورقة لأول مرة باللغة الإنجليزية في عام 2009. ولم يحدث اي تعديل أو إضافات على الإصدار الأصلي قبل ترجمته الى اللغة العربية. وقد قام بترجمة هذه الورقة الدكتور / محمد مصطفى الفولي والسيد / تامر محمد الفولي، وتقوم المبادرة بإتاحة هذه الورقة بدون اي مراجعة للترجمة، وأن المحتوى بالكامل، واي إضافات أو حذف هى مسئولية المؤلفين ولا تعبر عن رأي المبادرة، أو المانحين، أو مجلس المفوضين. وأي أخطاء أو حذف حدث خلال الترجمة – تعد مسئولية المترجمين.

يمكن الإشارة الى هذه الورقة كما يلي :

Suggested citation: Chambers, R, Karlan, D, Ravallion, M and Rogers, P, 2009. *Designing impact evaluations: different perspectives, 3ie Working Paper 4*. New Delhi: International Initiative for Impact Evaluation (3ie)

المحررين الحصرين لسلسلة أوراق العمل : بيرلي ليتش - عمانويل جيمينيز

مدير التحرير: كانيكا جها

المترجمون : د. محمد مصطفى الفولي / تامر محمد الفولي

صورة الغلاف : ادوين هوفمان ، البنك الدولي

© International Initiative for Impact evaluation (3ie), 2015

المحتويات:

مقدمة

هوارد وايت ، المدير التنفيذي ، المبادرة الدولية لتقييم الأثر

وهكذا يصبح الفقراء أكثر أهمية: إستخدام الأساليب التشاركية من أجل تقييم الأثر

روبرت تشامبرز ، معهد الدراسات التنموية ، جامعة ساسكس

أفكار حول التجارب العشوائية لتقييم التنمية: عرض في مؤتمر القاهرة للتقييم

دين كارلان ، جامعة يال و مبادرة ابتكارات من أجل الفقر / كلية جميل

تقييم

مارتن رافاليون ، البنك الدولي

تجانس تصميم دراسات تقييم الأثر مع طبيعة المبادرة والهدف من التقييم

باتريشيا روجرس ، تعاون من أجل بحوث واستشارات وتعلم في التقييم ، معهد ملبورن الملكي للتكنولوجيا

مقدمة

وصلت المناقشات بشأن اساليب تصميم دراسات تقييم الأثر في السنوات الأخيرة على ما يبدو إلى طريق مسدود. وكان من أهداف المؤتمر الدولي "وجهات نظر حول دراسات تقييم الأثر"، 29 مارس – 2 أبريل ، القاهرة، والذي تم تنظيمه من قبل المبادرة الدولية لتقييم الأثر، شبكة شبكات تقييم الأثر، رابطة التقييم الأفريقية واليونيسيف، هو جمع الأصوات المختلفة والعمل على توافق الآراء. وفي اطار هذا النهج، وفي جلسة رئيسية عامة، تم طرح سؤال لخبراء ذوي وجهات نظر مختلفة عن "كيفية اعداد دراسات تقييم لثلاثة تدخلات : التحويلات النقدية المشروطة، مشروع البنية التحتية، برنامج لمكافحة الفساد". كان الدافع للجلسة هو ان المناقشات تتعثر عندما تظل على المستوى المفاهيمي / النظري، ولكن يمكن أن يتحقق قدر أكبر قدر من التوافق عند الإنتقال إلى تفاصيل تصميم دراسة تقييم معينة. ويسرني جدا أن العارضين الأربع وافقوا على عرض آرائهم كتابة حتى يمكن نشرها على نطاق أوسع.

شكر لهوج وادينجتون و ريزوانا صديقي للمساعدة في اعداد هذه المجموعة

هوارد وايت

المدير التنفيذي، المبادرة الدولية لتقييم الأثر

وهكذا يصبح الفقراء أكثر أهمية: إستخدام الأساليب التشاركية من أجل تقييم الأثر

روبرت تشامبرز، معهد دراسات التنمية، جامعة ساسكس، المملكة المتحدة[1]

ملخص

إن نقطة البداية لاجراء تقييم هو أن نسأل لماذا يتم اجراء هذا التقييم؟ ومن المستفيد من ذلك؟ وماهو أثر التقييم؟ وكيف يتم ذلك؟ إن المناهج والأساليب التشاركية تتطابق على نموذج تعددى وتطورى ومكرر. فهى تشمل تحليل أصحاب المصلحة وراوية القصص الفردية، واعداد الخرائط الإجتماعية التشاركية، والإرتباط السببى، وعمل رسم بيانى عن الإتجاه والتغيير، وتقييم وتسجيل الدرجات والعصف الذهنى حول نقاط القوة ونقاط الضعف فى البرنامج. تعد الأساليب التشاركية المصممة بشكل جيد والميسرة دقيقة جدا، إلى جانب تقديمها رؤى نوعية، حيث يمكنها أن تُعد الذى لا يعد، وتجرى إحصائيات عن الأبعاد ذات الصلة التى من ناحية أخرى قد يتم التغاضى عنها أو 'ينظر إليها على أنها نوعية بحتة. فهذه الأساليب تفتح المجال لسماع أصوات أولئك الأكثر تأثرا بمشروع مما يبدو غير ممكن بإستخدام الأساليب التقليدية، ويستطيع أن يجعل واقع وتجارب الفقراء تلقى أهمية أكثر.

ملاحظات تمهيدية

على الرغم من أننى كنت سعيدا لدعوتهم لي، لكن عدم الكفاءة النسبية جعلتنى مترددا فى أن أكون فى هذه اللجنة. قام غالبية الحاضرين فى هذا المؤتمر بتصميم دراسات تقييم (فقد طلبنا من كل من قام بعمل تصميم لتقييم أن يرفع يده، وقام أكثر من نصف الحاضرين برفع يده). أنا قمت بعمل تصميم واحد فقط: وكان ذلك فى سنة 1970 وكان كارثيا. وكنت مترددا أيضا لأننا نعيش فى زمن الإبتكارات الهائلة فى المنهجيات التشاركية، بمافيها المتابعة والتقييم. والبرامج الثلاثة التى تم إختيارها تلائم هذه المنهجيات التشاركية بشكل أقل من البرامج المعنية - يمكنكم القول - بالتنمية المجتمعية أو الزراعة أو إدارة الموارد الطبيعية. أشعر بالحرج أيضا، مثل

[1] البريد الالكتروني: robertc@ids.ac.uk

المتحدثين الآخرين، أن أكون شخص آخر من الشمال. قال أحد الأفارقة فى إحدى الندوات الصغرى: "أى شئ يأتى من الشمال هو الحقيقة".

دعونا نعترف أن أكثر 'الحقيقة' – الإبداع والإبتكار – بأن المنهجيات التشاركية أتت وتأتى من الجنوب، ومن آسيا، ومن أمريكا اللاتينية وخاصة من إفريقيا.

مع ذلك، تشجعت بسبب تأكيد سالى جاريبا فى ملاحظاته الإفتتاحية، حول 'تمكين المجتمعات' ومناشدة إرما مانونكورت بأن 'نفتح الباب لمشاركة وتمكين الناس'.

فى الحالات الثلاثة جميعها التى يتم تناولها هنا – تحويل نقدى مشروط (CCT) وتنمية البنية التحتية ولجنة مكافحة الفساد – قد أناقش منهج كان تعدديا وتطوريا ومتكررا. قد يتم إستخدام أساليب مختلطة. قد تكون نقطة البداية هى السؤال عن الإقتصاد السياسى للتقييم: من قد يكسب؟ من قد يخسر؟ وكيف؟، وخصوصا، كيف كان المقصود والمتوقع هو أن هذه الإكتشافات قد تصنع فرقا. هذا قد يتطلب بشكل جيد ورشة عمل عصف ذهنى مع الموظفين من وكالة التمويل. إن كانوا غير مستعدين أو غير قادرين أن يجدوا وقتا لهذا أو لإلقاء الضوء على هذه الأسئلة أتمنى أن تكون لدى الشجاعة والموارد لرفض هذه المهمة. قد أتفاوض بشأن مذكرة التفاهم (MOU) لتشمل خطوات أخرى. إحدى تلك الخطوات قد تكون تحليل أصحاب المصلحة والمفاوضات لإشراك أصحاب المصلحة وذوي الصلة فى هذه العملية.

سأحصر إقتراحاتى بشكل كبير على الأساليب التشاركية التى قد يتم إستخدامها، إما بدلا من أو مكملة لأشياء أخرى. بالنسبة لهذه الأشياء، ستكون هناك أولوية للبحث عن ميسرين مبتكرين ذوي جودة. قد أتفاوض بشأن الوقت (ربما اطلب مدة أطول مما كان يتوقع الممولين) بالنسبة لورش العمل، والعمل الميدانى لتطوير وعمل إختبار تجريبى للمنهج التشاركى، مع تلافى الإغلاق السابق لأوانه، وإستعارة عنوان كتاب إيرين جويجيت ،

'البحث عن مفاجأة'. وعلاوة على كل ذلك، وخلال هذه العملية، قد يكون هناك سؤال 'ثم ماذا؟' عن مسارات التأثير من أجل تقييم الأثر نفسه وفاعلية تكلفتها والقيمة المضافة.

حالة 1: التحويل النقدى المشروط (CCT)

قد يتم إستكشاف تطبيق الأساليب والمناهج التشاركية التالية:

- تخطيط التعداد التشاركى فى المجتمعات المحلية الممثلة. قد يكون من الأهمية أن يتم إنتقاء المجتمعات المحلية بعناية. قد يغطى التخطيط جميع الأشخاص فى كل مجتمع محلى، وبذلك يتم تجنب مشكلات أخذ عينات من المجتمع. قد يتم تيسير تحديد الإناث اللاتى فى سن المدرسة ، والملتحقات بالمدرسة وغير الملتحقات بالمدرسة قبل البرنامج، والملتحقات وغير الملتحقات بالمدارس الأن. قد يعمل ذلك على إيجاد إحصائيات يتم مراجعتها وفقا إلى السجلات الرسمية التى قد تشمل التغييرات فى أداء الفتيات.
- متابعة الحالات الفردية، وتشمل الأشخاص الذين يقيمون بعيدا عن مكان عملهم والغاضبون، إن وجدوا.
- المجموعات البؤرية، والمقابلات شبه الهيكلية حسب الاحتياج مع الفتيات والمدرسين والآباء والإداريين...
- دعوة الفتيات وآخرين لإبتكار وأداء عمل درامى عن حياتهم وتجاربهم قبل وبعد
- تيسير عمل رسم بيانى عن الإرتباط السببى
- تجميع القصص
- عصف ذهنى عن نقاط القوة ونقاط الضعف للبرنامج، وكيف يتم تحسينها.

قد يمكث الميسر/الباحثين بضعة أيام وليال فى المجتمعات المحلية، وقد يجتمعوا من أجل عمل مقارنة للملاحظات. قد يتم إضافة القضايا والأسئلة المتوقعة إلى جدول الأعمال.

حالة 2: تقييم بعد انهاء المشروع لبرنامج قطاع النقل فى دولة بجنوب آسيا

قد يتم البحث عن تطبيق الأساليب والمناهج التشاركية التالية:

- المجموعات البؤرية والمقابلات شبه الهيكلية مع أصحاب الشأن والمتأثرين من الناس، على سبيل المثال: صغاررجال الأعمال من القطاعات الرسمية وغير الرسمية ومقاولى النقل، والسائقين، وأعضاء نقابات العمال فى الميناء، وكبار الموظفين.
- بالنسبة لأثر الطرق الفرعية فى المناطق الريفية، إختيار مجموعة من الظروف والمجتمعات المحلية، التى تحدد الرابحين والخاسرين بشكل غير رسمى. وبعد ذلك، مقابلات شبه هيكلية و/أو مجموعات بؤرية مع المتأثرين من الناس مع تيسير: تحديد أكثر للرابحين والخاسرين، والجداول الزمنية، وعمل رسم بيانى للإتجاه والتغيير قبل وبعد تسجيل درجات المصفوفة، وعمل رسم بيانى للإرتباط السببى للتغيير وتسجيل درجات الإرتباطات.
- التقديرات الرقمية للمكاسب (والخسائر)، وحيثما تقتضى الضرورة، عمل مقابلة لمناقشة الرسم البيانى، للحصول على رؤية أعمق.

يتعارض عمل التقديرات الرقمية مع وجهة النظر القائلة بأن المناهج التشاركية تستطيع فقط أن تنتج بيانات نوعية. مع ذلك، شهد العقد الماضى نموا فى 'أعداد المشاركة'، أى ان المناهج التشاركية هى التى تقوم بإنشاء البيانات الكمية (راجع شامبرز، 2007).

حالة 3: تقييم دعم الجهة المانحة للجنة مكافحة الفساد بدولة أفريقية

فى مجال حساس مثل الفساد، قد يعتبر التقييم نفسه (أكثر من المعتاد) تدخلا، ويتم معاملته على أنه فرصة. قد يتم انعكاس ذلك طوال التقييم لتعزيز الإستفادة وتقليل الضرر. وقد تكون احدى الخطوات المبكرة هي تحديد الأشكال والمستويات (المستوى العالى والمستوى المنخفض) للفساد، والإرتباطات السببية الإفتراضية بين دعم

الجهة المانحة والتغييرات التى قد تحدث. قد تكون مصادر المعلومات الأساسية هى المنظمات غير الحكومية التى تعمل فى المناصرة وحقوق الإنسان والصحفيين.

ومع الذين قاموا بعمل جولات دراسية وأعضاء اللجنة:

- مناقشة تأملية فى مجموعات بؤرية (لكن أنظر الإجتماعات الفردية التالية)

مع الفساد منخفض المستوى، إذا كان هناك أى إرتباط ظاهرى، إبحث عن المصادر المتنوعة للأدلة والرؤى:

- بطاقات تسجيل رضاء المواطنين أو ما يعادلها إن توفرت.
- مجموعات بؤرية للغاضبين من الناس
- مناقشات عارضة فى المقاهى إلخ.

مع الفساد عالى المستوى قد يكون الجزء الرئيسى

- مناقشات خاصة من شخص إلى شخص بدون أخذ ملاحظات. لدعم ذلك، فى أفضل تقليد للتيسير البرئ على طريقة مكيافيللى (أى التيسير بنية التلاعب) قد يتم تقديم إذن من المحكمة للقائمين بالتقييم.

ملاحظات ختامية

تم طرح سؤال عن عَدْ الذى لا يُعدْ. ان الوسائل التشاركية لديها قدرة غير متعارف عليها بشكل كبير لإنشاء الأرقام التى يمكنها، أيضا، أن تكون قابلة للقياس، ويمكن معاملتها مثل أى إحصائيات أخرى. من خلال التحكيم والتقدير والتعبير عن القيم، يقوم الأشخاص بقياس النوعية. يتأخر الإعتراف بإمكانات هذه الأساليب. حيث أنه دائما ماتكون هناك قضايا أخلاقية. الأساليب التشاركية الميسرة بشكل جيد يمكن أن تكون فائزة فى تمكين الأشخاص، وكذلك فى تقديم رؤى يمكن تصديقها والإعتماد عليها لصانعى القرار.

تم إثارة منهجيات السلطة والإقتصاد السياسى كقضايا فى هذا المؤتمر. سمعت عن أحد المخاوف، التى تعبر عن أن إختيارات التدخل قد تتأثر بتعرضها لتقييم الأثر بأساليب تنتسب إليها أشكال معينة من الدقة. هذا شئ لا أستطيع أن أحكم عليه. لكن، على سبيل المثال، إذا أدى هذا المنهج إلى تفضيل البرامج ذات المعطيات البسيطة والمدخلات سريعة التأثير على برامج أخرى أكثر تعقيدا وطويلة المدة وتعددية وتشاركية وممكنة فإن تكاليف الفرصة للفقراء قد تكون عالية.

أخيرا، من المثير للإنتباه فى هذا المؤتمر هو أن نادرا ماورد ذكر لصدارة وقدرات الفقراء. الغرض من تقييم الأثر هو التعلم والتغيير الذى يجعل الحياة أفضل لهم. ولتحقيق ذلك، نحتاج إلى أساليب مختلطة وتعددية. العديد من المناهج والأدوات يمكن أن تكون ويجب أن تكون تقييم الأثر.

مهما كانوا، يجب أن يعترفوا دائما بأنهم هم الذين يعيشون فى فقر، هؤلاء هم الضعفاء، وهؤلاء المهمشين. وهم أفضل الحكام على وهم السلطات الأولى المسؤولة عن حياتهم، وعن سبل معيشتهم وكيف تأثروا. نحن نعرف الأن، كما لم نعرف منذ عقدين، أنهم لديهم قدرات تحليلية أكبر بكثير مما كنا نعتقد. نعرف أنهم ' يستطيعون أن يفعلوها'. لكى نيسر تحليل تمكينهم. نعرف الأن أننا لدينا ثروة من المنهجيات التشاركية. نحتاج إلى أن نستفيد منها بشكل أكبر، وبشكل أفضل. مرارا وتكرارا، تتكرر الوصية: إسألوهم.

المراجع:

Chambers, Robert (200?) 'Who Counts? The Quiet Revolution of Participation and Numbers' *IDS Working Paper 296*. Falmer: Institute of Development Studies.

أفكار حول التجارب العشوائية لتقييم التنمية: عرض في مؤتمر القاهرة للتقييم

دين كارلان[2]
جامعة ييل
إبتكارات للعمل من أجل الفقر
معمل جاميل للعمل من أجل الفقر

خلاصة

'طلب منا أن نناقش طرق منهجية محددة لتقييم ثلاثة من التدخلات الإفتراضية. يستخدم هذا المقال هذا المنتدى لمناقشة ثلاثة مفاهيم خاطئة عن التصميمات العشوائية. أولا، لا أحد يجادل بأن التصميمات العشوائية مناسبة فى جميع البيئات ولجميع الأسئلة. كل فرد يوافق على أن طرح السؤال الصحيح هو الأولوية الأعلى. ثانيا، القرار عن ما الذى يجب قياسه؟ وكيف يتم قياسه؟، وذلك يتم من خلال الأساليب النوعية أو التشاركية إزاء المسح الكمى أو أساليب البيانات الإدارية، وهذا القرار قرار مستقل عن القرار حول هل يتم عمل تصميمات عشوائية؟. ثالثا، يمكن إستخدام التصميمات العشوائية لتقييم العمليات المعقدة والديناميكية، ليس مجرد التدخلات البسيطة والساكنة. يجب على المقيمين أن يهدفوا إلى الإجابة على الأسئلة الأكثر أهمية بالنسبة للقرارات المستقبلية وأن يفعلوا ذلك على نحو موثوق به بقدر الإمكان. تتحسن الإعتمادية بالتصميمات العشوائية متى أمكن ذلك، وبالإنتباه إلى النظرية والإختبارات الأساسية عن سبب نجاح أو فشل التدخلات، حتى يمكن نقل الدروس بأفضل مايمكن إلى البيئات الأخرى.

1. مقدمة

لماذا نقوم بالتقييم؟ هناك ثلاثة أسباب بارزة: لمعرفة أين ننفق مواردنا المحدودة، ولمعرفة كيف نُحسّن برامجنا، ولتحفيز من لديهم المال أن يعطوا أو يستثمروا أكثر.

أحب أن أبدأ بتجربة فكر من فيلسوف النفعية بيتر سينجر. هل تقوم بانقاذ طفل يغرق إذا كان هذا الأمر سيكلفك 100 دولار بسبب تلف الملابس أو بسبب فقدان موعد؟ أكثر الناس سيقولون نعم لهذا السؤال. لكن هل ترسل

[2] البريد الالكتروني: dean.karlan@yale.edu

أيضا الـ100 دولار الأن لمنظمة غير حكومية فى دولة فقيرة لتنقذ طفلا؟ العديد من الناس يقولوا لا، مجادلين بأنه لا أحد يعرف هل الـ100 دولار التى يرسلونها يمكن أن تنقذ طفلا أو قد تضيع. هذا عذر شائع للتراخى. التقييم يدحض هذا العذر.

هناك كثير من النقاش حول إستخدام التجارب العشوائية المخططة[3] (RCTs) مقابل الأساليب الأخرى لبرامج التقييم. لكن، فى العديد من الحالات، يتم إصطناع هذه التفرقة بين المناهج التجريبية والمناهج الأخرى ويجعل من التداخل بين الأساليب التجريبية والنوعية التى يمكنها أن تميز التقييم الجيد صعبا. فى هذه الورقة، أبدأ بشرح بعض سوء الفهم المرتبط بطرق القياس، والإيعاز وجدوى التجارب العشوائية المخططة. بعد ذلك، سوف أصف ثلاثة أمثلة لبرامج التنمية الشائعة – التحويلات النقدية المشروطة والبنية التحتية وإجراءات مكافحة الفساد – والظروف التى يجب أو لايجب توظيف التجارب العشوائية المخططة فيها كجزء من إستراتيجية التقييم.

2. المفاهيم الخاطئة للتجارب العشوائية

أحد المفاهيم الخاطئة هو أن المقيم قد يختار أكثر إما أن يقوم بعمل تقييم نوعى أو محاولة تجربة عشوائية. يكمن وراء ذلك طيف خاطئ من دقة "الايعاز" بالتجارب العشوائية المخططة من جهة، وبأساليب نوعية من جهة أخرى. فى الواقع، المنهجيات النوعية ليست عكس محاولات الرقابة العشوائية المخططة. فمن ناحية، غالبا ما يشتمل التقييم الجيد للتجربة العشوائية المخططة على تقييم دقيق لكيفية عمل البرنامج، وتصميمه الأولي، ونظرية التغيير، ومشاركة المستفيدين .. إلخ.

يلزم لتوضيح المناقشات حول أساليب التقييم، الفصل بين تجميع البيانات وقياس النتائج – ماذا نقيس وكيف نقيس ومن نشمله فى هذه العملية – عن كيف نقوم بإنشاء سببية بين النتائج والتدخل. تقوم التجارب العشوائية المخططة بايجاد سببية من خلال قياس الواقع المضاد: ماذا قد يحدث إذا لم يوجد هذا البرنامج أو هذه السياسة.

[3] التجارب العشوائية المخططة – Randomized Control Trials (RCTs)

وكممارسة معيارية فى التجارب الطبية، فإنهم يحققوا ذلك بتحديد أشخاص في مجموعات بدون علاج وعلاج بطريقة عشوائية، وتكون هذه المجموعات متشابهة حول السمات الملحوظة أو غير الملحوظة إذا كان حجم العينة كبير بشكل كاف ، فيما عدا البرنامج العلاجي الذي يتم تجربته.

2-1 إنشاء السببية

يعتبر الإختيار العشوائى مفيدا حتى لا نقع في الإنحياز عند الإختيار، أو بمعنى آخر لأن المشاركين فى برنامج ما غالبا ما يكونوا مختلفين عن غير المشتركين. إذ أن بدلا من أن كان يجب علينا أن نقارن بين هؤلاء الذين قد إشتركوا فى برنامج لكنهم لم يفعلوا، فإننا قد نصل إلى مقارنة إثنين من مجموعتين محتملتين مختلفتين جدا من الأشخاص. من السهل أن نرى تلك المجموعات التى قد تختلف فى طريقتين هامتين، لكن من الصعب قياسهما. هؤلاء الذين يلتحقون بالبرنامج قد يتم تحفيزهم أكثر على تحسين وضعهم، أو على تمكينهم أكثر، أو على أن يتم تعليمهم بشكل أفضل. وقد يكون لديهم وقت فراغ أكثر. غالبا مايحاول الباحثون أن يتحكموا فى تلك الإختلافات، لكن حتما هناك متغيرات محذوفة أو أشياء أخرى مثل التحفيز الذي قد يشكل مشكلة فى قياسه. هذه الفروق تعنى أن تقديرات أثر التدخل متحيزة لأن الإختلافات فى التأثيرات على مجموعات المقارنة والمتأثرة بالتدخل قد ينتج عنها تلك الصفات غير الملحوظة بدلا من كونها حدثت بسبب التدخل.

2-2 البيانات والقياس

تعتبر القياسات الكمية للتأثير مفيدة للتقييمات لأنها تسمح للباحثين بأن ينشأوا أهمية إحصائية لأثر البرنامج. لكن التجارب العشوائية المخططة لاتحدد أسلوب واحد لتجميع البيانات. يمكن إستخدام كل من البيانات الكمية والنوعية داخل إطار التجارب العشوائية المخططة في نفس التقييم. يمكن إستخدام أساليب من علم الإقتصاد وعلم الإجتماع وعلم النفس أو من أنظمة أخرى، وأيضا عمليات تشاركية بإشراك أصوات محلية (على سبيل المثال، أنظر تشاتوبادياي ودافلو، (2004) الذين وجدوا أن النساء فى غرب البنغال كانوا أكثر إحتمالا أن يشتركن فى عملية صناعة السياسة إذا كانت قائدة مجلس القرية إمرأة)، من بين الآخرين، وحتى "الأشخاص

الذين محل إقامتهم الدائم بعيد عن مكان عملهم" كما ناقشها تشامبرز فى هذا المنتدى (كارلان وزينمان، 2009).

أحد المفاهيم الخاطئة الشائعة الموجهة إلى مناصري التجارب العشوائية المخططة هو أننا نقترح أنهم يمكن ويجب أن يتم إشراكهم فى كل برنامج. تعتبر التجارب العشوائية المخططة أداة بحث هامة لأن السببية التى تنشأها هذه المحاولات تقدم مقياس دقيق لأثر البرنامج، ولذلك تساعد على معرفة هل يجب أن يتم مضاهاتها فى مكان آخر، وأيضا كيف يجب أن يتم تحسينها. مع ذلك، فان التجارب العشوائية المخططة ليست دائما ممكنة التنفيذ. حيثما تعتبر التجارب العشوائية المخططة مناسبة يتوقف ذلك جزئيا على الوضع، وعلى السؤال المطروح أيضا. وكما يناقش مقال رافاليون (2009) فى هذا المنتدى، يجب على الفرد ألا يبدأ الربط بالطريقة أو المنهج وانما يحدد أولا الأسئلة التى تحتاج إلى إجابة، وبعد ذلك يفحص الأداة الأكثر ملاءمة للإجابة على تلك الأسئلة. وعندما يتم تنفيذها، فإن التجارب العشوائية المخططة تقدم أكثر تقدير غير متحيز لأثر البرنامج، لكن مجرد كونه ممكن التنفيذ لا يعني تنفيذه بأى شكل لمجرد أنه ممكن. حيثما لا توجد إستراتيجيات مناسبة لتحديد المبحوثين، بدون شك، تعتبر التجارب العشوائية المخططة الوسيلة الأكثر عملية لإجراء بحث يمكن تصديقه.

2-3 المدخلات الخلاقة فى التجارب العشوائية المخططة

بينما نؤكد أن التجارب العشوائية المخططة لا تصلح فى كل مكان، فإن العديد من الحالات التى تبدو غير ممكنة هى فى الحقيقة ممكنة مع قدر قليل من الإبداع. على سبيل المثال، تستطيع التدخلات أن تستفيد غالباً من محددات التنفيذ، وتعمل بطريقة عشوائية فى المجتمع المحلى أو على مستوى جغرافى آخر بدلا من إنتقاء الأفراد بشكل عشوائى فى المجموعات التي تعرضت للتدخل والمقارنة. هناك تقييمات عديدة تقيس أثرالتمويل متناهي الصغر الذى يستخدم هذا الإتجاه. فى حالات أخرى، الفروق فى كثافة تسويق برنامج بالنسبة لمجالات مختلفة (تشجيع التصميم) يمكن أن يتم إستغلاله فى التجارب العشوائية. المعيار الأساسى للتجارب العشوائية

المخططة هو حجم العينة والذي يجب أن يكون فى وحدات منفصلة كافية بحيث يمكن قياس الآثار غير المباشرة والآثار المتوازنة العامة. إذا تم تخطيطها كما ينبغى، وإذا لم تكن الأثار قد تجميعها بشكل مفرط (على سبيل المثال، على مستوى الدولة)، يمكن استخدام تصميمات التجارب العشوائية الدقيقة لقياس كل من الآثار المباشرة للتدخل، وأيضا الآثار غير المباشرة، الإيجابية والسلبية على مجموعات من خارج المستفيدين المباشرين. وهذه المجموعات فى الواقع هى بعض من محاولات التجارب العشوائية المخططة الأكثر إثارة للقراءة، لأنها تساعدنا على فهم ليس فقط هل فكرة ما تعمل على فرد معين ولكن كيف تؤثر على نطاق أوسع من خلال الآثار المباشرة وغير المباشرة.

4-2 مناهج التنفيذ الثابتة مقابل مناهج التنفيذ الديناميكية

مفهوم خاطئ شائع آخر للتجارب العشوائية المخططة هو أن هذا التدخل يجب أن يكون متجانس وثابت. حقيقة أن التدخلات 'الناشئة، المعقدة أو المعقدة'، مثل تلك التى ناقشها روجرز (2009) فى هذا المنتدى، ليست أكثر صعوبة بالنسبة للتجارب عشوائية المخططة أن نتناولها في تجارب غير عشوائية. تفشل المناقشات التى تقترح التعقيد وعمل عملية ديناميكية وتلحق ضررا بالتجارب العشوائية، فى التعرف بالضبط على ما تحدثه لنا التجارب العشوائية. ببساطة تساعد التجارب العشوائية فى إنشاء مجموعة مقارنة موضوعية يتم مقارنة التغييرات الناشئة عن التدخل بالتغيرات الطبيعية في مجموعة المقارنة. بالطبع التدخل ذاته يمكن أن يكون ثابتا وبسيطا أو معقدا ومتغيرا. إذا كان معقدا ومتغيرا، فى هذه الحالة يتم وصف التقييم بالطبع كما يلى: كتقييم لعملية، كفرصة معها بعض الموارد، كتدخل ديناميكى وسلس تم قيادته بطريقة معينة، إلخ. الحل هنا هو العملية ككل وليس الأنشطة الفردية التى يشملها البرنامج. إذا كان المشروع ناجحا، إذن الذى لابد أن يتكرر هو عملية وضع الموارد فى مكانها وتيسير إستخدامها، إلخ. هذا يشبه بكثير تدخلات تنمية مجتمعية عديدة يتم فيها توفير موارد مثل التدريب والمساعدة الفنية المخصصة للمجتمعات المحلية، ويتم وضع تدريبات تيسير موضع التنفيذ لمساعدة المجتمعات المحلية على النمو والإزدهار.

نقوم بعمل مجرد تقييم كهذا بإستخدم منهج التجارب العشوائية كاملا بأدوات نوعية وكمية، كما في مشروع الجوع فى غانا وبرنامج التنمية الذى يقوده المجتمع المحلى فى سيراليون. بالطبع، من الأهمية أن نفهم أن ما يتم تقييمه هنا هو العملية التعاونية، وليس من التدخلات المحددة بوضوح. ليس من الممكن أن نعرف مقدما ماهى المدخلات التى سيختارها الفاعلون المعينون ولا أن نتوقع أن نفس العملية فى مكان آخر قد تعطى نفس الإختيارات. لذلك فإن الدروس المستفادة من تقييم كهذا هى عن التغييرات التى يمكن أن يتوقعها الفرد من عملية كهذه - وليس من إختيارات وإستثمارات محددة يختار الفاعلين أن يعملوها ، وإنما من عملية التيسير و/أو التمويل القروي، بينما تقوم بإعداد البرنامج بأنفسهم. ويعنى هذا، إذا كان مسؤولي البرنامج أو المديرين مهتمين بقياس أثر الأنشطة التى تشكل التدخل، قد يتم تصميم تجربة عشوائية لتحقيق نتائج حصيفة من التدخلات المعقدة. قد يتطلب ذلك عمل تنوع عشوائى لمكونات التدخل فى مجموعات المعالجة المتعددة. المقارنة المحتملة قد تكون أثر مجموعة أساسية من الخدمات، مع أو بدون تفاعل مع أحد المكونات الإضافية أو أكثر.

للتجارب العشوائية المخططة ميزة هامة على الأساليب الأخرى ، لأنها يمكن أن تخاطب تحيزات إختيار متأصلة فى برامج إجتماعية عديدة، وفى دراسة الآثار المباشرة للأنشطة المختلفة فى تصميم متعدد الأنشطة. على سبيل المثال، إذا قام شخص بعمل تقييم لتدريب تجارى ووجد زيادة فى الأرباح، وخصوصا بين هؤلاء الذين يسجلون عملياتهم بشكل أفضل، هل هذا يؤدي الى إقتراح وجوب ترويج التدرب على حفظ السجلات؟ ربما يكون حفظ السجلات هو العنصر الأساسى للتدريب، أو يحتمل أن رجال الأعمال الآفضل بالطبيعة قد شاركوا فى البرنامج التدريبي لحفظ السجلات. تستطيع التجارب العشوائية المخططة أن تفصل فى هذه القضايا بتخصيص مشاركين لتلقى تدريبا مع أو بدون برنامج خاص لحفظ السجلات.

مفهوم خاطئ آخر للتجارب العشوائية (RCTs) هو أنها تقيس آثار التدخل على متوسط العينة متجاهلة الآثار المتفاوته على القطاعات المختلفة من العينة. فى الحقيقة، فرضا أن حجم العينة كاف، وخطة أخذ عينات تشمل

مجموعة متنوعة من الناس قد تكون صالحة للبرنامج الأوسع، فإن التجارب العشوائية يمكنها أن تساعد فى تحديد الفئات التى تأثرت أكثر من البرنامج، والفئات التى تأثرت أقل، أو حتى تلك التي تأثرت سلبيا. على سبيل المثال، كانت إحدى النتائج المدهشة من قياسات التجارب العشوائية هو قياس أثر برنامج تدريبى لأصحاب مشروعات متناهية الصغر فى بيرو، بأن أصحاب الأعمال الذين أبدو عدم إهتمام بالتدريب الإضافى إستفادت بالفعل لحد ما أكثر من البرنامج (كارلان وفالديفيا، 2008).

3. ثلاثة أمثلة

يقدم مقال رافاليون فى هذا المنتدى نظرة عامة لأنواع الأسئلة التى يجب على الفرد أن يطرحها فى بداية عملية التقييم، لكى يحدد الهدف والمجال الخاص بالتقييم، وكذلك أسئلة البحث الأساسية. وأثناء نقاشه، وأعتمادا على المهمة، سيتحدد ما اذا كانت العشوائية ممكنة ام لا. تقدم هذه الأمثلة الثلاثة حالات ممتازة لتلك النقطة بالضبط. سأناقش هنا كل من الخطط العريضة لكل منها عن كيف يمكن تقييمها، وبعد ذلك أفكارا معينة فى نطاق كل منها عن كيف تتم الإجابة على الأسئلة الثانوية والأسئلة المرتبطة بعمليات التنفيذ من خلال التجارب العشوائية، حتى في حالات إذا كان التدخل الأساسى يوظف أساليب أخرى لتقدير الأثر العام. هذه الأفكار ليست عوضا عن التقييم غير التجريبى الشامل، لكنه يستطيع أن يقدم أساليب مفيدة لإيجاد بيانات موضوعية دقيقة لتساعد مع أسئلة التنفيذ الهامة في المستقبل.

نعود إلى المثال الأول، التحويلات النقدية المشروطة[4]، الأسلوب الموصى به لتقييم الأثر هو محاولة التجارب العشوائية التى تشمل تجميع البيانات الكمية والنوعية. وقد إستُخدامت فى بلدان عديدة. حيثما كانت الحكومات لديها موارد محدودة للتوسع فى برامج التحويلات النقدية المشروطة، فإن التوزيع العشوائى يعتبر بشكل خاص الطريقة العادلة والشفافة لتوزيع الإستفادات على مراحل. تبين الأبحاث الأخيرة أنه إذا تم تصميمها بشكل مناسب، فإن التحويلات النقدية المشروطة يمكنها أن تكون وسيلة فعالة لتحقيق الأهداف الهامة للسياسة العامة. ورغم ذلك، فإن السؤال عن كيفية تنفيذ هذه البرامج بأفضل طريقة لازال مطروحا بلا إجابة شافية. على سبيل

[4] التحويلات النقدية المشروطة – Conditional Cash Transfer (CCT)

المثال، أسئلة التنفيذ: كم مرة يتم السداد، وهل يجب أن نراعى إضافة خدمات إدخار، وهل يجب أن يتزامن الدفع مع مواعيد تسديد نفقات التعليم.[5]

بالنسبة للمثال الثانى، البنية التحتية، هناك العديد من الخيارات للتصميمات الممكنة فنياً، ولكنها تتطلب درجة متنوعة من الإلتزام من جانب المسؤولين الحكوميين القائمين بإدارة البرامج. سأناقش اعادة تأهيل الميناء والطرق الرئيسية والطرق الفرعية فى المناطق الريفية. لسوء الحظ، القائمين بالتقييم كثيرا جدا ما 'يطلب منهم أن يقوموا بالتقييم بعد فوات الأوان. بغض النظر على الأسلوب الذي يتم إستخدامه، من المفضل جدا أن يتم إعداد طرق واساليب التقييم مبكرا، وأن تكون لدينا أهدافا واضحة وشاملة عن بماذا وكيف نقوم بقياس النتائج.

قد يشمل تقييم إعادة تأهيل الميناء والطرق الرئيسية تركيزا شديدا على عملية منهجيات التقييم. ستكون الخطوة الأولى اعداد إطار منطقى بأهداف مختلفة، على سبيل المثال: عدد أيام الإنتظار بالميناء، وعدد الأيام اللازمة للنقل، وتكلفة الشحن والنقل البرى، وقيمة البضائع التى يتم شحنها، وكمية البضائع التى يتم شحنها، وعدد السفن والشاحنات والسيارات القادمة للميناء والخارجة منه. هناك إحتمال إستخدام أدوات إقتصادية للقياس إستنادا إلى الأثار المتناولة على الصناعات والتعريفات الجمركية على سبيل المثال. هذه ببساطة هى متابعة للبرنامج، وهى مهمة لكل من إدارة التنفيذ والمساءلة عن النتائج.

يمكن فى هذه الحالات، توظيف التجارب العشوائية المخططة للمساعدة على إجابة للجوانب النقدية لنظرية التغيير الخاصة بالبرنامج، لكنها ليست من المحتمل أن تشمل التدخل بأكمله. على سبيل المثال، السؤال الأساسى لتأهيل ميناء في دولة نامية قد يكون "هل يؤدى إنخفاض تكاليف النقل إلى زيادة نمو الصناعة فى المناطق الريفية؟". فى هذه الحالة، يمكن اختبار التجارب العشوائية التى تدعم تكاليف النقل بشكل عشوائى فى بعض المناطق من أجل فحص التغيير فى النشاط الإقتصادى كأحد النتائج.

[5] من أجل عمل نقاش جيد عن كيفية تصميم بيئات الإختيار التى تساعد الناس على أن يختاروا بشكل أخلاقى، راجع دفعة "Nudge" التى قام بكتابتها ريتشارد تيلر وكاس سنستين)

الإجابة على الأسئلة الأساسية للسياسة حول أثر برامج الطرق الفرعية من خلال التجارب العشوائية يمكن أن تكون ممكنة فنيا، لكنها أيضا من المحتمل أن تتطلب قدرا كبيرا من الإلتزام من جانب صانعى السياسة. هذا المثال هو احد الأمثلة التى يمكن أن نستفيد استفادات كثيرة من حيث الدروس المستفادة من تجارب البلدان الأخرى، لكنه أيضا أحد الأمثلة التى نعترف أنها قد تكون صعبة التحقيق سياسيا. إذا كان هناك طرقا كافية وظروف الجغرافيا وتكاليف البناء تسمح بذلك، فإن هناك إحتمال للعشوائية في إنشاء الطرق. تخيل خطة مدتها عشر سنوات لتحسين أو بناء طرق فرعية فى المناطق الريفية.

تعد عشوائية الترتيب (أ) مقبولة و (ب) يمكن تقييمها بسهولة. يمكن تنفيذ ذلك بدمج الأولويات الخاصة بالطرق ضمن خطة عشرية، إذا كانت بعض الطرق أكثر أهمية لأسباب إقتصاديا وجغرافية (أو سياسية). صانعى السياسات قد يعترفون بأن أحد مزايا التجارب العشوائية، فى هذا السياق، هى أنها تتجنب المحسوبية السياسية لكى يتم إتخاذ قرار بشأن الترتيب (بمعنى، قد يتم إختيار الطرق للمرحلة 1 أو المرحلة 2 فى عملية تداولية تشاركية شفافة، ثم يمكن ترتيب إنشاء الطرق ضمن كل مرحلة عشوائيا، وبذلك تكون مقبولة). فى هذه الحالة، يكون استخدام التجارب العشوائية المخططة هى إحدى جوانب تصميم التقييم. فهى أيضا، قد تشمل إستخدام أساليب القياس الإقتصادى بمافيها من مناهج الفروق، قبل وبعد، أوالمقارنة المستعرضة لما تم إنشاؤه إزاء ما لم يتم إنشاؤه (على سبيل المثال: مقارنة المدن التى تبعد 5 أميال عن الطريق الذى تم إصلاحه أو الذى تم إنشاؤه، إزاء المدن التى تبعد 5 أميال عن الطريق الذى لم يتم إصلاحه أو لم يتم إنشاؤه).

المثال الأخير، قياسات مكافحة الفساد، لا تكون دائما هي المجال الذى يتم فيه البحث عن تقييم إسلوب الإعزاء، مع أنه من الممكن أن يكون هناك بعض النتائج التى يمكن قياسها لهذه العملية مثل القيام بإلقاء القبض على المخالفين أو طرد السياسيين من وظائفهم. لكن بداخل مربع كيف ولماذا يلجأ الموظفون العموميون إلى تكتيكات غير قانونية، فهناك الكثير الذي يمكن تعلمه. وعلاوة على ذلك، هذه هى إحدى المجالات التى تأخذ فعلا فيها

الشفافية أهمية فى منهجية التقييم. على سبيل المثال، تم عرض التدقيقات المحلية فى البرازيل على شاسة التليفزيون. العمل الذى كتب عنه أولكن (2007) فى إندونيسيا هو مثال جيد آخر، يمكننا من أن نتعلم الفاعلية النسبية للأساليب المختلفة لمكافحة الفساد، من خلال مزيج من جمع البيانات النوعية (مفهوم الفساد والأساليب التشاركية في إجتماعات القرية) وجمع البيانات الكمى.

4. الخلاصة

إن إحدى المزايا النهائية لمنهج التجارب العشوائية هى الإستقلالية التى تسمح بها، التى يستطيع الفرد فيها أن يُنشأ إختبارات إحصائية واضحة قبل وبعد، وبعد ذلك يجعل البيانات تتحدث هل هى تعمل أم لا. أخيرا، الهدف للتقييم هو المساعدة على إتخاذ قرار لما يجب أن يتم عمله فى المستقبل. وهذا سواء للمانحين الذين يحتاجون أن يعرفوا أين يجب أن يضعوا أموالهم، وللمتشككين الذين يريدون رؤية هل هذه البرامج يمكن أن تعمل وتؤدي الى المردود المرتقب، وللمنفذين الذين يحتاجون أن يعرفوا كيف يتم تصميم هذه البرامج أفضل تصميم. بعض الأعمال الأكثر إثارة تستخدم أساليب مختلطة بدمج الأساليب النوعية فى التجارب العشوائية وبإستخدام التجارب العشوائية لتقييم العمليات الديناميكية والمعقدة، مثل برامج التنمية المجتمعية.

فى هذا المقال، ركزت على إثنين من القضايا الأساسية فى الحوار حول أساليب تقييم الأثر: تحليل ما الذى يجب قياسه مقابل ما الذى يجب مقارنته. بالنظر إلى تلك الأسئلة، نستطيع أن نرى تركيز محاولات التجارب العشوائية على السؤال الأخير (ما الذى يجب مقارنته)، وهذه المحاولات مرنة فى تضمين أساليب العديد من الأساليب التشاركية والنوعية والكمية للسؤال الأول (ما الذى يجب قياسه). لقد حاولت، أيضا، أن أدحض بعض المفاهيم الخاطئة الشائعة عن طرفى المناظرة. حتى مؤيدى التجارب العشوائية (RCTs) لايدافعون عن أنه يتم إستخدامها فى كل مكان، ومن أجل كل برنامج. إذا كنت أخاطر بعمل تخمين، فقد يتم إنفاق أقل من 1% من ميزانيات التقييم للتجارب العشوائية. أعتقد أنه يجب أن يتم عمل أكثر من ذلك ولكن ليس 100% (أو حتى 99%).

المراجع:

Chattopadhyay, Raghabendra, and Esther Duflo. 2004. Women's leadership and policy decisions: evidence from a nationwide randomized experiment in India. Econometrica, 72(5), 1409–1443

Karlan, Dean, and Martin Valdivia. 2008. Teaching Entrepreneurship: Impact of Business Training on Microfinance Institutions and Clients. Yale University Economic Growth Center working paper.

Karlan, Dean, and Jonathan Zinman. 2009. Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts. Review of Financial Studies.

Olken, Benjamin. 2007. Monitoring Corruption: Evidence from a Field Experiment in Indonesia. Journal of Political Economy 115: 200-249.

Ravallion, 2009. Evaluating Three Stylized Interventions, Journal of Development Effectiveness, forthcoming.

Rogers, Patricia J., 2009, Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation, Journal of Development Effectiveness, forthcoming.

Thaler, Richard, and Cass Sunstein , 2008, Nudge: Improving Decisions About Health, Wealth, and Happiness, New Haven, CT: Yale University Press.

تقييم ثلاثة أساليب من تدخلات

مارتين رافاليون[6]
مجموعة بحوث التنمية
البنك الدولي ، واشنطون ، 20433 ، الولايات المتحدة الامريكية

ملخص

طلب مني، جنبا إلى جنب مع متحدثين آخرين خلال إحدى جلسات هذا المؤتمر، مناقشة تصميم التقييم لثلاثة أساليب من التدخلات: التحويلات النقدية المشروطة، برنامج قطاع النقل، ولجنة مكافحة الفساد. تُسجل هذه الورقة ردودي، وتوضح بعض النقاط، بما في ذلك إشارات إلى الأدبيات والمراجع. سأبدأ ببعض الاقتراحات العامة بشأن قضايا للتفكير فيها قبل أي تقييم. وبعد ذلك، سأحاول توضيح هذه النقاط، مع الإشارة إلى أساليب التدخلات الثلاثة.

1. مقدمة

طلب من المشاركون في هذه الجلسة مناقشة كيفية تقييم ثلاثة تدخلات:

1. "التحويلات النقدية المشروطة" في إحدى بلدان أمريكا الوسطى، حيث تتلقى الأسرة مبلغاً شهرياً طالما بقيت الفتيات في سن التعليم في المدرسة وتلتزم بالحضور ومتطلبات الأداء.
2. "برنامج قطاع النقل" في إحدى بلدان جنوب آسيا والذي يشمل على إعادة تأهيل الميناء وإعادة تأهيل طرق النقل، والاستثمارات الجديدة في الطرق الفرعية الريفية.
3. لجنة مكافحة الفساد[7] في بلد أفريقي. ويتضمن البرنامج المساعدة في وضع مبادئ توجيهية، وتحسين البنية التحتية وجولات دراسية. ويجري تنفيذ برامج مماثلة في ستة بلدان أخرى.

هذه التدخلات الثلاثة مختلفة جدا في كثير من النواحي، ولكن الفرق الأكثر أهمية في سياق التقييم هي في الدرجة التي يمكن علاجها على أنها برامج مستهدفة، بمعنى أن بعض وحدات الرصد (الأسر والشركات

[6] هذه هي آراء الكاتب ولا يجب أن تنسب إلى البنك الدولي أو أي منظمة منتسبة. العنوان: mravallion@worldbank.org. الكاتب يعبر عن امتنانه للمشاركين في المؤتمر على تعليقاتهم، وفيل كيفر، نوربرت شادي، دومينيك فان دي وال وهوارد الأبيض للمناقشات والتعليقات المفيدة.

[7] لجنة مكافحة الفساد – Anti-Coruption Committee (ACC)

والقرى والمناطق) تتلقى البرنامج، ولكن البعض الأخر لا يتلقاها. وفرق وثيق هو مدى تأثيرها على الأخرين، حيث تتأثر وحدات غير مستهدفة (سلباً أو إيجاباً) من خلال البرنامج. حقيقة أن البرنامج مستهدف لا يعني بالطبع أن الوحدات غير المستهدفة لن تتأثر. تسرب الآثار يمكن أن تكون مصدرا خطيرا للتحيز في أساليب التقييم الكلاسيكية.[8]

على النقيضين، برنامج التحويلات النقدية المشروطة هو برنامج مستهدف، يستهدف أسر محددة مع احتمال لتسرب متواضع للآثار، ولكن لجنة مكافحة الفساد تؤثر على مستوى الاقتصاد، على الرغم من أنه قد يكون لها بعض الجوانب الميتهدفة، مثل عندما تستهدف قطاعات محددة من الحكومة أو الشركات الخاصة. بين هذين النقيضين، يعتبر برنامج النقل هو هجين، حيث أن معظم المكونات مستهدفة، ورغم ذلك فإن التسرب للآثار يمكن أن تكون له أثر كبيرة.

هناك حاجة إلى أدوات تقييم مختلفة للبرامج المستهدفة مقابل البرامج غير المستهدفة، على الرغم من أن المبادئ الأساسية للتقييم هي نفسها - بما في ذلك الحاجة إلى تقييم الأثر مقارنة بالواقع المغاير. يسهل البرنامج المستهدف المقارنات الرصد، حيث تستخدم بعض العينات المختارة من الوحدات غير المستهدفة في محاولة للاستدلال على الواقع المغاير واضعا في الاعتبار افتراضات محددة.[9]

في كل حالة، وقبل مناقشة قضايا التقييم لهذه التدخلات، سأبدأ بطرح بعض الأسئلة الأساسية حول سياق التدخل. من الطبيعي أنه سيكون من الصعب أن نكون محددين جدا دون معرفة المزيد عن الإعداد، ولكنني سأحاول.

أولا، بعض التعليقات العامة للتدخلات الثلاثة.

2. الأسئلة العامة

[8] انظر، على سبيل المثال، تشن وآخرون (2009) .في سياق الردود على الإنفاق العام المحلي على برامج التنمية في المناطق الفقيرة في الصين ، وأودور وآخرون (2009) بشأن تسرب الآثار للتسويق اجتماعية لعلاج الملاريا في كينيا.

[9] لمزيد من المناقشة عن الافتراضات والأساليب المستخدمة للبرامج المستهدفة والمراجع وأدبيات (موسعة) أنظر رافاليون (2008)

الأسئلة الرئيسية التي أود أن أطرحها في بداية أي تقييم هي ما يلي:

- لماذا هذا التدخل؟ إن فهم الأساس المنطقي لتدخل معين مهم لتصميم تقييم مفيد، ولكنه مهم أيضاً في صنع السياسات الجيدة على النطاق الواسع (وهو بعد كل شيء الهدف النهائي للتقييم). التحقيق في مبررات التدخل، قد يؤدي إلى تدخل مختلف.

- ما الذي نعرفه حاليا عن هذا النوع من التدخل وما هي أهم الثغرات في هذه المعرفة؟ هناك تقريباً دائماً بعض التجارب السابقة ذات الصلة. في البداية، يمكن اجراء مراجعة جيدة للأدلة السابقة حتى تكون كاشفة، وقد تؤثر على تصميم البرنامج وتنفيذه، وكذلك القضايا التي يتم اختيارها للتركيز عليها أثناء التقييم.

- ما هو الواقع المغاير ذو الصلة؟ والواقع المغاير الكلاسيكي هو غياب البرنامج، ولكن هذا لا يجب أن يكون للواقع المغاير ذو الإهتمام الأكبر لصناع السياسات الذين غالباً ما سوف ينفقون نفس الموارد على بعض البرامج الأخرى. قد يظهر برنامج معين أداءاً جيداً ضد خيار عدم القيام بأي شيء، ولكن سيئ أمام بديل أخر ذو جدوى. رسمياً، مشكلة التقييم أساسية لا تختلف إذا كان بعض البرامج البديلة هي الواقع المغاير، من حيث المبدأ، يمكننا تكرار تحليل كل بديل ممكن نسبة إلى "الواقع المغاير بعد التدخل"، ومقارنتها بعضها ببعض. هذا لا يتم كثيراً في تقييم مشاريع التنمية، ولكنه أكثر شيوعاً في الرعاية الصحية والتجارب الطبية (حيث تحصل على المجموعة الضابطة على التدخل القائم ومجموعة العلاج تحصل على التدخل الجديد).

- ما هي النتائج المرجوة؟ وخلال أي فترة من الزمن؟ وما هي النتائج غير المرغوب فيها المحتملة؟. من الواضح أنه من المهم جدا أن نعرف الأهداف، وكيف يمكن أن تترجم إلى نتائج محددة قابلة للقياس

(كماً أو نوعاً). وليس أقل أهمية أن نعرف خلال أي فترة من الزمن نتوقع النتائج (الإيجابية والسلبية).[10]

- ما هي المعايير التقدير ذات الصلة؟ التقييمات الكلاسيكية تركز على اثنين فقط، وهي متوسط تأثير هذا التدخل على الوحدات التي يتم إعطاءها الفرصة للتأثربالتدخل (المعاير "المستهدف من التدخل")، ومتوسط تأثير على أولئك الذين يحصلون عليه. ومع ذلك، صانعي السياسات عادة لا يهتمون فقط بهذين المعيارين. أسئلة أخرى مفيدة تشمل: هل يعمل التدخل على النحو الذي كان مخطط له؟، ما هي الفئات من الناس التي استفادت، وما هي الفئات التي خسرت؟، ما هي نسبة المستفيدين للمشاركين؟، ماذا سيحدث عند تعميم البرنامج؟ وكيف يمكن أن يعاد تصميمه بحيث يعزز الأثر؟

- ما هي آليات الانتقال المتوقعة؟ إن تقييم الأثر الكلاسيكي يمكن أن يكون "الصندوق الأسود" الذي سيخبرنا القليل عن كيف يمكن للبرنامج أن يكون له تأثير أم لا. من أجل تصميم تقييم يمكن أن يلقي الضوء على هذا الصندوق الأسود، لابد أن نفهم المنطق النظري للبرنامج - الطرق المحددة التي من المتوقع للتدخل أن يُحسّن حياة الشعوب. (سأناقش مثل هذا التقييم "القائم على نظرية "في وقت لاحق). يمكن أن يحيط الغموض لصناع السياسات حول تلك الآليات، والتحقيق من قبل مقيمين قد يكون مفيداً في الكشف عن مبررات التدخل، وأحيانا يكشف أن الأساس المنطقي مشكوك فيه.

- ما هي الخيارات المنهجية الممكنة في أطار هذا الإعداد؟ سيكون هناك القيود ذات الصلة (فنية واقتصادية وسياسية وأخلاقية) للتقييم وهذه ينبغي أن تحدد في البداية من الناحية المثالية. بعض أساليب التقييم الأخرى المرغوب فيها قد تكون غير ذات جدوى في إطار معين. من بين الخيارات الممكنة، يجب أن يعتمد الاختيار على الإجابات لجميع الأسئلة السابقة، بدلا من التفضيلات المنهجية للمقيم. قد يبدو هذا منطقياً، ولكنه ليس مطبقاً دائماً. وفي كثير من الأحيان، يأتي المقيم بمجموعته الخاصة المفضلة من الأدوات لهذه المهمة، ويختار الأسئلة التي يمكن معالجتها من خلال هذه الأدوات المحددة

[10] انظر كينج وبيرمان (2009) لمناقشة مفيدة حول هذه النقطة.

(وليس العكس). أحيانا تكون الأدوات المقبولة متناثرة بشكل ملحوظ، ويبدأ المقيم بأداة واحدة مفضلة لديه (مثل التوزيع العشوائي، أسلوب الاقتصاد القياسي غير التجريبي أو أداة النوعية المفضلة) ويتطلع لطرح الأسئلة التي يمكن معالجتها بهذه الأداة. وكقاعدة عامة، هذا لا يجعل التقييمات مفيدة كثيراً.

الإجابة على هذه الأسئلة غالباً ما تكون أساسية لتصميم التقييم، ولكنها سوف تكون محددة لكل برنامج والإعداد له. بعيداً عن الإعداد، ليس هناك طريقة صحيحة واحدة مناسبة لأي تدخل. في ضوء ذلك، إسمحوا لي أن أقدم بعض الأفكار حول قضايا التقييم التي أثارتها هذه التدخلات الثلاثة المختلفة إلى حد ما.

3. برنامج التحويلات النقدية المشروطة

هذا هو أسهل البرامج الثلاثة لسببين: أولا، برنامج التحويلات النقدية المشروطة هو برنامج مستهدف، وربما يكون من المعقول أن نفترض أن فرصة وصول تأثيره لغير المستهدفين تكاد تكون معدومة. ثانيا، (يرجع ذلك جزئياً الى السبب الأول) كان هناك الكثير من البحوث التقييمية عن التحويلات النقدية المشروطة، ولهذا نعرف الكثير حول القضايا المتعلقة بهذه البرامج وأفضل السبل لتقييمها. ومع ذلك، لا تزال هناك فجوات في المعرفة.

الفكرة الأساسية لبرنامج التحويلات النقدية المشروطة هي أن الأسرة المستفيدة يجب أن تثبت الحضور المدرسي الملائم (وفي بعض الأمثلة) الرعاية الصحية؛ يتم تحويل الدفعة في حالة تحقق هذه الشروط (وتسمى احيانا "المشاركة في المسؤولية / المسؤولية التشاركية).[11] احد الأمثلة المؤثرة كان "برنامج الغذاء مقابل التعليم[12]" ببنجلاديش؛ و"البرنامج التعليمي – الصحة والغذاء[13]" (والمسمى الأن "الفرص") بالمكسيك، و"حقيبة المدرسة[14]" بالبرازيل. وهناك أدلة من تقييمات الأثر لبرامج التحويلات النقدية المشروطة مثل تحقيق

[11] مصطلح "التحويلات النقدية المشروطة" هي تسمية خاطئة لأن معظم ما يقرب من جميع برامج تحويل والحماية الاجتماعية تفرض شروطا على المستفيدين، ولكن تم تطبيق هذا المصطلح فقط لهذا النوع المحدد من البرنامج.

[12] Food-for-Education Program

[13] PROGRESA (Programa de Educacion, Salud y Alimentacion) program (now called Oportunidades)

[14] Bolsa Escola

منافع غير ضئيلة للأسر الفقيرة، سواء من حيث الدخل الحالي أو المستقبلي، من خلال زيادة الاستثمارات في تعليم الأطفال والرعاية الصحية؛ لمراجعة حديثة للأدلة من التقييمات السابقة أنظر فيزباين وشادي (2009).

الأساس المنطقي لبرنامج التحويلات النقدية المشروطة ليس واضحاً كما قد يتصور المرء. والسؤال منذ فترة طويلة هو لماذا تفرض الشروط، بدلا من القيام بتحويلات غير مشروطة؟. كثيرا ما استشهد بفشل سوق الائتمان كسبب لبرنامج التحويلات النقدية المشروطة، حيث أن الأسر الفقيرة لا تستطيع أن تقترض لتمويل تعليم أطفالهم، ولكن قد تكون هناك سياسات أفضل لمعالجة هذه المشكلة، بما في ذلك التحويلات غير المشروطة (داس وآخرون، 2005). قضايا عدم المساواة داخل الأسرة المعيشية والاقتصاد السياسي غالباً ما تهيمن على المبررات الأخرى بما في ذلك فشل سوق الائتمان.

واحدة من الثغرات الأساسية في المعرفة حول هذه البرامج هو مدى نجاحها في البلدان متدنية الدخل (معظم البرامج والتقييمات تأتي من البلدان متوسطة الدخل، ولا سيما أمريكا اللاتينية). هذه الفجوة المعرفية هي في طور شغلها، بناءاً على برامج التحويلات النقدية المشروطة يجري تقييمها حالياً (ولا سيما في أفريقيا)، مع العديد من النتائج الجديدة المحتملة في السنوات القليلة المقبلة.

نحن بحاجة إلى معرفة المزيد عن تكوين حزمة الأليات للتحويلات النقدية المشروطة، في محاولة لمعرفة كيفية تحديد التكوين الدقيق للتحويلات (كم، وهل كان نقداً أو عيناً) والحوافز لتغيير السلوك (هي سلوكيات مطلوب تشجيعها. هناك بعض الأدلة على أن تحول الميزانية المحايدة من إعانة الالتحاق في "البرنامج التعليمي – الصحة والغذاء" من المرحلة الابتدائية إلى المرحلة الثانوية قد أدت الى تحصيل دراسي أعلى، من خلال زيادة نسبة الأطفال الذين يستمرون حتى المدرسة الثانوية (دي جانفري وسادوليت، 2006). وهذا مثال لوجهة نظري السابقة أن المقيمين في حاجة إلى فهم أفضل بالضبط عن كيفية تحقيق البرامج لآثارها، حتى نتمكن من

تقديم مشورة أفضل لكيفية تحسين التصميم.

وبينما كان هناك بعض البحوث التقييمية الجيدة عن الاستجابات السلوكية، فقد ركزت أكثر على السلوك المقصود من قبل مصممي البرنامج (أي الامتثال مع المسؤولة التشاركية للبرنامج). نحن بحاجة إلى معرفة المزيد عن استجابات أخرى، بما في ذلك عرض العمل وقرارات الادخار من الآباء والأمهات.

وقد تركت الأبحاث السابقة أيضا فجوات في المعرفة حول تأثيرات الرفاهية للتغييرات السلوكية التي تسببها. وينبغي أن تتضمن نتائج كل من الفقر الحالي والمستقبلي. الفقر الحالي هو أسهل للقياس (وإن كان لا يزال ليس من السهل قياسه). ويمكن أن يكون من المفيد التركيز على "النتائج الوسيطة" مثل عمالة الأطفال والحضور في المدارس، ولكننا نعرف أقل عن النتائج النهائية. بالتأكيد، المزيد من الاطفال من الأسر الفقيرة يذهبون إلى المدرسة الابتدائية (بمعنى عندما تخفض تكلفة التعليم من خلال الشروط). لكنهم هل تعلموا شيئاً مفيداً؟ أم أنه فات الأوان للتأثير على قدراتهم التعليمية؟ وماذا عن الآثار المترتبة على أرباح العمل بعد المدرسة؟

أهمية العوامل السياقية المتعلقة بالعرض حيزا كبيرا في برامج التحويلات النقدية المشروطة. كشفت بعض التقييمات نتائج مخيبة للآمال من حيث تعلم الأطفال وحالة التغذية يمكن ربطها بعوامل جانب المعروض – سوء جودة المدارس والعيادات الصحية - والتي تختلف من مكان إلى آخر. التغيرات السلوكية الناجمة عن التحويلات النقدية المشروطة قد لا تكون كافية لتحقيق النتائج المرجوة دون رعاية تحسينات جانب المعروض (فيزباين وشادي، 2009).

فجوات معرفية هامة تتعلق بالآثار طويلة المدى للتحويلات النقدية المشروطة. وهذه بطبيعة الحال تتطلب وقتاً. لكن يجب القيام بالاستثمار في البيانات الآن. للأسف، العوامل الخارجية المعنية، وما يترتب على ذلك من صعوبات شديدة في التمويل ودراسات الحفاظ على الآثار طويلة الأجل.

هناك فجوة أخرى في معرفتنا حول التحويلات النقدية المشروطة تتعلق بمرونتها في تكيف الاستحقاق للتغيرات في الحاجة. وهذا أمر مهم إذا كانت هذه البرامج هي لخدمة وظيفة شبكة الأمان، مثل: أثناء الأزمة المالية العالمية الحالية. يتوقع المرء أنه من الأسهل تنفيذ زيادة للمدفوعات المؤقتة للمستفيدين الحاليين بدلاً من التوسع مؤقتاً في عدد المستفيدين أثناء الأزمة. ولكن، هناك حاجة إلى مزيد من البحوث حول هذه المسألة، وكيف أن برامج التحويلات النقدية المشروطة يمكن أن تكون أكثر مرونة في الممارسة العملية.

أثناء التفكير في الخيارات للتقييم، من المهم أن ننظر إلى مرحلة التخطيط لبرنامج التحويلات النقدية المشروطة. حتى إذا لم يتم تنفيذ البرنامج على أساس تجريبي، فإنه لن يكون دائماً ذو جدوى (لأسباب مالية وفنية) لتنفيذ البرنامج دفعة واحدة على الصعيد الوطني. بعض المناطق الجغرافية قد تحصل على البرنامج قبل مناطق أخرى. المهم فهم أسس اختيار المناطق المستهدفة. البحث عن خيارات لإيجاد مجموعة مراقبة من المراقبة مماثلة للمناطق غير المشاركة. أو إذا كان هناك صيغة لاختيار المناطق المطبق فيها البرنامج أولا، بناءاً عليه، يمكن للمرء النظر في تصميم الانقطاع، الذي يحدد الأثر المرجو في الحي ونقطة نهاية الاستحقاق. مع توسع البرنامج، سوف تنضم منطق المقارنة، ولكن قبل ذلك، قد تكون هناك خيارات لتحديد الآثار. هنا يمكن أن يكون المهم أن نعرف ما إذا كان أهل مناطق المقارنة يعرفون أنهم سوف ينضمون للبرنامج ومتى؟، لأن هذا قد يؤدى الى آثار مغلوطة، حيث تتأثر وحدات المقارنة من قبل التحسب للانضمام إلى البرنامج.

في كثير من النواحي، تقييم "البرنامج التعليمي – الصحة والغذاء" الأصلي هو نموذج لبرامج التحويلات النقدية المشروطة الأخرى. ميزة التصميم الأساسي هي أن المجموعة "العشوائية الخارجة" تصرفت مثل مجموعة التحكم في فترة ادخال البرنامج. طول عمر هذا البرنامج (نجا من التغيرات في الحكومة) ونفوذه في مجتمع التنمية ينبع بشكل واضح من جوهر الجهد الذي بذلوه في تقييمه. اختير ثلث المجتمعات المؤهلة للانضمام للبرنامج بشكل عشوائي كعينات لتشكيل مجموعة المراقبة التي لم تحصل على البرنامج في الفترة الأولية، في

حين تلقت ثلثي المجتمعات الاخرى تدخلات البرنامج. وصول الجمهور إلى بيانات التقييم سهلت عددا من دراسات قيمة. يمكن العثور على لمحة شاملة عن تصميم وتنفيذ ونتائج تقييم "البرنامج التعليمي – الصحة والغذاء" في سكوفياس (2005).

هناك مجالا لتحسين تصميم "البرنامج التعليمي – الصحة والغذاء". كما نوقش أعلاه مطلوب مسألة أكثر اكتمالا لمتغيرات النتائج المرغوب فيها. هناك مخاوف بشأن احتمال تأثر مجموعة المقارنة، لا سيما من خلال الآثار المتوقعة (خصوصا عندما يتم تعميم البرنامج على المستوى الوطني كما في "البرنامج التعليمي – الصحة والغذاء"، وبالتالي يكون له صورة عامة). أيضا، كان يمكن لتصميم التقييم التحدث أكثر عن القضايا الخاصة بتصميم البرامج، مثل ما إذا كان ينبغي أن يكون الدعم في المرحلة الابتدائية أو الثانوية. أيضا أظهرت أبحاث لاحقة أهمية العوامل في جانب العرض، وساعد الاستخدام الأفضل للمسوحات للمنشأة المرتبطة للمدارس والعيادات الطبية.

4. برنامج قطاع النقل

هناك عدد من القضايا الهامة السياقية لابد التفكير فيها عند تقييم برامج قطاع النقل. منطقيا، فإن الخطوة الأولى هي أن نفهم الحاجة إلى التدخل، والتي (في جملة أمور) سوف تتصل بالتاريخ الاقتصادي والجغرافيا والاقتصاد السياسي. أهملت موانئ ووسائل نقل معينة بشكل انتقائي، مما أدى إلى الحاجة لإعادة تأهيلهم. وإذا كان الأمر كذلك، فلماذا؟ أسئلة متعلقة بالمالية العامة من المرجح أيضا أن تلوح في الأفق. كيف سيتم تمويل البرنامج، وهل من المرجح أن يكون هناك آثار تبادلية أو صائدة الذباب (حيث أنه على الرغم من التبادلية الجزئية، فإن المساعدات لا تزال "عالقة" للقطاع المخصص له)؟ قل نحن نقييم التدخل المناسب؟

يمكن التوقع بالآثار على كل من المدى القصير والمدى الطويل. سيكون هناك تأثيرات مؤقتة على دخل العاملين (لكل من العاملين بالمشاريع وغيرهم من العاملين الذين قد يستفيدون من تشديد سوق العمل)، والتي قد

تكون مهمة جدا في بعض السياقات، مثل عندما يكون مشروع النقل هو جزء من الاستجابة لأزمة. سيكون هناك أيضا تكاليف مختلف في مرحلة البناء، بخلاف تكاليف البناء، مثل الآثار الاجتماعية والبيئية (بما في ذلك الأسر المرحلة).

لكن الجزء الأكبر من هذه التأثيرات متوقع بعد الانتهاء من البناء، ولابد من تخطيط جهود التقييم وفقا لذلك في خطة العمل. وهذه التأثيرات يمكن أن تكون واسعة النطاق، ويصعب قياسها كمياً. وفورات في تكاليف النقل (سواء تكاليف غير مخططة وقيمة توفير وقت السفر للأشخاص والشحن) لا شك ستكون ذات أهمية أولى. أما بالنسبة للمشروعات الكبيرة، من المحتمل أن يكون هناك أيضا تأثيرات أوسع على النشاط الاقتصادي(التوازن العام) بما في ذلك الجغرافيا، وسأعود إليه لاحقاً.

التقييم في حاجة إلى النظر في عمل وآثار ذلك على الأسواق. نحن سنريد أن نعرف الى مدى تستفيد السلع المتداولة جغرافيا من الأسواق القائمة، مما قد يثير احتمال لتحديد التأثيرات من خلال الأسعار. (واحدة من الطرق الرئيسية التي تؤثر على الرفاهية بسبب مشروع النقل الرفاه هي التغيرات في الأسعار.) يجب علينا أيضا استكشاف مدى جودة عمل الأسواق الريفية الموجودة في الإطار المحدد، والتي قد توفر مجالا لتحديد التأثيرات من خلال تحرير أسعار الأراضي. (نموذج فون تونن الكلاسيكي يتوقع انخفاض إيجارات الأراضي مع المسافة إلى وسط المدينة، والتي تعكس تكاليف النقل؛ للحصول على مثال لكيفية استغلال هذا النوع من النموذج في تقييم مشاريع النقل (أنظر جاكوبي، 2000).

ربما نرغب أيضا في معرفة كيفية تأثير التدخل على جغرافيا النشاط الاقتصادي، بما في ذلك السوق المحلية والتنمية المؤسسية. هل التحسن في الطرق الريفية، على سبيل المثال، جذب الأسواق والمؤسسات والنشاط الاقتصادي الجديد الى المناطق الفقيرة المعزولة، أو تشجع مزيد من التركيز الجغرافي لهذه الأنشطة؟ كيف غَيّر مشروع النقل الكبير من جغرافية النشاط الاقتصادي؟

نظرا للأهمية الواضحة لتكاليف النقل في اتخاذ قرارات اختيار الموقع، فإن مشروع النقل الكبير من المتوقع أن يغير مشهد النشاط الاقتصادي، وستستفيد منها بعض المواقع والأنشطة، ولكن ربما يؤدي إلى انخفاض الاستفادة لآخرين. تتوقع أن الجغرافيا الاقتصادية الجديدة قد تُخفض تكاليف النقل للمنتجات (الزراعية وغير الزراعية). من خلال الاستثمارات في البنية التحتية للنقل، ستزيد من التركز الجغرافي للأنشطة الأخرى (غير الزراعية) في المناطق الحضرية لاستغلال اقتصاديات التكتل (فوجيتا وآخرون 2001، الفصل 7). بناءاً على إحدى دراسات التقييم لمشروع الطرق الريفية في فيتنام والدعم من البنك الدولي، وجد أدلة على أن المشروع حفز تطويرالأسواق المحلية في المناطق الفقيرة - جزئيا بعدم التمركز للنشاط الاقتصادي؛ أنظر مو وفان دي وال (2008). قد تختلف آثار المناطق الريفية المحلية على التمركز الجغرافي للنشاط عن التأثيرات الإقليمية، بما في ذلك المناطق الحضرية.

احد الأسئلة التي تظهر هو هل ينبغي أن يركز التقييم على نتائج الكفاءة الاقتصادية الكلية فقط. يقال في بعض الأحيان، أن صناعة سياسة التنمية لا ينبغي أن تكون قلقة بخصوص التوزيع المكاني للأنشطة، ولا يجب أن تتدخل في التوزيع الذي تحدده اقتصاديات السوق، في حين تستخدم أدوات سياسة عامة أخرى (غير مهتمة مكانياً) لمعالجة المخاوف بشأن العدالة، بما في ذلك مخاوف من أن يكون لها بُعد جغرافي، مثل المناطق الفقيرة النائية؛ أنظر البنك الدولي (2009) للحجج الخاصة بذلك. بتطبيق هذا النظرية على برامج النقل سنكون مهتمين فقط بكفاءتها، بقياس تأثيرها على متوسط الدخل.

ومع ذلك، فإنني أشك في أن العديد من صانعي السياسات سيقبلوا هذا الرأي، ولأسباب وجيهة. فهم يعرفون أنهم ليس لديهم المجموعة الكاملة للأدوات السياسة المطلوبة للتعامل مع إعتبارات العدالة بطريقة غير مهتمة مكانيا.[15] في الواقع، في البلدان الفقيرة، الموقع هو واحد من الأبعاد المستخدمة على نطاق واسع لتحديد الفقراء

[15] من المتعارف به في الاقتصاد منذ فترة طويلة انه في حاجة الى شروط قوية جدا للفصل الصارم بين اليات العدالة والكفاءة لتحقيق الأهداف العامة للرعاية الاجتماعية. ومن المسلم به على نطاق واسع أن هذه الشروط (لاسيما جدوى التحويلات الكاملة غير المشوهة) لا تتحقق في الممارسة العملية.

لأغراض إعادة التوزيع، بدلاً من معلومات أفضل عن مستويات الرفاهية الفردية. لبعض الوقت في البلدان النامية، يتوقع المرء أن التقييمات مشاريع النقل ذات الصلة بالسياسات سوف تحتاج إلى النظر في جوانب العدالة، وأبعاد أخرى "غير الدخل" (بما في ذلك، على سبيل المثال، الآثار على صحة الأطفال من خلال تحسين فرص حصول الأمهات على العناية الصحية اللازمة بسبب تحسن النقل).

تصميم التقييم لبرنامج النقل سوف يحتاج للتمييز بين المكونات التي يمكن أن تُعامل على أنها برامج مستهدفة (مثل الطرق الريفية المحلية)، وتلك التي من المحتمل أن يكون لها تأثيرات مكانية متناثرة (الطرق الرئيسية على سبيل المثال)، وهو سيتطلب نهج خاص بالقطاع أو على مستوى الاقتصاد الكلي.

تقييم المكونات المستهدفة وغير المستهدفة تتطلب تصاميم مختلفة للمكونات المستهدفة، مثل تحسين الطرق الريفية المحلية، سوف نرغب في التعرف على "منطقة التجمعات" ذات الصلة التي تؤمن تحقق التأثيرات المحتملة (فان دي وال، 2009). ونحتاج، أيضا، إلى فهم كيفية تحديد الموقع الجغرافي، وإيجاد مناطق مماثلة كمناطق مراقبة لا تتلقى إنفاق مثل اللازم للطرق الريفية الجديدة أو إعادة التأهيل. من شبه المؤكد أن الملاحظات على المدى البعيد طولية ستكون مطلوبة (للسماح بقدر من تحديد "الفرق في الاختلاف"؛ انظر رافاليون 2008، لمزيد من المناقشة وأمثلة). لابد للمقيمون، أيضا، تقييم المصادر المحتملة للتحيز في الاختيار. إذا كان الاختيار قائما على الملاحظة بصفة كبيرة، فمن الممكن جمع البيانات الصحيحة لتصحيح هذا التحيز. إذا كان الاختيار غير قابل للملاحظة، سوف يكون من الضروري النظر في نطاق الأساليب الإحصائية مثل تقدير اليات المتغيرات. (التوزيع العشوائي ونادرا ما يكون ذو جدوى لأسباب واضحة.)

إذا كان مشروع النقل كبير، فبالتالي ستكون منطقة التجمعات، وبالتالي سيكون من الصعب العثور على مناطق مقارنة جيدة، لم تتأثر من قبل بالبرنامج. سوف نحتاج إلى اللجوء إلى أدوات نوعا ما مختلفة.[16] في اقصى الطرف (من حيث التجمع) تم استخدام انحدارات النمو العام للدولة في محاولة لتحديد الآثار المترتبة على البنية

[16] يمكن ايجاد نظرة عامة عن خيارات المنهجيات في أوسترهافن وكناب (2003).

التحتية، بما في ذلك النقل؛ انظر، على سبيل المثال، كالديرون وسيرفين (2008). البيانات الجغرافية الوطنية الفرعية المرتبطة بالوحدات السكنية والشركات يمكن أن توفر نظرة أدق؛ انظر، على سبيل المثال، جالان ورافاليون (2002) حيث وجدوا أن الطرق الريفية في الصين أثرت إيجاباً على عملية نمو الاقتصاد الجزئي. المواقع المحددة بناءاً على النماذج العامة المتوازنة، والمرتكزة على الجغرافيا الاقتصادية الجديدة[17]، قد تلعب أيضا دورا هاما في المستقبل (فوجيتا وآخرون، 2001).

هذه النماذج مناسبة تماماً من حيث المفهوم لمشكلة تقييم آثار مشاريع النقل الكبيرة، على الرغم من أنها هي أيضا نماذج معقدة وتتطلب الكثير من البيانات والمعايرة خاصة بها. وقد طور الاتحاد الأوروبي نموذج فريق الخبراء الاستشاري (CGE) المكاني لتقييم التحسن في النقل؛ وذلك للتطبيق في سياق تقييم مشروعات النقل الكبيرة (مشروع السكك الحديدية في هولندا)، أنظر أوسترهافن وكناب (2000).

يجب الاعتراف بأن النماذج القديمة لجغرافيا الاقتصاد الجديدة جعلت بعض الإفتراضات الخاصة بنماذج تكلفة النقل غير قابل للتصديق، وهذه الإفتراضات قد تلعب أيضا دوراً في الدروس المستقاة للآثار من التحسينات في النقل.[18] يمكن أن يتوقع في المستقبل مزيد من التقدم في بناء النماذج التشغيلية بناءاً على إفتراضات أكثر واقعية.

5. لجنة مكافحة الفساد

هي هيئة مستقلة لها سلطة غير عادية للتحقيق والإدعاء في الفساد، وعادة ترسل التقارير لرئيس الدولة أو البرلمان.[19] سؤالان رئيسيان في هذا السياق لابد النظر فيهما في تقييم لجنة مكافحة الفساد. أولاً، هناك حاجة إلى تقييم ما إذا كان التدخل يعالج الأسباب المعروفة للفساد. هل الأسباب الجذرية للفساد تقع في آليات التطبيق أو في الحوافز التي تواجه الجهات السياسية بالتغاضي عن الإشراف على الموظفين العموميين؟ وهل يجب

[17] New Econimic Geography (NEG)

[18] أود أن أشير هنا إلى "وظايفة تكلفة النقل الفائضة" وهو ما يعني ضمنا أن أسعار تسليم البضائع تزيد أضعافا مضاعفة مع المسافة التي يتم شحنها، وهو ما يتعارض مع الأدلة؛ لمزيد من المناقشة أنظر ماكان (2005).

[19] يمكن العثور على لمحة عامة مفيدة عن تاريخ وسجلات عن لجنة مكافحة الفساد في روز-أكرمان (1999، فصل 9).

علينا أن نتحدث بدلاً منه عن إصلاح المؤسسات الائتمانية، وخاصة إدارة القطاع العام؟. ثانياً، لابد أن يكون واضحاً للمقيمين الحوافز لهذا التدخل. هل هي مجرد رغبات للجهات المانحة للعمل على معاقبة المسؤولين الفاسدين، أو زعيم سياسي حريص على قمع المعارضة، أو هل تم التدخل بناءاً على تحليل أعمق لهذه المشكلة؟

يمكن تصميم التقييمات لتقييم التأثير على الفساد بناءاً على أنواع الأدوات المستخدمة من قبل لجنة مكافحة الفساد. يمكن العثور على نموذج مبتكر في أولكن (2007) لدراسة الفساد في مشاريع الطرق باندونيسيا باستخدام تصميم عشوائي. وجد أولكن أن زيادة المراجعات الحكومية يؤدي الى تقليل "النفقات المفقودة" قياساً بالفرق بين تكاليف المشاريع المسجلة رسميا والتقديرات من قبل المهندسين المستقلين. على النقيض من ذلك، وجد أولكن (2007) أن النهج التشاركي للرصد ليس له تأثير يذكر.

في حين أن مثل هذه الدراسات التقييمية يمكن أن توفر مدخلات مفيدة، فهي لا تحدد تأثير لجنة مكافحة الفساد. دراسة أولكن تخبرنا أن الرصد من قبل مدققي حسابات الحكومة يمكن أن تساعد في مكافحة الفساد بإندونيسيا. ومع ذلك، فدراسة تقييم واحدة (قدتكون مؤرخة)، لا تظهر نوعية الجهود الفعالة الخاصة بلجنة مكافحة الفساد بإندونيسيا في تسهيل مثل هذه المراقبة والعقاب، بخلاف بعض الاعتقالات لأفراد فاسدين معروفين مرتبطين بالنظام السياسي السابق، والتي حظيت بتغطية إعلامية (شيرلوك، 2002).

لجنة مكافحة الفساد الموذجية ليست برنامج مستهدف مثل التحويل النقدي المشروط أو بعض مكونات برنامج النقل التي نوقشت أعلاه. قد يكون هناك بعض المجال لزيادة التغطية الجغرافية و / أو وزارية تدريجياً، عن طريق تكليف بعض سلطات لجنة مكافحة الفساد إلى بعض الحكومات أو الوزارات المحلية وليس غيرهم. إذا كان التدخل التدريجي ممكناً، والآثار المتوقعة غير محتمل أن تكون شديدة جداً، فمقارنة نتائج إختبارات الفساد الموضوعية (مثل التي يستخدمها أولكن 2007)، بين مراقبة وكالات حكومية مماثلة (في المناطق الحكومية المحلية المختلفة)، بالاختلاف أن بعضها تخضع للجنة مكافحة الفساد ولكن البعض الأخر ليس خاضعاً، قد

يساعد ذلك في تقييم تأثير لجنة مكافحة الفساد. ومع ذلك، وبعيدا عن الجدوى (الحكومة قد تفضل أيضا إعطاء لجنة مكافحة الفساد العنان من البداية)، وهناك مخاوف جدية حول آثار الترقب والتحيزات في الاختيار. (علامة التحيز في الاختيار غامضة دون معرفة المزيد عن الإعداد، رئيس ملتزم بمكافحة الفساد يفترض أن يختار الأماكن أو القطاعات التي تعتبر المشكلة الأعظم؛ والرئيس الذي يستخدام لجنة مكافحة الفساد كأداة سياسية سيختار بشكل مختلف، وقد يتجنب مراكز الفساد).

نهج أكثر واعد (وربما مكمل) هو النظر في القنوات المحتملة لتأثير لجنة مكافحة الفساد، والمبني على نموذج نظري أكثر أو أقل وضوحاً لسبب وجود الفساد. ثم يمكن لجهود التقييم التركيز على ما يبدو أنه الروابط الرئيسية في السلسلة السببية المتوقعة، على نحو مماثل لنوع من "التقييم القائم على نظرية" نوقشت في فايس (2001)، روجرز (2009) ووايت (2009). على سبيل المثال، جادل كليتجارد (1988) بأن الفساد هو نتيجة لثلاثة عوامل: احتكار بعض الموارد، والسلطة التقديرية للمسؤولين في السماح للوصول إلى تلك الموارد، وغياب أو فشل آليات مساءلة هؤلاء المسؤولين. ويمكن للمرء بعد ذلك، أن يبدأ تقييم مدى تناول لجنة مكافحة الفساد لكل عنصر من هذه العناصر، عن طريق تغيير الحوافز التي تواجه المسؤولين في بيئة مؤسسية محددة. سوف يتطلب ذلك تقييم ما تواجه الحوافز المسؤولين الأفراد وكيف تغيرت مع لجنة مكافحة الفساد. هذا لن يكون سهلاً، لأنه قد يكون هناك بعض الحوافز التي تكشف الحقيقة للمقيّم. ولكن على الأقل عن طريق بناء في جهود التقييم لفهم السبب في أننا نرى الفساد في المقام الأول (الذهاب بعمق أكبر من الطعون إلى "فجور" المسؤولين) يمكن أن يكون للمرء بعض الفرص لتحديد ما إذا كان للجنة مكافحة الفساد أي تأثير حقيقي على المشكلة.

أعضاء لجنة مكافحة الفساد مغرمون بتقديم عدد من المسؤولين للمحاكمة كمقياس للنجاح. هذا يمكن أن يكون خادعاً، نظرا لإمكانية أن عمل لجنة مكافحة الفساد قد ينطوي على الاستهداف الانتقائي للخصوم السياسيين. وإدراكاً لهذا، فإن التقييم قد يركز بشكل مفيد مباشرة على الدعم السياسي أو المعارضة للجنة مكافحة الفساد.

يمكن جمع بيانات عن الانتماء السياسي لأعضاء لجنة مكافحة الفساد، جنبا إلى جنب مع استطلاعات الموظفين على الاتصالات مع السياسيين، وأسباب مغادرة الموظفين غادرت. سيتم بعد ذلك بحاجة إلى عمليات مسح المتابعة (بعد المسح الأساسي)، وطلب الأسر والشركات والمسؤولين الحكوميين حول مستوى تسييس لجنة مكافحة الفساد مقارنة مع مؤسسات أخرى مثل البنك المركزي، وزارة المالية ومكتب المدعي العام.

6. **الخاتمة**

فن التقييم الجيد هو أن نسأل الأسئلة الصحيحة في البداية، بناءاً على الفجوات المعرفية القائمة وتفصيل البيانات وتحليلها للإجابة على تلك الأسئلة في سياق معين. لا يمكن للمرء أن يتوقع كل الأسئلة الأكثر أهمية لتقييم المحدد في ورقة مثل هذه، أو توقع كل الخيارات لأساليب التقييم، وكلاهما سوف يعتمد بطبيعة الحال على السياق المحدد. ومع ذلك، نأمل بأن تكون هذه الورقة قد قدمت بعض نقاط الانطلاق المفيدة للتفكير.

المراجع:

Calderon, Cesar and Luis Servén, 2008, Infrastructure and Economic Development in Sub-Saharan Africa, Policy Research Paper 4712, World Bank, Washington DC.

Chen, Shaohua, Ren Mu and Martin Ravallion, 2009, Are There Lasting Impacts of Aid to Poor Areas? Evidence from Rural China, *Journal of Public Economics* 93: 512-528.

Das, Jishnu, Quy-Toan Do and Berk Ozler, 2005, A Welfare Analysis of Conditional Cash Transfer Schemes, *World Bank Research Observer*, 20(1): 57-80.

De Janvry, Alain and Elisabeth Sadoulet, 2006, Making Conditional Cash Transfer Programs More Efficient: Designing for Maximum Effect of the Conditionality, *World Bank Economic Review* 20(1): 1-29.

Fiszbein, Ariel and Norbert Schady, 2009, *Conditional Cash Transfers for Attacking Present and Future Poverty,* World Bank Policy Research Report, World Bank, 2009.

Fujita, Masahisa, Paul Krugman and Anthony Venables, 2001, *The Spatial Economy*, Cambridge, Mass.: MIT Press.

Jacoby, Hanan, 2000, Access to Markets and the Benefits of Rural Roads, *Economic Journal* 110: 713-737

Jalan, Jyotsna and Martin Ravallion, 2002, Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China, *Journal of Applied Econometrics* 17(4): 329-346.

King, Elizabeth M. and Jere R. Behrman, 2009, Timing and Duration of Exposure in Evaluation of Social Programs, *World Bank Research Observer* 24(1): 55-82.

Klitgaard, Robert, 1988, *Controlling Corruption*, University of California Press, Berkeley CA.

Knaap, Thijs and Jan Oosterhaven, 2000, The Welfare Effects of New Infrastructure: An Economic Geography Approach to Evaluating New Dutch Railway Links, mimeo Erasmus University Rotterdam.

McCann, Philip, 2005, Transport Costs and the New Economic Geography, *Journal of Economic Geography* 5: 305-318.

Mu, Ren and Dominique van de Walle, 2008, Rural Roads and Local Market Development in Vietnam Policy Research Working Paper, World Bank.

Oduor, Jacob, Anne Kamau, and Evan Mathenge, 2009, Evaluating the impact of microfranchising the distribution of anti-malarial drugs in Kenya on malaria mortality and morbidity , *Journal of Development Effectiveness*, this volume.

Olken, Benjamin A., 2007, Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy* 115(2): 200-249.

Oosterhaven Jan and Thijs Knaap, 2003, Spatial Economic Impacts of Transport Infrastructure Investments, in *Transport projects, programmes, and policies (edited by*

A. D. Pearman, Peter J. Mackie, John Nellthorp) Interdisciplinary Centre for Comparative Research in the Social Sciences, Ashgate Publishers.

Ravallion, Martin, 2008, Evaluating Anti-Poverty Programs. In Paul Schultz and John Strauss. eds., *Handbook of Development Economics Volume 4*, Amsterdam: North-Holland.

_____, 2009, Evaluation in the Practice of Development, *World Bank Research Observer* 24(1): 29-54.

Rogers, Patricia J., 2009, Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation, *Journal of Development Effectiveness*, forthcoming.

Rose-Ackerman, Susan, 1999, *Corruption and Government: Causes, Consequences and Reform*, Cambridge University Press, Cambridge.

Sherlock, Stephen, 2002, Combating Corruption In Indonesia? The Ombudsman And The Assets Auditing Commission, *Bulletin of Indonesian Economic Studies,* 38(3): 367–83.

Skoufias, Emmanuel, 2005, *PROGRESA and Its Impact on the Welfare of Rural Households in Mexico*, Research Report 139, International Food Research Institute, Washington DC.

van de Walle, Dominique, 2009, Impact Evaluation of Rural Road Projects, *Journal of Development Effectiveness*, 1(1): 15-36.

Weiss, Carol, 2001, Theory-Based Evaluation: Theories of Change for Poverty Reduction Programs. in Osvaldo Feinstein and Robert Piccioto, eds, *Evaluation and Poverty Reduction*, New Brunswick, NJ: Transaction Publications.

White, Howard, 2009, Theory-based Impact Evaluation: Principles and Practice, *Journal of Development Effectiveness*, this volume.

World Bank, 2009, *Reshaping Economic Geography*, World Bank, Washington DC.

مطابقة تصميم تقييم الأثر لطبيعة التدخل والغرض من التقييم

باتريشيا روجرز، أستاذ تقييم القطاع العام في "CIRCLE" (التعاون للبحوث متعددة التخصصات، استشارات والتعلم في التقييم) بمعهد ملبورن الملكي للتكنولوجيا، ملبورن، أستراليا[20].

ملخص

التصميم المناسب لتقييم الأثر إستجابة للواقع يتطلب مطابقة التصميم للإحتياجات والقيود والفرص الخاصة بقضية معينة. لابد للتصميم أن يعكس طبيعة التدخل والغرض من تقييم الأثر. يحتاج تقييم الأثر، على وجه الخصوص، إلى معالجة الجوانب البسيطة والمعقدة والمركبة للتدخل. الجوانب البسيطة يمكن أن تكون محددة بإحكام وموحدة ؛ تعمل الجوانب المعقدة كجزء من حزمة السببية؛ تكون الجوانب المركبة ديناميكية وتكيف بشكل مناسب. ينصح بتصاميم مختلفة لكل حالة، بما في ذلك التجارب العشوائية المخططة، انقطاع الإنحدار، والمقابلات المجتمعية غير المنظمة، تقارير قصة الأداء التشاركية، والتقييم التنموي.

نهج الاستجابة للواقع في تقييم الأثر

كان الهدف من هذه الجلسة بالمؤتمر عرض لتطبيق مناهج مختلفة لتصاميم تقييم التأثير. في حين أن العروض ركزت بشكل خاص على تحليل القياس والتحليل السببي، يجدر بنا أن نتذكر أن هناك مهام أخرى يجب أن يعالجها تقييم الأثر (روجرز، 2008a). يحدد التصميم الشامل لتقييم الأثر كيف أن التقييم سوف يؤدي الى مجموعة كاملة من المهام التي تدخل في تقييم الأثر، والتي هي "تحديد شامل للآثار الهامة؛ الجمع المنهجي للبيانات والدفاع عنها، وتحليل أدلة هذه الآثار؛ إستنتاجات سليمة حول مساهمة التدخل في تحقيق هذه الآثار؛ والإدارة الفعالة للتقييم، بما في ذلك الإبلاغ بشفافية عن المنهجية؛ وعند الملائمة، القيام بالتقييم الفوقي الرسمي

[20] للأتصال: Patricia.Rogers@rmit.edu.au 124 Latrobe Street Melbourne VIC 3000 Australia

(شبكة شبكات تقييم الأثر[21] - المجموعة الفرعية 2، 2008). لكل من هذه المهام، هناك مجموعة من الخيارات، وتصميم تقييم الأثر يحتاج إلى إختيار الطريقة الأكثر مناسبة أو الجمع بين أكثر من أسلوب مناسب لكل مهمة.

تركز بعض الأساليب لتقييم الأثر على استخدام تصميم أو طريقة معينة لجمع البيانات وتحليلها أو شكل من أشكال الحوكمة. أفضل وصف لتوجهي لتصميم تقييم الأثر يمكن أن يوصف بالـ"الاستجابة للواقع" (باتون، 2008a).

هناك إعتراف متزايد في التنمية وفي قطاعات أخرى من التقييم، أن حالات التقييم المختلفة يمكن معالجتها بصورة أفضل من خلال التخطيط المناسب من مجموعة من الأساليب والتقنيات. على سبيل المثال، جاء في بيان شبكة شبكات تقييم الأثر[22] بشأن تقييم الأثر في عام 2008 أن:

" تدعو شبكة شبكات تقييم الأثر إلى نهج إنتقائي ومفتوح لإيجاد أفضل الطرق للقيام بمهمة تقييم الأثر - بالإعتماد على مجموعة واسعة من التقنيات المتاحة من مختلف التخصصات" (شبكة شبكات تقييم الأثر، 2008)

في محاضرة عن "أحدث فنون قياس فعالية التنمية" بمؤتمر البنك الدولي (مجموعة التقييم المستقلة) "قياس فعالية التنمية: التقدم المحرز والعقبات"، لخص مايكل كوين باتون هذا الأمر بأنه بمثابة إعادة صياغة ما نفهمه تحت "المعيار الذهبي" لتقييم الأثر "إن المعيار الذهبي للمنهجية هنا هو الملائمة، وليس فقط أسلوب معين" (باتون، 2008b).

[21] شبكة شبكات تقييم الأثر – Network of Networks for Impact Evaluation (NONIE)

[22] شبكة شبكات تقييم الأثر تتألف من منظمة التعاون الاقتصادي ولجنة التنمية للمساعدة الإنمائية (OECD / DAC) شبكة التقييم، ومجموعة التقييم التابعة للأمم المتحدة (UNEG)، وفريق تعاون التقييم (ECG)، والمنظمة الدولية للتعاون في مجال التقييم (IOCE)، وهي شبكة تكونت من اتحادات التقييم الإقليمية.

تتضمن الاستجابة للواقع مطابقة التصميم للإحتياجات والقيود والفرص لوضع معين.وبالتالي، فإن إثنين من الأسئلة الرئيسية التي لابد من الإجابة عليها قبل وضع تصميم تقييم الأثر هما "ما هي طبيعة التدخل؟" و"لماذا سيتم اجراء تقييم الأثر؟".

ما هي طبيعة التدخل؟

نوع التدخل وحجمه من المحددات الهامة لتصميم تقييم الأثر المناسب. هل هو مشروع صغير يتم تجريبه لإحتمالية التكرار والتوسع فيه؟، هل هو برنامج مستمر ومن المرجح أن يستمر بشكل ما؟، هل هو تدخل موحد أو مجموعة من المبادرات المختلفة؟ وهل التدخل محدد بإحكام وموحد أو أنه يختلف في المواقع المختلفة إستجابة للظروف والإحتياجات والفرص المحلية؟. هذه الأسئلة لها إنعكاسات على نوع تقييم الأثر المطلوب، وكذلك مدى التوفر المحتمل للموارد المطلوبة لتقييم الأثر، والخيارات من حيث تصميم البحوث.

ما هي طبيعة الآثار التي نبحث عنها؟، هل كانت تنتج مباشرة بسبب التدخل (مثل*لطخة*) أو غير مباشرة (مثل الرزاز)؟، هل هي أثار قصيرة الأجل سوف تكون واضحة خلال عمر المشروع والتقييم (مثل أداء الأطفال في المدارس)، أو آثار على المدى الطويل والتي ستظهر في وقت لاحق بعد عدة سنوات (مثل فرص العمل في مرحلة ما بعد المدرسة، أو تقليل ملوحة التربة)؟، هل هي آثار تحويلية، التي اذا تحققت مرة من غير المرجح أن ترتد (مثل تعلم كيفية القراءة أو ركوب الدراجات)، أو آثار هشة يمكن الإرتداد عنها بسهولة (مثل التغذية الكافية أو انضمام الإناث للمدرسة)؟ وهل التأثيرات المحتملة تكون نتيجة تدخل "رصاصة فضية"، التي تحقق النتائج بغض النظر عن السياق، أو "تدخل اصطفاف البط"، الذي يحقق النتائج فقط بالإقتران مع الظروف المواتية، قد تشمل ايضاانواع أخرى من التدخلات.

يمكن تلخيص هذه الخصائص المختلفة من حيث التصنيف الى ثلاثة أجزاء: بسيطة ، معقدة ، ومركبة (ستايسي 1992؛ جلوبرمان، 2001؛ جلوبرمان وتسيمرمان 2002 وكورتز وسنودن، 2003). وقد تبين أن هذا سيكون مفيداً عند التخطيط وتحليل التقييم (جويجت، 2008؛ باتون، 2008a وروجرز، 2008b). هذا التصنيف مفيد بشكل خاص عندما يتم إستخدامه لتصنيف جوانب التدخلات بدلاً من التدخل كله.

في هذا التصنيف، فإن مصطلح "مركب" له معنى محدد ومهم، وهو ليس دائماً المتعارف عليه في الإستعمال الشائع. مصطلح التقييم "مركب" يُستخدم أحياناً كمرادف ل"معقد"، ويُستخدم أحياناً للإشارة إلى أي شيء صعب، وأحياناً يُستخدم كذريعة لعدم ملائمة التخطيط. في هذا التصنيف فإن "مركب" تشير إلى الجوانب الديناميكية للتدخلات، التي عليها التكيف والإستجابة للإحتياجات والفرص الناشئة. الجوانب البسيطة للتدخلات يمكن تحديدها بإحكام وتكون موحدة - على سبيل المثال: منتج أو تقنية أو عملية معينة. الجوانب المعقدة للتدخلات لها مكونات متعددة، وهي جزء من تدخل أكبر متعدد المكونات، أو تعمل بشكل مختلف كجزء من حزمة سببية أكبر، على سبيل المثال في بيئات تنفيذ معينة، لفئات معينة من المشاركين، أو بالإقتران مع تدخل آخر.

الجوانب المختلفة للتدخلات لها آثار هامة على كيفية عمل التدخلات، كيف يمكننا أن نفهمها، وكيف يمكننا إستخدام هذا الفهم، كما هو مبين في الجدول (1).

الجدول (1): الآثار المترتبة على الجوانب البسيطة والمعقدة والمركبة للتدخلات.

الجوانب	**الآثار المترتبة على:**		
	تنفيذ التدخل	**عمليات سببية**	**إعداد التقارير واستخدام نتائج تقييم الأثر**
بسيطة	منظمة واحدة	سلسلة سببية واحدة لازمة	رسالة واحدة - ماذا يعمل

		لتحقيق الآثار	
معقدة	منظمات متعددة في علاقة تعاقدية مع أدوار محددة بوضوح	سلاسل سببية متعددة لازمة لتحقيق الآثار: تدخلات متعددة متعاقبة أو تدخلات متعددة في وقت واحد أو مستويات متعددة من التدخل أو آليات سببية مختلفة تعمل في سياقات مختلفة	رسالة طارئة - ماذا يصلح لمن في أي من الحالات
مركبة (ديناميكية وناشئة)	منظمات متعددة في تطوير علاقة شراكة	يمكن ارداد السبب، مع حلقات ردود الفعل النتائج طارئة - الكل هو أكثر من مجموع الأجزاء	رسالة ديناميكية، ناشئة - ماذا يتم عمله

لماذا يتم تقييم الأثر؟

لابد منالنظر الى الغرض من تقييم الأثر عند وضع تصميم تقييم الأثر. من هم المستخدمون المستهدفون من التقييم؟، ماذا سيعتبرونه كأدلة موثوق بها من حيث الآثار التي ينبغي إدراجها، وأساليب القياس التي ستستخدم والنهج المتبع في التحليل السببي؟ ومن الذي مطلوب منه أن يُشارك في تحديد معالم التقييم؟.

سيتم إستخدام قيم مَنْ في التقييم؟، ماذا سيعتبر تأثيرات كبيرة، إما إيجابية أو سلبية؟، ماذا سيعتبر التوزيعات المرغوبة للتكاليف والفوائد؟، هل سيكون التركيز على متوسط أثر، أو التأثير على الأكثر حرماناً؟ ما هي الإستخدامات المقصودة من التقييم؟، هل يجري القيام به لتبرير الإنفاق بأثر رجعي؛ وفي هذه الحالة ستكون التقديرات موثوقة للفائدة الصافية كافية؟ أم أنه يتم إجراؤه للإبلاغ عن التوسع الممكن للتجربة - وفي هذه الحالة، سيكون هناك حاجة إلى معلومات جيدة عن كيف يعمل؟، هل يتم بغرض الإبلاغ عن التغيير التدريجي أو إعادة صياغة لبرنامج أو سياسة؟.

وأخيراً، قبل تصميم تقييم الأثر، لابد من معالجة المسائل اللوجستية. متى لابد الانتهاء من التقرير؟، ما هي الأدلة المتاحة بالفعل عن هذا التدخل؟، وماذا عن تدخلات مماثلة؟ وما هي الموارد الإضافية المتاحة للقيام بتقييم الأثر؟.

عند إستخدام نهج الاستجابة الواقعية، عندما يكون لدينا فقط إجابة لكل هذه الأسئلة، يمكن معالجة السؤال "كيف ينبغي أن يتم ذلك؟". من الواضح، أنه في التقييم الحقيقي، سيتم الإجابة على كل هذه الأسئلة قبل، أو أثناء عملية وضع التصميم. في عملية التصميم هذه، نقوم بملء الثغرات في وصف الحالات لإعداد سيناريوهات أكثر تحديداً.

1. التحويلات النقدية المشروطة

وقد وصفت هذه الحالة على النحو التالي:

التحويلات النقدية المشروطة في أحد بلدان أمريكا الوسطى، حيث تتلقى الأسر المبلغ الشهري إذا بقيت الإناث في سن التعليم في المدرسة، وتلبية متطلبات الحضور والأداء المحددة.

لدى برامج التحويلات النقدية المشروطة بعض الجوانب الهامة التي من الأفضل وصفها بأنها بسيطة - وهي، أنها منفصلة، وتدخلات موحدة يعتزم تنفيذها بنفس الطريقة في مواقع مختلفة. ومع ذلك، أظهر تقييم برامج التحويلات النقدية المشروطة عدم تجانس كبير للنتائج (على سبيل المثال: سواريس وآخرون، 2009). قد تحتاج التحويلات النقدية المشروطة إلى حزمة من التدخلات الأخرى لتحقيق النتائج المرجوة. على سبيل المثال: قد تكون التحويلات النقدية المشروطة ناجحة في تحقيق الهدف الأولي من الالتحاق بالمدارس ولكن التحسينات في الآثار على المدى الطويل مثل نتائج التعلم ، والتخرج والعمل للطلاب سوف يعتمد، أيضا، على

وجود مدارس فعالة. ولذلك، سيكون من المهم النظر في الجوانب المعقدة من التدخل، وشمول قياس وتحليل عناصر أخرى (مثل ميزات معينة لبيئة التنفيذ وخصائص المشاركين) اللازمة لتحقيق الآثار المرجوة.

بإفتراض أنه في هذه الحالة يكون الغرض من تقييم الأثر هو أن يقرر ما إذا كان التوسع في تطبيق برنامج تجريبي، فان التجربة العشوائية المخططة[23] قد تكون التصميم المناسب للتحليل السببي، خاصة إذا كانت مقرونة بالجمع المنهجي للبيانات عن عوامل أخرى، مثل نوعية التعليم، التي قد يكون هناك حاجة اليها لتحقيق الآثار المقصودة، والتحليل، مثل التصنيف، لتحديد مساهمتها. تطوير واختبار نظرية البرنامج التي شملت عوامل إضافية، وآثار تفاضلية، وكذلك النتائج الوسيطة، من شأنه أن يُحسِّن نوعية التحليل. إذا كانت التحويلات النقدية المشروطة فعالة في تحسين الأداء المدرسي، ولكن فقط مقرونة بالمدارس الفعالة، فلن يكون هناك حل ذاتي سريع. كبديل وحيث أن برامج التحويلات النقدية المشروطة مناسبة لمستوى معين من الدخل، فقد يوفر نهج إنقطاع الإنحدار أدلة دامغة على إسناد السببية على نفقة أقل من التجربة العشوائية المخططة، على الرغم من أنها قد تحتاج أيضا إلى تصنيف حسب بيئة التنفيذ.

من المهم أن يتضمن تقييم الأثر بالإضافة إلى قياس الآثار المقصودة تأثيرات كبيرة أخرى، بناء على تعريف لجنة المساعدة الإنمائية[24] للأثر كـ"إيجابي وسلبي، آثار ابتدائية وثانوية على المدى البعيد ناتجة عن التدخل الإنمائي، بشكل مباشر أو غير مباشر، مقصودة أو غير مقصودة".

يمكن توقع بعض الآثار المحتملة غير المقصودة وإدراجها في جمع البيانات المستهدفة، وذلك بإستخدام "نظرية البرنامج السلبية" (فايس، 1997). من المرجح أن يكون ملاحظة التأثيرات غير المتوقعة، خاصة التأثيرات السلبية، من خلال الأساليب التشاركية (تشامبرس، 2009)، بناءاً على تكرار المقابلات والملاحظات ومراجعة

[23] Randomized Control Trial
[24] Development Assistance Committee (DAC)

الوثائق قد يكون هناك حاجة الى دراسات الحالة للكشف عن الآثار غير المقصودة (إيجابية أو سلبية)، تتبعها عمليات مسح واسعة النطاق أو مراجعة الإحصاءات الإدارية بهدف إظهار تقديرات تكرار هذه الآثار.

2. برنامج البنية التحتية للنقل

وقد وصفت هذه الحالة على النحو التالي:

هو "التقييم اللاحق لبرنامج قطاع النقل في بلد بجنوب آسيا، والذي يشمل إعادة تأهيل ميناء وإعادة تأهيل الطرق الرئيسية، والإستثمارات الجديدة في الطرق الفرعية الريفية".

إننا نفترض في هذه الحالة أن الغرض من تقييم الأثر كان لفهم الأثر الكلي للإستثمار، وفي المقام الأول لتقديم التقارير إلى الممولين، ولكن أيضا للمجتمع، وأنه لا يوجد تقييم قد تم تصميمه في بداية البرنامج. يظهر هذا النقص، بجانب طبيعة التدخل (وهو متعدد الأوجه، ومتنوع، ويؤثر على المنطقة بأسرها) تحديات كبيرة للتحليل السببي، حيث أن الأساليب التجريبية وشبه التجريبية الكلاسيكية من المحتمل أن يكون من المستحيل أن تنفذ بفعالية (رافايلون، 2009).

يحتاج تقييم تأثير هذه الحالة الى وسيلة لجمع أدلة حول عدد كبير من المكونات المختلفة معا، ومعالجة مسألة الإسناد (الإيعاذ) دون مجموعة مقارنة ومراقبة. قد يكون من المناسب في هذه الحالة إستخدام نهج جديد لتقارير قصص الأداء التشاركية[25]، والتي عرضت في هذا المؤتمر، (دارت، 2009) والتي يمكن أن تؤدي هذه الوظائف. يفهم بعض الناس أن مصطلح "تشاركي" على أنه تطبيق نهجاً أقل صرامة لجمع وتحليل البيانات؛ لذلك، من المهم أن نفهم القاعدة التجريبية القوية جداً التي تكمن وراء هذا النهج.

[25] Participatory Performance Story Reporting (PPSR)

تعتبر تقارير قصص الأداء التشاركية هي تطوير وتنظيم للأدلة متعددة الخطوط والمستويات[26]، وهو نهج للتحليل السببي مصمم للحالات عالية المخاطر، حيث يكون فيها بناء مجموعة مراقبة أو مجموعة مقارنة غير ممكن، ولكن يتوفر أدلة كثيرة عالية الجودة، والتي تتطلب تحليل سببي منهجي.

"خط الأدلة هو نوع من الأدلة، مثل سمة النظام البيئي (مثل وفرة الأسماك وثراء الكائنات اللافقارية متناهية الصغر، والكتلة الحيوية للنباتات الصغيرة)، ويتم التحقق منه فيما يتعلق بالضغوط أو تدخل بشري. مستوى الأدلة هو قيمة قوة الأدلة المستخدمة لتحديد الحالة لإستنتاج أن النشاط البشري يؤدي إلى تغيير البيئة أو الجيومورفولوجي." (كوتنجهام وآخرون، 2005)

لم يظهر هذا النهج في معظم الأدلة الارشادية لتقييم الأثر، ولكن إستخدامه يعود إلى الإستراتيجيات المستخدمة في الستينات لتحليل الأدلة حول العلاقة بين التدخين وسرطان الرئة. كان من المهم التمكن من إستخلاص النتائج حول تأثير التدخين على الصحة مع وجود بيانات علمية كبيرة، ورغم عدم وجود أدلة من التجربة العشوائية المخططة. تم إستخدام الأدلة متعددة الخطوط والمستويات في تقييم المخاطر البشرية والبيئية وإدارة الموارد الطبيعية في السنوات الأخيرة (على سبيل المثال: كيو وآخرون، 2002؛ بويز، 2006).

تشمل الأدلة متعددة الخطوط والمستويات على التحقيق بشكل منهجي قوة الحجة السببية التي تربط تدخل أو سبب بآثاره، وذلك من خلال تحليل مجموعة مراقبة من حيث معايير سببية معينة، وعن طريق تحديد وإستبعاد تفسيرات بديلة ممكنة. للربط بين التدخين وسرطان الرئة، إعتبر هيل (1965) قوتها ، وإتساقها، ونوعيتها، وتزامنها، وترابطها مع أدلة أخرى مقبولة معقولا، وقياسا على تدخلات مماثلة. للتحقق من الإستجابات البيئية لتدخل في إدارة البيئة، يعتبر نوريس وآخرون (2005) المعقولية البيولوجية للإستجابة، والجرعة، والإتساق البيولوجي، وتكوين المجموعات. يتم إجراء هذا التحقيق من قبل لجنة من الخبراء الموثوقون يغطون مجموعة واسعة من التخصصات ذات الصلة ويطلب منهم الحكم على مصداقية الأدلة والتحليل السببي وذلك نظرا

[26] Multiple Lines and Levels Evidence (MLLE)

لطبيعة الأدلة العلمية المتخصصة، متداخلة التخصصات. تمتد تقارير قصة الأداء التشاركية للأدلة متعددة الخطوط والمستويات لتشمل أيضا جمع البيانات الأولية، وإجتماع مع أصحاب المصلحة حيث يتم مراجعة البيانات المتاحة، والبيانات الأولية الإضافية، وتحليل فريق الخبراء لإعداد تقرير تقييم أثر متفق عليه، والذي يتضمن أثر قابل للتدقيق كدليل (على سبيل المثال: كلير هوريزن وآخرون، 2008).

قد يبدأ تطبيق تقارير قصص الأداء التشاركية لحالة البنية التحتية للنقل من خلال تطوير سلسلة النتائج، أو نظرية البرنامج، وتحديد الآثار المقصودة وسلسلة من النتائج المتوسطة وربط البرنامج مع هذه الآثار. ثانيا، سوف يتم تحديد الأدلة المتوفرة حول تنفيذ البرنامج، وتحقق النتائج المتوسطة والآثار النهائية، ومساهمة البرنامج في ذلك، وتقييم تلك الأدلة. وسيجري تقييم الأدلة من حيث الجودة نسبة إلى نوع المعلومات، وليس من حيث التسلسل الهرمي للأدلة القائمة على نوع التصميم البحثي المستخدم. على سبيل المثال، سيتم تقييم المقابلات مع مستخدمي وسائل النقل من حيث أساليب أخذ العينات المستخدمة، والأسئلة المطروحة، والعمليات المستخدمة للحد من التحيز، وإمكانية مراجعة حسابات البيانات. ثالثا، سيتم جمع أدلة إضافية لسد ثغرات معينة في البيانات المتاحة. على سبيل المثال، إذا كانت معظم البيانات المتاحة في شكل إحصاءات رسمية مثل كمية ونوع البضائع التي يتم شحنها، وعدد من الشاحنات التي تستخدم الطرق والنشاط الاقتصادي الإجمالي، فجمع بيانات إضافية قد تشمل جمع صور[27] توضيحية لتأثير برنامج النقل على الفرد والأسر والقرى. كما يهدف جمع البيانات الإضافية إلى إلتقاط المعلومات من مجموعة من وجهات النظر حول الآثار غير المقصودة، سواء كانت إيجابية أو سلبية. على سبيل المثال، تحسين الطرق المحلية يمكن أن يحسن من إمكانية الوصول إلى المرافق الطبية المركزية، تقليل حالات مضاعفات الولادة. يمكن أن تتبع أمثلة الآثار الناجمة عن مجموعة من وجهات النظر عن طريق فحص الإحصاءات الصحية الرسمية.

بعد ذلك، تقوم لجنة من الخبراء في مجموعة المجالات ذات الصلة على مراجعة البيانات من حيث مصداقية إدعاءاتها سواء وصف الآثار وحول نسبها (إما وحدها أو بالاقتران مع عوامل أخرى) إلى البرنامج. في حين

[27] vignettes

أن نهج تقارير قصص الأداء التشاركية لا يربط تحديدا خبراء المراجعة مع جمع بيانات إضافية، سيكون من المرغوب فيه إدراج البيانات الإضافية في هذه المراجعة، وللجنة التوصية بإجراء مزيد من جمع البيانات وتحليلها لمعالجة الثغرات في قاعدة الأدلة. وأخيراً يعقد إجتماع يضم ممثلين عن كافة المجموعات الرئيسية من أصحاب المصلحة لعرض تقرير تقييم الأثر. ويتم نشر التقرير النهائي في إصدارات مختلفة - تقرير موجز بالنتائج الرئيسية وأمثلة للأثر، وتقريرا مفصلا مع ربطه بكل الأدلة التي إستندت إليها الإستنتاجات. لذا، من شأن تقارير قصص الأداء التشاركية أن توفر محاسبة أكثر شمولا لمختلف أصحاب المصلحة، بما في ذلك المجتمع، للتأثيرات المختلفة لبرنامج النقل، مقارنة بإستخدام مناهج أخرى.

3. برنامج مكافحة الفساد

وقد وصفت هذه الحالة على النحو التالي:

هي "دعم المانحين للجنة لمكافحة الفساد في بلد أفريقي. ويتضمن البرنامج المساعدة في وضع مبادئ توجيهية، وتحسين البنية التحتية وجولات دراسية. ويجري تنفيذ برامج مماثلة في ستة بلدان."

إنه ليس تدخل موحد، كما أنه لا يبدو أنه محدد بوضوح مسبقا. بدلا من ذلك من المحتمل أن تظهر الأهداف المحددة للبرنامج، ووسائل تحقيق هذه، أنها حصيلة البرنامج، ويتم تطوير فهم أفضل للأولويات والإمكانيات. لهذا السبب، أود أن أقترح إستخدام التقييم "التنموي" (باتون، 1994، 2008)، والذي لا يهدف إلى تقديم تقرير في نهاية تنفيذ تدخل موحد وثابت، ولكن لتوفير المعلومات أثناء تنفيذ تدخل متغير بإستمرار مع الجوانب المعقدة الهامة.

"يشير التقييم التنموي الى علاقة شراكة طويلة بين المقيمين والعاملين في مجال المبادرات المبتكرة والتنمية. تشمل عمليات التقييم التنموية طرح أسئلة تقييمية وجمع المعلومات لتوفير رد الفعل ودعم إتخاذ القرارات التنموية وتصحيحات المسار على طول الطريق الناشئ. المقيم هو جزء من الفريق الذي يتعاون أعضاؤه لوضع تصور وتصميم وإختبار أساليب جديدة من خلال عملية طويلة المدى ومستمرة للتحسين المستمر، والتكيف والتغيير المتعمد. الوظيفة الأساسية للمقيم في الفريق هو توضيح مناقشات الفريق مع الأسئلة التقييمية والبيانات والمنطق، وتسهيل عمليات التقييم القائم على البيانات للأمورأين هي، وكيف تتكشف الأمور، وأي الاتجاهات حققت الوعد، وما هي الإتجاهات التي يجب أن يتم التخلي عنها، وما هي التجارب الجديدة المطلوب تطبيقها – أو بعبارة أخرى، إتخاذ القرارات المستندة إلى البيانات في كشف وعمليات الابتكار التنموية (ويستلي وآخرون، 2006).

ينطوي تقييم الأثر في هذه الحالة على العمل مع منفذي البرنامج والمشاركين على تطوير نظرية برنامج يمثل فهمهم للمشاكل الرئيسية من حيث الفساد، وكيف يهدفون للتصدي لها. وستطور نظرية البرنامج، وتنقح مع مرور الوقت لكل مبادرة محددة التي يتم تطويرها كجزء من البرنامج، جنبا إلى جنب مع قياسات وطرق مناسبة لاختبار سببية الإسناد. على سبيل المثال، إذا كانت مبادرة تركز على خط ساخن للجمهور للإبلاغ مجهولا عن حالات الفساد، ينبغي أن يكون تقييم الأثر هو متابعة هذه الحالات (والتي يمكن أن تشمل جميع الحالات، أو عينة عشوائية طبقية، اعتمادا على العدد والموارد المتاحة) لمعرفة ماذا حدث من حيث الى اين تم إرشادهم، وما هي التحقيقات التي تمت، وهل كانت الدعوى مسببة. وإذا كان الأمر كذلك، ما هي النتائج المترتبة على الجاني.

يكون من الممكن تحديد نظريات مختلفة للتغيير من خلال تطوير نظريات لكل مبادرة واضحة - على سبيل المثال، هل كانت التدخلات تهدف إلى العمل من خلال تحسين فهم المسؤولين لمتطلبات النزاهة، من خلال تحديد ومعاقبة وإبعاد المسؤولين الفاسدين، أو من خلال ردع الفساد عن طريق زيادة مخاطر الكشف عن

طريق المراجعات الرسمية أو من خلال جعل إتاحة اكبر للجمهور للإبلاغ عن الفساد؟ هل يمكن تطبيق نظريات تغيير معينة لأنواع مختلفة من ممارسات الفساد أو الفساد في أنواع مختلفة من البرامج؟

إذا كان الغرض الرئيسي من التقييم تحديد الممارسات الجيدة وترجمة هذه الى اطارات أخرى، يمكن أن يكون نهج "لانحراف الإيجابي" فعالا. وهذا ينطوي على تحديد المواقع أو الحالات التي يتم فيها تحقيق نتائج غير عادية، والتحقق من ذلك، وتحليل ما يحقق النتائج. إحدى المميزات الكبيرة للـ "الإنحراف الإيجابي" هو أن هذا التحقيق يتم من قبل أولئك الذين يسعون إلى التعلم من الممارسات الجيدة. انه لن يتم من قبل فريق التقييم ثم نشرها قت لاحقا لمستخدمي المعلومات. لقد أستخدم الانحراف الإيجابي في الصحة العامة والتغذية وختان الإناث، والتعليم، والتنمية الزراعية، مع أمثالة في مجال الفساد، وتحديدا طلبات ابتزاز الموظف العام (هورويتز، 2006).

إحدى تحديات تقييم برامج مكافحة الفساد هو تأثير الفساد نفسه على القدرة على جمع أدلة دقيقة، وإتخاذ الإجراءات المناسبة كنتيجة للنتائج. في هذه الحالة، لابد من إعطاء عناية خاصة لإدارة التقييم لضمان إستقلالها وسلامة المقيمين. وقد يشمل ذلك خط مباشر للتقرير إلى مستوى ثقة من الحكومة، لتجنب دفن أو تغيير النتائج، و/أو إشراك منظمات تأييد المواطن في جمع أو التحقق في مزاعم الفساد.

4. الخلاصة

من المثير للإهتمام، أن كل من هذه الحالات تشمل بعض مكونات التقييم القائم على النظرية، على الرغم من أن نظرية البرنامج قد تستخدم بطرق مختلفة جدا. في الحالة الأولى يتم إستخدامه لتحديد النتائج الوسيطة التي يمكن أن تشير إلى تحقيق تأثيرات المدى الطويل، وتحديد العوامل السياقية التي ينبغي التحقيق منها في التحليل. في الحالة الثانية، يتم إستخدامه كإطار مفاهيمي لجلب أدلة متنوعة معا حول مجموعة متنوعة من المكونات.

في الحالة الثالثة، يتم إستخدامه كإطار مفاهيمي لتوجيه تصميم متطور لجمع وتحليل البيانات للإعلام عن التغيير المستمر.

فيما يلي تلخيص لتحليل الحالات، وآثارها على تقييم الأثرفي الجدول (2). وفي الوقت الذي يركز فيه هذا الجدول على الجوانب الرئيسية للتدخلات، فمن المرجح أن الجميع سيكون لديه درجة معينة من الجوانب الأخرى (على سبيل المثال، فإن جميع الحالات من المحتمل أن يكون لديها بعض الجوانب البسيطة).

الجدول (2): تحليل الحالات من حيث الخصائص والمقاصد.

	الجوانب الرئيسية للتدخل	**أغراض التقييم**
حالة 1 – التحويلات النقدية المشروطة	جوانب بسيطة - تدخل موحد جوانب معقدة - يعمل جنبا إلى جنب مع عوامل وبرامج اخرى	تعلم ما اذا كان يعمل – وفي اي سياق يعمل - لإعلام السياسة الجارية
الحالة 2 - البنية التحتية للنقل	معقدة - مكونات متعددة ومتنوعة التي بحاجة إلى العمل معا بشكل فعال	تقييم الآثار الشاملة للبرنامج المنتهي
حالة 3 – مكافحة الفساد	مركبة - تدخل مكيف وناشئ، استجابة لاحتياجات ومشاكل وفرص	فهم وتحسين برنامج مستمر ومتغير

وقد أظهرت تصاميم الموضوعة لهذه الحالات الثلاث أن الإستجابة الواقعية تتطلب معرفة مجموعة واسعة من الأساليب والتقنيات. حتى لو كان التطبيق الفعلي لتقييم الأثر تم التعاقد عليه مع مقيم خارجي ذو خبرات وتجارب ذات الصلة، فإن الجهة المكلفة تحتاج الى فهم كاف للطريقة لتكون قادرة على تطوير الشروط المرجعية، وإختيار الإستشاري المناسب، وإدارة العقد بفعالية، بما في ذلك تقييم جودة العمل. الآثار المترتبة على هذا بناء قدرات التقييم هي أننا في حاجة كل من العمق (المتخصصين في طرق معينة) والإتساع (فهم ووجود مجموعة من الأساليب، ومتى قد تكون الأنسب).

شكر وتقدير

شكراً لمايكل باتون، كاي ستيفنز، هوارد وايت، وبوب ويليامز لتعليقاتهم المفيدة على الإصدار السابق لهذه الورقة.

المراجع:

Boyes, B. (2006). Determining and managing environmental flows for the Shoalhaven River, Report 1 - Environmental Flows Knowledge Review. NSW Department of Natural Resources, May 2006. Retrieved 14 May 2009 from www.dwe.nsw.gov.au/water/pdf/monitor_sholahaven_sh003.pdf

Chambers, R. (2009) Participatory methods. Journal of Development Effectiveness, this issue.

Clear Horizon and O'Connor NRM (2008) Performance Story Report: A study of the Mount Lofty Ranges Southern Emu-Wren and Fleurieu Peninsula Swamps Recovery Program and how it contributed to biodiversity outcomes in the Adelaide and Mount Lofty Natural Resources Management region. Canberra: Commonwealth of Australia. Retrieved 16 May 2009 from http://www.nrm.gov.au/publications/books/pubs/psrmount-lofty.pdf

Dart, J. (2008) 'Report on outcomes and get everyone involved: The Participatory Performance Story Reporting Technique'. Paper presented at the 2008 Australasian Evaluation Society conference, Perth. Retrieved 16 May 2009 from http://www.aes.asn.au/conferences/2008/papers/p100.pdf

Dart, J. (2009) 'Participatory Performance Story Reporting Technique'. Paper presented at the 2009 Impact Evaluation conference, Cairo.

Downes, B. J. L. A. Barmuta, P. G. Fairweather, D. P. Faith, M. J., Keough, P. S. Lake, B. D. Mapstone and G. P. Quinn (2002) Monitoring Ecological Impacts: Concepts and Practice in Flowing Waters Cambridge: Cambridge University Press.

Glouberman, S. (2001) 'Towards a New Perspective on Health Policy', CPRN Study No. H/03, Canadian Policy Research Networks Inc., Ottawa.

Glouberman, S. and B. Zimmerman (2002) Complicated and Complex Systems: What Would Successful Reform of Medicare Look Like? Commission on the Future of Health Care in Canada, Discussion Paper 8. Retrieved 14 May 2009 from http://www.hcsc.gc.ca/english/pdf/romanow/pdfs/8_Glouberman_E.pdf

Guijt, I (2008) 'Seeking surprise : rethinking monitoring for collective learning in rural resource management'. PhD thesis. Wageningen, the Netherlands.

Hill, A.B. (1965) 'The Environment and Disease: Association or Causation', Proceedings of the Royal Society of Medicine 1965 May;58:295-300.

Horowitz, B. (2006) Bridge Ogres, Little Fishes and Positive Deviants: One-on-one deterrence of Public Functionary Extortion Demands. Retrieved 14 May 2009 from http://www.positivedeviance.org/projects/law/Bridges_final.doc

Kurtz C, Snowden D. (2003) The new dynamics of strategy: Sense-making in a complex and complicated world. IBM Systems Journal 2003;42(3):462-483.

Land and Water Australia ( ) Improving the Natural Resource Management System for Regions. Canberra: Australian Government Publishing Service. Retrieved 14 May from http://www.rkrk.net.au/images/3/34/PR061220.pdf

NONIE Subgroup 2 (Network of Networks on Impact Evaluation) NONIE Impact Evaluation Guidance. Retrieved 14 May 2009 from http://www.worldbank.org/ieg/nonie/docs/NONIE_SG2.pdf

NONIE (Network of Networks on Impact Evaluation) Statement on Impact Evaluation. Retrieved 14 May 2009 from http://www.worldbank.org/ieg/nonie/docs/IE_statement.doc

Norris, R.; Liston, P.; Mugodo, J.; Nichols, S. (2005) Multiple Lines and Levels of Evidence for Detecting Ecological Responses to Management Intervention. Paper presented at the American Geophysical Union, Spring Meeting 2005.

Patton, M. Q. (1994) 'Developmental Evaluation', Evaluation Practice, Vol 15, No. 3: 311-319.

Patton, M.Q. (2008a) Utilization Focused Evaluation, 4th ed. Text. Thousand Oaks, CA: Sage Publications.

Patton, M. Q. (2008b). State of the Art in Measuring Development Assistance. Address to the World Bank Independent Evaluation Group. Conference, 10 April 2008, Washington, DC. Retrieved 14 May 2009 from http://www.worldbank.org/ieg/conference/results/patton.pdf

Ravaillon, M. (2009) Evaluating Three Stylized Interventions. Journal of Development Effectiveness, this issue.

Rogers, P.J. (2008a) Four key tasks in impact assessment of complex interventions, Keynote address. Workshop on Rethinking Impact. Understanding the Complexity of Poverty and Change, Consultative Group on International Agricultural Research (CGIAR)

Cali-Colombia. Cali, Colombia. Retrieved 14 May 2009 from http://www.prgaprogram.org/riw/files/papers/Rogers%20material%20for%20workshop.ppt

Rogers, P.J. (2008b) 'Using programme theory for complicated and complex programmes' Evaluation: the international journal of theory, research and practice. 14 (1): 29-48.

Soares, F.V, R. Perez, G.I. Hirata (2009) 'Achievements and Shortfalls of Conditional Cash Transfers: Impact Evaluation of Paraguay's Tekoporã Programme'. Paper presented at the 2009 Impact Evaluation Conference, Cairo.

Stacey, R. (1992). Managing the Unknowable. San Francisco: Jossey-Bass.

Weiss, C. (1997). Evaluation: Methods for Studying Programs and Policies. (2nd ed.) Upper Saddle River, NJ: Prentice Hall.

Westley, F., B. Zimmerman, M.Q. Patton (2006) Getting to Maybe: How the World Is Changed. Random House Canada. Extract retrieved 14 May 2009 from http://innovationlabs.com/r3p_public/rtr3/pre/preread/Patton.Developmental%20Evaluation.pdf

Contacts
International Initiative for Impact Evaluation (3ie)
202-203 Rectangle One
D-4 Saket District Centre
New Delhi - 110017,India
Tel:+91 11 4989 4444
www.3ieimpact.org