

Evaluation Methodology

The Evidence Network Inc.



The
Evidence
Network

Introduction

The Evidence Network Inc. (TEN) was founded in response to the need for a practical approach to evaluating the effectiveness of organizations and programs that support business, research, innovation, and entrepreneurship. We have developed and proven a universally applicable approach to capturing the knowledge-based and economic impacts of innovation enabling organizations on the firms that are their clients.

TEN has provided evaluations of more than 100 organizations and programs including economic development organizations, business incubators and accelerators, research institutes, innovation funding programs, and technology transfer and commercialization organizations. Innovation enabling organizations use our evaluations to support their requests for funding, to communicate their accomplishments to stakeholders, and to inform refinements to strategy and operations. Funders use our evaluations for learning and accountability, and to support evidence-based decision-making in resource allocation.

Our methodology, which has been identified as promising by the US-based Center for American Progress¹, is grounded in an understanding of the research literature and knowledge of user needs based on practical experience.² Our impact evaluations are customizable and can be tailored to suit various innovation investments. They can focus on a single organization, examine changes in impact over time, or benchmark the impact of multiple peer programs.

We differentiate between improvements to firm performance that are a consequence of support services and funding, and improvements to firm performance that would have occurred in the absence of interventions. We identify not only impacts on firm performance, but also the mechanisms through which impact is achieved. Our evaluations capture the full range of technical, social, and economic impacts.

In this document we discuss alternative approaches to presenting the results of investments in business support programs, and present TEN's methodology.

¹ Holly, Krisztina. 2012. *Universities in Innovation Networks – The Role and Future Promise of University Research in U.S. Science and Economic Policymaking*. Center for American Progress, p 27. Available at

http://www.americanprogress.org/wp-content/uploads/issues/2012/01/pdf/dwwsp_university_innovation.pdf

² Dalziel, M., and S. Parjanen, 2012: Measuring the Impact of Innovation Intermediaries: A Case Study of Tekes. In Melkas, H. & Harmaakorpi, V. (eds.) *Practice-based innovation: Insights, applications and policy implications*, Part 1, 117-132, Springer.

Alternative Evaluation Methodologies and Substitutes

Approaches to assessing the results of investments in business support programs range from state of the art evaluation methodologies to substitutes for evaluations such as ‘success stories’. State of the art evaluation methodologies are highly rigorous but are very demanding in terms of data requirements. As a consequence, they are rarely feasible and are used infrequently outside of academia. The most frequently used approaches to reporting on the results of investments are not evaluations, but are often used in place of evaluations. These include success stories, the presentation of firm performance data, the presentation of client satisfaction data, and economic impact analyses, which are attempts to estimate the total impact of interventions on the GDP of a region. In the middle ground between these two extremes are methodologies that are both reliable and feasible. Such approaches include matched sample approaches that seek to identify differences in the performance of treated and untreated firms, and approaches that rely on the judgement of survey respondents to distinguish between differences in performance that are attributable to interventions, and differences in performance that would have happened in the absence of interventions.

Figure 1, below, illustrates the trade-off between rigor and feasibility in reporting the results of investments in business support programs. Highly rigorous approaches include experimental and quasi-experimental designs that are constructed to address the issue of causality. Easily-implemented approaches include biased by design success stories, presentations of data that make no attempt to address causality, and economic impact analyses. Methods that balance rigor and feasibility include matched sample approaches, and the use of expert judgement to assess attribution. In the following we discuss each of these types of approaches.

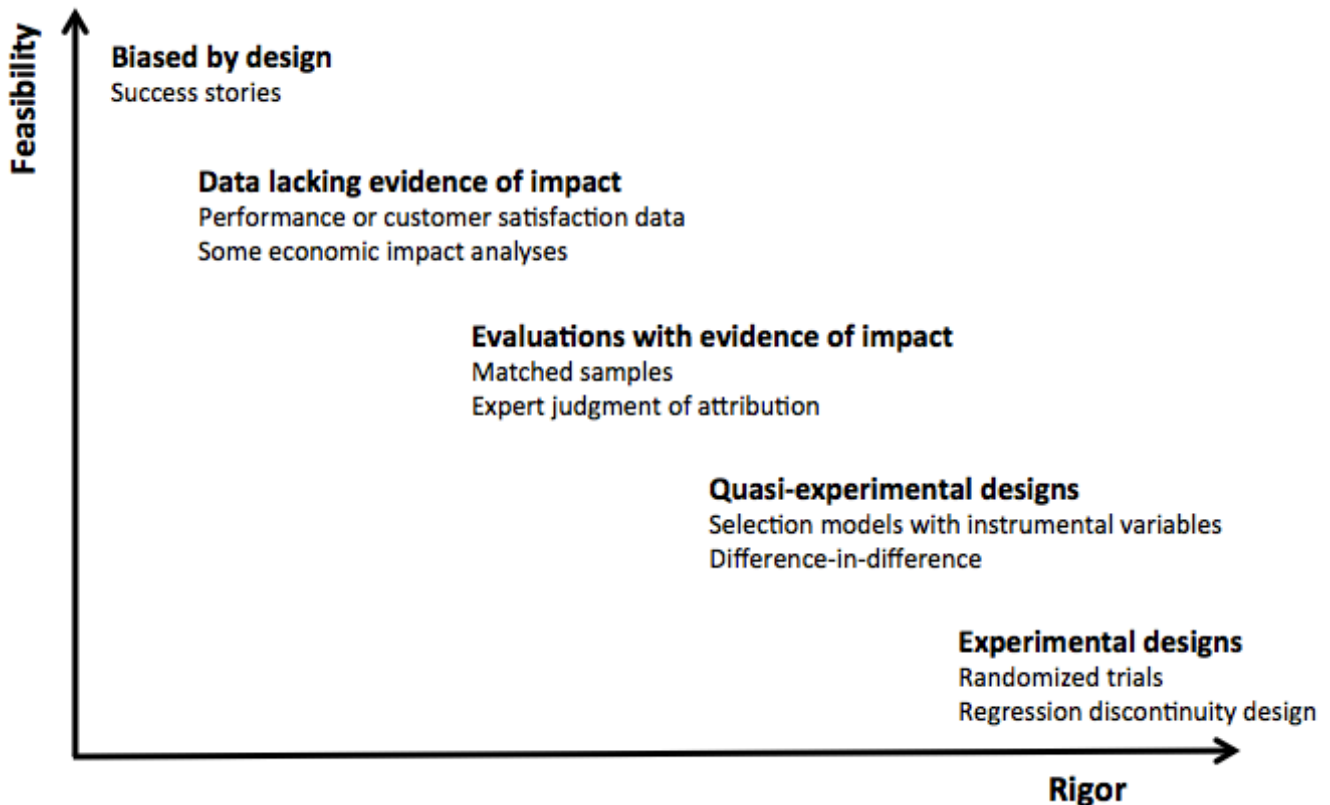


Figure 1: Trade-off Between Rigor and Feasibility

Highly Rigorous Methodologies

Experimental Designs

The critical issue in evaluation is causality, distinguishing between changes that are a consequence of the intervention and those that would have happened in the absence of the intervention. Many regard experimental design approaches, particularly randomized controlled trials (RCTs) as the gold standard in evaluation. RCTs are frequently used to assess the effectiveness of medical interventions and have been used to assess interventions in education, labour training, and international development. A randomized controlled trial is an experiment that randomly allocates treatment to eligible subjects, in our case support services or funding to client firms. By randomly allocating treatment, the possibility for selection bias, which arises when subjects selected for treatment differ systematically from subjects that are not selected for treatment, is eliminated. This makes the determination of the effects of treatment relatively straightforward.

However, RCTs are rarely feasible in the evaluation of business support programs because client companies are unlikely to tolerate the randomized distribution of standardized treatments. While medical interventions are consistently applied across patients, the interventions of business support programs can be highly variable and specific to the recipient firm. Insisting on standardized treatments would seriously compromise the effectiveness of business support. Also, while patients care deeply about the effectiveness of medical interventions, the recipients of business support—particularly financial support—are happy to continue receiving the support regardless of whether or not that support is effective in transforming their behaviour. Finally, while recipients of medical, international development, and educational interventions may be sick, poor, or young and relatively voiceless, the recipients of business support are amongst the most influential members of society. If they are unhappy with the nature of business support and the manner in which it is distributed, politicians will listen. For these reasons almost no business support programs are designed as RCTs. An exception is NESTA's Creative Credits program.³ Regression discontinuity designs (RDDs) lessen the requirement for randomized distribution of treatment, but retain the requirement for standardized treatment.

Quasi-Experimental Designs

RCTs and RDDs are experimental designs and as such require the *a priori* design of the treatment-evaluation experiment. Where the evaluation is being conducted *ex post*, after treatment has occurred, quasi-experimental designs provide a high level of rigor. Like experimental designs, quasi-experimental designs compare the outcomes of supported and unsupported firms. But the allocation of support can be based on merit, as is conventionally the case, rather than randomly assigned. Quasi-experimental designs have their own requirements in terms of longitudinal data and the identification of an instrumental variable that eliminates the endogeneity problem. But where such a variable can be identified, and suitable data is available, quasi-experimental methodologies yield reliable results.

Easily-Implemented Methodologies

Success Stories

Reports on the success of business support programs are abundant, typically comprising popular descriptions of how stakeholders have benefited from particular types of support. Success stories are biased by design as they only report the impact or successes of carefully selected stakeholders.

³ Bakhshi, Hasan, et al. "Creative credits: a randomized controlled industrial policy experiment." (2013).

Performance and Customer Satisfaction Data

Another common approach to impact evaluation is to analyse client performance data and report on changes that have occurred since the intervention (e.g., valuations of accelerator program graduates). While determining the change in performance since an intervention is important, it does not address the issue of causality. For example accelerator program graduates may have obtained high valuations without participating in the accelerator program.

Client satisfaction data is sometimes relied upon as a proxy for outcome data notwithstanding the obvious difference between client satisfaction and client outcomes. For example, while clients may be satisfied with a networking event, the event may or may not have had an impact on their ability to find, for example, new suppliers.

Economic Impact Analyses

Economic impact analyses attempt to estimate the change in economic activity in a specified region as a consequence of an intervention. Historically such analyses were used to estimate the impact (or forecast the impact) of new sports centres, shopping centres, or pipelines. Because the long-term goal of many business support programs is economic development, economic impact analyses have been used to estimate this effect. The limitation of economic impact analyses is reliability. Estimates of direct impacts are grossed up through the use of input-output multipliers. The result is a large number, but a number that may be subject to several unstated assumptions.

Methodologies that Balance Rigor and Feasibility

Matched samples

Matching is a technique that is used to evaluate the effect of an intervention by comparing supported and unsupported firms in an observational study. Propensity-score matching (PSM) is one such approach. PSM uses longitudinal performance data to estimate the effect of an intervention by accounting for the covariates that predict receiving the intervention. It attempts to reduce bias by comparing outcomes from firms that received the intervention to those that did not. The challenge in matched sample techniques is to find suitable control group firms. This can be challenging when it is difficult to identify treated and control group firms in secondary data, difficult to rule out the possibility that treated or control group firms received other types of business support besides the type of support being evaluated, and, if surveys are used to collect data, when it is difficult to persuade control group firms to respond to the survey.

Expert Judgment of Attribution

An alternative to the foregoing is to use expert judgement to identify the effects of a business support program. Attribution of impact can be elicited directly from client firm managers, those individuals best able to determine the effect of an intervention on their firm. Multiple measures of improvements can be considered including impacts on firm resources and capabilities, and impacts on changes in firm performance. This addresses the issue of causality and can also accommodate situations where firms benefit from multiple types of support, making statistical inferences of impact difficult. It also allows impacts to be identified even in cases when firms are pre-revenue or when the lag between intervention and impact on firm performance is pronounced, in which case impacts on standard firm performance measures, such as revenues, are unlikely. In such cases, measures of impact on firm resources and capabilities provide a useful complement or substitute.

The recognition of this novel solution, as an opportunity to meet the unique and high-stakes challenge of measuring the impact of innovation enabling organizations, is what led to the development of The Evidence Network Inc.

TEN's Assessment Methodology

The Evidence Network Inc. has developed and refined a methodology to impact assessment based on the expert judgment of attribution approach. TEN uses survey respondent estimations of specific impacts to reliably measure the attributed impacts of innovation support programs. Attribution of impact is elicited directly from firm managers, researchers, clinicians, and other impact targets – i.e., those individuals best able to determine the effect of an intervention on their performance. Multiple measures of improvements can be considered including short-term impacts on resources and capabilities, and medium-term impacts on changes in performance. This addresses the issue of causality and can also accommodate situations where stakeholders benefit from multiple types of support. It also allows impacts to be identified even in cases when the lag between intervention and impact on stakeholder performance is pronounced, in which case impacts on standard performance measures, such as revenues or employment, are unlikely. In such cases, measures of impact on stakeholder resources and capabilities, embedded in our methodology, are the most appropriate measures. TEN's methodology in terms of impact on stakeholder performance has been validated using Propensity Score matching, a Quasi-Experimental approach.

TEN's approach to measuring innovation impact is based on the premise that innovation-enabling organizations can be described as an overarching class of organizations whose members share common goals. Despite their diversity, innovation enabling organizations, ranging from small economic development organizations to large and sophisticated research institutes, seek to make their members or clients more innovative, in the interests of facilitating increases in their viability, profitability, or other measures of their success.

The logic model shown on the following page illustrates how innovation-enabling organizations work to fulfill their missions, and how TEN measures their impact.⁴ As shown at the top-left of the diagram, innovation-enabling organizations express their purpose in terms of national competitiveness, regional economic development, industry strength, or viable new ventures, and conduct activities to achieve direct impact on resources and capabilities, indirect impact on performance, and long-term impacts in the form of socio-economic benefits. The direct impacts of these organizations are improvements to the resources or capabilities of clients, indirect impacts are improvements to market performance of clients, and long-term impacts affect communities, industries, economies, and societies.

⁴ TEN's innovation enabling logic model shown uses companies as the impact target. It has been adapted and successfully used to assess the impact of healthcare providers, researchers and clinicians, non-profit organizations, and other impact targets.



Figure 2: Innovation Enabling Organization Logic Model

Understanding the Mechanisms of Change

Figure 3 below depicts a logic model in a U-shaped format to emphasize the mechanisms of change. The leading indicators of impact are improvements to resources and capabilities. Impact on resources and capabilities occurs in the short term (i.e., typically less than one year), and is the result of the support services or funding. Impact on performance occurs in the longer term (i.e., typically within one to three years), and is the result of the improvements to resources and capabilities.

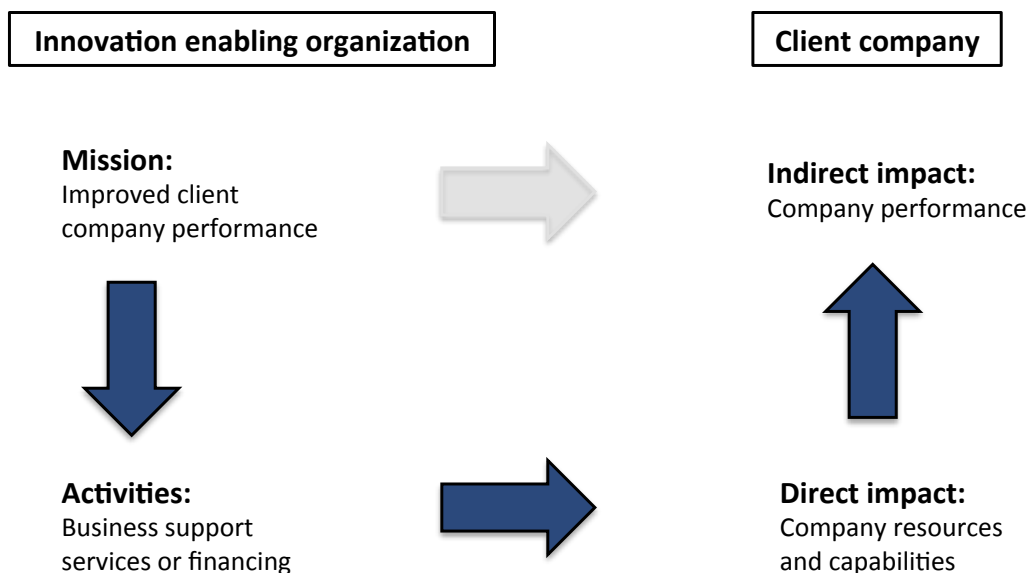


Figure 3: Leading Indicators of Impact

TEN measures the intensity of use of services, and direct and indirect impact on performance using a customized survey instrument. TEN determines impact on resources and capabilities, and performance, by identifying the degree to which improvements are attributable to the support services offered by business support programs. Through statistical examinations of the relationships between the intensity of use of services, direct impact on resources and capabilities, and impacts on performance, we can determine which services and direct impacts are significantly associated with improvements to clients' performance in the market. This enables us to report both the degree to which an intervention has impacted stakeholders, and **how** the intervention has achieved that impact.

Benchmarking

TEN conducts its impact evaluations using a systematic and standardized approach, and all of our assessments are underpinned by our theory of change, as depicted in our logic model. As a consequence, we have the ability to benchmark peer organizations against one another on the following indicators: impact on strategy, operations, access to financing, and linkages, revenues, employment, investment capital raised, and time to market.

Evaluation Process

TEN's methodology and approach ensures that the appropriate measures of impact, developed in the client specific logic model created at the outset of each of our evaluation projects, feed directly into a customized, condensed impact assessment survey.

Impact Surveys

The impact survey is used to elicit information from stakeholders on the importance of innovation support initiatives, and the subsequent impact of support on improvements to stakeholders' resources and capabilities and performance on a number of specific measures.

TEN's impact surveys elicit the following information directly from stakeholders, enabling us to segment impact according to respondent attributes to determine if there are significant differences between sub-samples:

- Stakeholder demographic data (investments in R&D, size, etc.)
- The intensity of use of services (participation in research projects, events, access to facilities, etc.)
- The impact stakeholders attribute to the support program on improvements to their resources and capabilities (improvements to knowledge, new industry or research linkages, improved access to financing, etc.)
- The impact stakeholders attribute to the support program on improvements to their performance (new products, processes, or services developed, increased revenues, patent applications, etc.)

Impact Assessment Report and Infographic

Our final deliverable is a succinct Impact Assessment Report and accompanying Companion Reports that include findings, interpretations of results, and recommendations to facilitate continuous improvement. With information on the relationships between services, impact, and the attributes of companies we are able make specific policy recommendations to help build on the strengths of support programs, while also addressing potential gaps. We also produce infographics that visually communicate impact to a variety of audiences.

Conclusion

TEN's assessment methodology is both reliable and nearly universally applicable. It can be used to assess the impact of all direct business support programs, ranging from those that support pre-revenue ventures to those that support the commercialization of scientific discoveries. It is not suitable for the assessment of indirect support such as tax credits.