

## **PRÁTICAS AVALIATIVAS EM POLÍTICAS DE CT&I: ANÁLISE COMPARATIVA DO CASO NACIONAL**

**Rodrigo Nunes Peclat  
André Tortato Rauen**

### **RESUMO**

Avaliar políticas públicas quanto aos seus impactos é essencial para a compreensão da sua efetividade. A fim de melhor entender como o perfil das avaliações brasileiras na esfera federal dos programas de CT&I se comparam com a comunidade internacional, é realizada uma comparação com o repositório SIPER – o qual reúne um conjunto de avaliações de políticas públicas de CT&I de vários países do mundo, incluindo aqueles na fronteira tecnológica. Assim, a partir da extração dos dados desse repositório e da revisão sistemática de um conjunto de avaliações nacionais produzidas por instituições significativas realizou-se a comparação entre estatísticas selecionadas como momento de avaliação, impactos avaliados, desenhos avaliativos empregados. Entre as principais constatações, observou-se que uma grande diferença entre os perfis no que é referente à consideração da teoria do programa como subsídio às avaliações. Além disso, verificou-se que o retorno do investimento é uma prática bem menos utilizada no Brasil do que no âmbito do SIPER. Por fim, riscos são identificados pelos dados que sugerem falta de integração entre métodos qualitativos e quantitativos nos estudos analisados.

### **PALAVRAS-CHAVE**

Impacto; Avaliação; SIPER; Ciência; Tecnologia; Inovação.

### ***EVALUATION PRACTICES IN ST&I POLICIES: A STUDY OF THE NATIONAL CASE SUMMARY***

*Impact evaluation is essential to understand the real effects of a public policy. This work intends to compare Brazilian federal ST&I evaluations practices to international community. This proposal utilizes SIPER – one of the main ST&I policies repositories in the World – to carry out this study. SIPER has evaluations from several countries around the World, including those in technological frontier. From SIPER's data and systematic review from Brazilian evaluations prepared by relevant institutions, it has been compared these two sets utilizing statistics from evaluation timing, impacts covered, evaluation design. Among the main findings, it has been observed that program theory in Brazilian evaluations have almost never been used. Besides that, return over investment is a practice much more frequent in SIPER than in Brazilian sample. Lastly, risks have been identified when the data suggested the absence of integration between quantitative and qualitative methods in Brazilian evaluations.*

### **KEYWORDS**

*Impact; Evaluation; SIPER; Science; Technology; Innovation.*

## 1. INTRODUÇÃO

É conhecido que as políticas de ciência, tecnologia e inovação (CT&I) no Brasil sofrem com a falta de constância na avaliação de seus impactos. Contudo, em razão de esforços individuais de especialistas e acadêmicos já existe um conjunto interessante, porém insuficiente de estudos no país. O que não se sabe é se essas práticas avaliativas nacionais estão em sintonia com o que faz países líderes tecnológicos.

As avaliações de impacto têm sido cada vez mais exigidas pela sociedade, o que tem se convertido em determinações de órgãos de controle, em especial na esfera federal. A insuficiência avaliativa dos programas de CT&I quanto à efetividade em relação às medidas propostas para o Plano Plurianual e em relação à superação dos desafios estratégicos da Estratégia Nacional de Ciência, Tecnologia e Inovação já foi objeto de análise e de recomendações para os ministérios gestores envolvidos (BRASIL, 2016).

Contudo, do ponto de vista internacional inclusive, há registro de que mesmo na presença de avaliações, sua utilidade no âmbito do processo decisório é limitada (OECD, 2009). Dessa forma, cresce a importância de se compreender como práticas avaliativas nacionais se comportam frente ao que é internacionalmente aceito para, então, corrigir rotas e adequar esse importante instrumento de *accountability* à natureza das atividades de CT&I.

Por isso, esse texto objetiva, a partir de critérios e indicadores já estabelecidos pela literatura internacional, comparar as características das avaliações de impacto feitas no Brasil com aquelas realizadas por outros países. Para isso, se empregará como referencial o repositório SIPER<sup>1</sup> - uma base mantida pela Universidade de Manchester, a qual, no momento deste estudo, apresentava cerca de 600 avaliações de políticas e programas de CT&I em vários países do mundo, inclusive brasileiras.

---

<sup>1</sup> Science and Innovation Policy Evaluation Repository (SIPER). Consulte: <http://si-per.eu/>. Último acesso: nov. 2018.

Para atingir seu objetivo, além dessa introdução, o texto encontra-se dividido em cinco seções. Inicialmente na Seção 2 apresenta-se uma discussão sobre as características avaliativas utilizadas ao longo deste estudo. Na Seção 3, a metodologia utilizada é apresentada. Na Seção 4, os principais resultados são discutidos, juntamente com análises críticas sobre a comparação com o SIPER. Por fim, uma seção de conclusão encerra este trabalho.

## 2. AVALIAÇÕES DE IMPACTO EM POLÍTICAS DE CT&I

De forma geral, avaliar o impacto de políticas públicas relaciona-se a estabelecimento de uma forma metodologicamente robusta de testar causalidade entre variáveis (GERTLER, 2016). Diferentemente de outras modalidades avaliativas, o foco deste tipo de avaliação é a de responder se a política, de fato, está sendo responsável ou não pela modificação no comportamento de dimensões e indicadores de interesse.

Especificamente quanto a avaliação de impactos em CT&I, primeiro identifica-se a necessidade de compreender seu relacionamento com os demais estágios presentes num ciclo de política pública. A partir daí busca-se diferenciar entre as entregas imediatas da política (seus produtos e resultados) dos impactos (efeitos propriamente ditos), bem como se identifica a necessidade de realizar uma discussão acerca das dificuldades relacionadas a essas atividades e que se tornam ainda mais acentuadas quanto a avaliações de CT&I.

O ciclo da política pública pode ser caracterizado em diferentes estágios que estabelecem diferentes tipos de relacionamentos entre si, a saber: definição de agenda, formulação, tomada de decisão, implementação e avaliação (WEGRICH, 2006). Os resultados decorrentes dos estágios de avaliação podem ter diferentes efeitos sobre cada uma das etapas desse ciclo, uma vez que ela pode ser utilizada para o estabelecimento de prioridades e a seleção de políticas, pode ser utilizada para o monitoramento do progresso das atividades da política, ou numa abordagem *ex-post* ser utilizada para avaliação

quanto ao alcance de resultados e impactos de médio e longo prazo (LINK; VONORTAS, 2013).

Resultados podem ser caracterizados como as entregas mais imediatas que decorrem no âmbito de uma política ou programa, diretamente decorrentes das ações realizadas pelos seus atores a partir dos recursos alocados. Já os impactos podem ser conceituados como os efeitos percebidos dessas entregas ou desses produtos sobre o ambiente, sobre a sociedade a médio e a longo prazo (GODIN; DORÉ, 2004).

Observa-se que além das dificuldades inerentes às avaliações de impacto de cunho geral, certas características tornam-se ainda mais desafiadoras quando se trata de políticas e programas de CT&I, dada a presença de forte incerteza tecnológica. O Quadro 1 apresenta uma visão geral dessas dificuldades.

<b>Desafio</b>	<b>Descrição</b>
Atribuição e Múltipla Causalidade	Relacionada à dificuldade de medir a porção exata do efeito que é decorrente da política implementada, descartando-se os demais efeitos influenciadores.
Horizonte de Medição	Relacionado ao fato de que certos efeitos tem um horizonte temporal bastante longo para se manifestarem, trazendo a questão sobre qual é o melhor momento de avaliar a política de C&I
Apropriabilidade	Relacionada à capacidade de medir os efeitos junto às partes interessadas ou às dimensões apropriadas, como a dimensão econômica, social ou cultural.
Não-Linearidade	Relacionada à dificuldade de expressão de modelos de impactos de políticas de inovação tendendo à simplificação para uma forma linear e assim perdendo atores e relacionamentos causais-chaves para o fenômeno a ser avaliado.
Endogeneidade	Relacionada à dificuldade de se estabelecer relacionamentos lineares entre variáveis explicativas e variáveis independentes devido à existência de relacionamentos circulares entre elas.

*Quadro 1 - Desafios sobre a avaliação de impactos de políticas de CT&I*

*Fonte – Elaboração dos autores a partir de (OECD WORKSHOP ON ASSESSING THE IMPACTS OF PUBLIC RESEARCH SYSTEMS, 2016b)*

## 2.1. POLÍTICAS DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO

A longo prazo, a inovação é um fator determinante para o crescimento econômico de um país. Reconhece-se que a capacidade de produção de bens e serviços de um Estado influencia o padrão de vida de sua população e em torno disso promove-se a mudança técnica necessária para condições melhores (MANKIW, 2014). A fim de promover a essa mudança, entre outros gastos inovativos, destacam-se aqueles direcionados para pesquisa e desenvolvimento (P&D), os quais, sob uma perspectiva evolucionária, constituem-se como práticas de busca permitindo a combinação de conhecimentos existentes em novidades técnico-sociais (CORAZZA; FRACALANZA, 2009).

Contudo, a produção de conhecimento tem características intrínsecas como incerteza, inapropriabilidade e indivisibilidade que tendem a influenciar, de forma adversa, os retornos privados das firmas pertencentes a um Estado, fato este que acaba por provocar subinvestimento privado em atividades de P&D (JAFFE, 1997). A fim de sanar essa falha de mercado, políticas públicas têm sido desenvolvidas em vários países para estimular o investimento – seja público ou privado – em pesquisa e desenvolvimento. Com o desenvolvimento do modelo de sistema nacional de inovação (SNI), o qual concebe esse fenômeno como decorrência da interação de um conjunto de organizações, e não apenas como o resultado do custeio de atividades de P&D, tem-se outras falhas como as de capacidade tecnológica, relacionadas à dificuldade de absorção de potencial inovativo, as institucionais, quando organizações-chaves ao SNI apresentam funcionamento inadequado (como escritórios de patentes), as de rede (relacionadas a dificuldades na interação entre as organizações do SNI), e as de *framework* (assim denominadas as condições subjacentes ao SNI como questões econômicas, regulatórias, culturais e sociais necessárias ao seu bom funcionamento) (ARNOLD, 2004). Para cada um dessas condições, políticas públicas focadas em inovação tratando cada uma dessas condições tornam-se necessárias.

Dessa forma, no âmbito de um sistema nacional de inovação, encontra-se um conjunto heterogêneo de políticas públicas, voltadas para diferentes tipos de falhas que surgem no âmbito desse sistema, motivando a ação do Estado. Contudo, devido ao longo prazo dos efeitos das políticas de inovação, as questões da dificuldade de atribuição dos efeitos e da múltipla causalidade tornam-se ainda mais relevante para essas políticas devido à

complementariedade ou à substituição que sofrem junto a outras medidas ou programas, exigindo que a avaliação dos seus impactos considere de forma criteriosa esses obstáculos de medição (OECD WORKSHOP ON ASSESSING THE IMPACTS OF PUBLIC RESEARCH SYSTEMS, 2016b).

Além disso, há de se ressaltar que, diferentemente de outros domínios do conhecimento, o estabelecimento de modelos de impacto para a avaliação de políticas de CT&I é mais custoso devido à sua natureza essencialmente não linear (OECD WORKSHOP ON ASSESSING THE IMPACTS OF PUBLIC RESEARCH SYSTEMS, 2016d), com inúmeros relacionamentos causais intermediários refletindo a auto-organização dos agentes em torno do fenômeno inovativo ao qual a política pública está atrelada (CORAZZA; FRACALANZA, 2009). Modelos econométricos, em particular, são vulneráveis a questões de endogeneidade, em que os relacionamentos dinâmicos entre variáveis – explicativas e independentes -, próprios em políticas de inovação, trazem às especificações estáticas da Econometria maiores desafios à mensuração dos efeitos de longo prazo de CT&I (OECD WORKSHOP ON ASSESSING THE IMPACTS OF PUBLIC RESEARCH SYSTEMS, 2016b).

Conforme o olhar, diferentes características avaliativas podem ser definidas. Considerando as especificidades das intervenções de CT&I é possível definir um conjunto que reúne as mais relevantes (Quadro 2). Sobre ele, cabe ressaltar que de acordo com OECD (2009) tem-se que as características são bastante dependentes do contexto em que são empregadas. Essa afirmação é confirmada por EDLER (2012) ao analisar um conjunto de avaliações realizadas no âmbito da União Europeia entre 2002 e 2007. Para cada uma dessas dimensões, é possível identificar uma diferente agregação de valor à discussão sobre perfis avaliativos.



<b>Atributos</b>	<b>Descrição</b>
Posicionamento	Relaciona-se ao posicionamento interno, externo (porém interno ao Governo), ou completamente externo à política avaliada demonstrada pela equipe de avaliação.
Momento da Avaliação	Relaciona-se ao fato de a avaliação ser <i>ex-ante</i> , <i>interim</i> (após o fechamento de uma fase), <i>ex-post</i>
Teoria do Programa	Relaciona-se com a modelagem lógica do programa – entradas, ações, produtos, resultados e impactos – orientando as questões avaliativas
Tipos de Impacto	Relacionam-se aos tipos de impacto que uma política de CT&I pode gerar. Como exemplos tem-se impactos econômicos, sociais, tecnológicos, de capacitação, ambientais e externalidades.
Retorno do Investimento	Relaciona-se à quantificação dos custos e dos benefícios da política avaliada permitindo uma comparação da relação custo x benefício
<i>Design</i>	Relaciona-se à aplicação de métodos experimentais ou quase-experimentais (baseados em contrafactuais – isto é, na estimação da realidade alternativa na inexistência do programa) ou na utilização de métodos não experimentais (ex: estudos de caso) para avaliação.
Coleta de Dados	Relaciona-se ao emprego de técnicas como entrevistas, <i>surveys</i> ou análises de bases de dados a fim de buscar informações de qualidade para a avaliação da política
Análise de Dados	Relaciona-se a um conjunto de técnicas que pode ser empregada para análise dos dados coletados. Como exemplos no SIPER tem-se análise econométrica, estudo de caso, análise de redes, análise custo-benefício, dentro outras.

Quadro 2 - Características avaliativas selecionadas

Fonte - SIPER

### 2.1.1. POSICIONAMENTO DO AVALIADOR

A área de CT&I não está isenta de questões relacionadas a conflito de interesses nas suas várias atividades avaliativas (MARCOVITCH, 2010). Ao mesmo tempo, há de se considerar os vários propósitos implícitos que podem estar motivando a realização de uma avaliação, desde o seu uso como um subterfúgio político para a postergação de uma tomada de decisão, até mesmo para a tomada de decisão baseada em evidências a fim de se decidir sobre ampliar, reduzir ou manter um programa (WEISS, 1998). Nesse contexto, a escolha sobre uma equipe de avaliação interna ou externa à política, traz reflexos sobre questões como: relacionamento de confiança junto aos gestores da

política, objetividade na análise das informações, capacidade de entendimento do programa, potencial da avaliação em alterar a realidade, autonomia da equipe avaliativa (WEISS, 1998).

Cabe ressaltar que dependendo do envolvimento e da relação estabelecida junto à equipe do programa (ex: um relacionamento financeiro de alta materialidade), estabelece-se um ambiente mais propício para que mesmo uma equipe externa produza avaliações com um menor nível de independência (WEISS, 1998).

### 2.1.2. MOMENTO DA AVALIAÇÃO

De acordo com o SIPER, as avaliações podem ser prévias (antes da implementação do programa), de acompanhamento (durante sua implementação), interim (após uma determinada fase) ou *ex-post* (após o término da política). Cada um desses tipos apresenta diferentes tipos de dificuldades para serem realizados, destacando-se as modalidades *ex-ante* e *ex-post*.

Avaliar impactos de uma política de forma *ex-ante* exige um exercício de antecipação sobre o futuro, o que se tratando de políticas de CT&I traz como obstáculos adicionais o longo prazo que elas podem alcançar e a diversidade de atores com opiniões divergentes sobre esses resultados futuros presentes em um sistema de inovação (MILES; SARITAS; SOKOLOV, 2016). Essa modalidade não fica restrita apenas ao desenvolvimento de novas políticas, mas também se apresenta comum ao a discussões sobre priorização de recursos, em que abordagens multicritérios são frequentes para situações mais complexas, as quais muitas vezes impossibilitam monetização dos impactos trazidos pela política em análise (MILES; SARITAS; SOKOLOV, 2016).

Avaliações *ex-post*, por outro lado, também são dificultadas não apenas pelo longo prazo dessas políticas, mas pela sobreposição de diferentes efeitos que ao longo do tempo influenciam os impactos que estão sendo avaliados (OECD WORKSHOP ON ASSESSING THE IMPACTS OF PUBLIC RESEARCH SYSTEMS, 2016). A atribuição da causalidade torna-se ainda mais desafiadora em políticas de CT&I devido ao caráter



evolucionário dos comportamentos empregados pelos agentes para solução de seus problemas (CORAZZA; FRACALANZA, 2009).

### 2.1.3. TEORIA DO PROGRAMA

Teoria do Programa pode ser conceituada como a variedade de maneiras de se construir uma cadeia de causalidade a partir de elementos de entrada que são processados por uma série de ações, visando objetivos estabelecidos. O detalhamento da Teoria de um programa, geralmente por meio de modelos diagramáticos, subsidia a avaliação orientada à teoria (ROGERS, 2000).

Uma ferramenta para implementar essa abordagem avaliativa é a construção de modelos lógicos – estruturação de dados relativos a dimensões como entradas, ações, produtos, efeitos de médio e de longo prazo - a fim de tornar mais transparente às partes interessadas a análise da teoria do programa em questão (LINK; VONORTAS, 2013). Contudo, existem aspectos de uma intervenção que podem levar a diferentes níveis de dificuldade no desenvolvimento de um modelo lógico. São exemplos dessas complexidades políticas com múltiplos atores, cadeias de causalidade paralelas e modelos de impacto não-lineares (ROGERS, 2008).

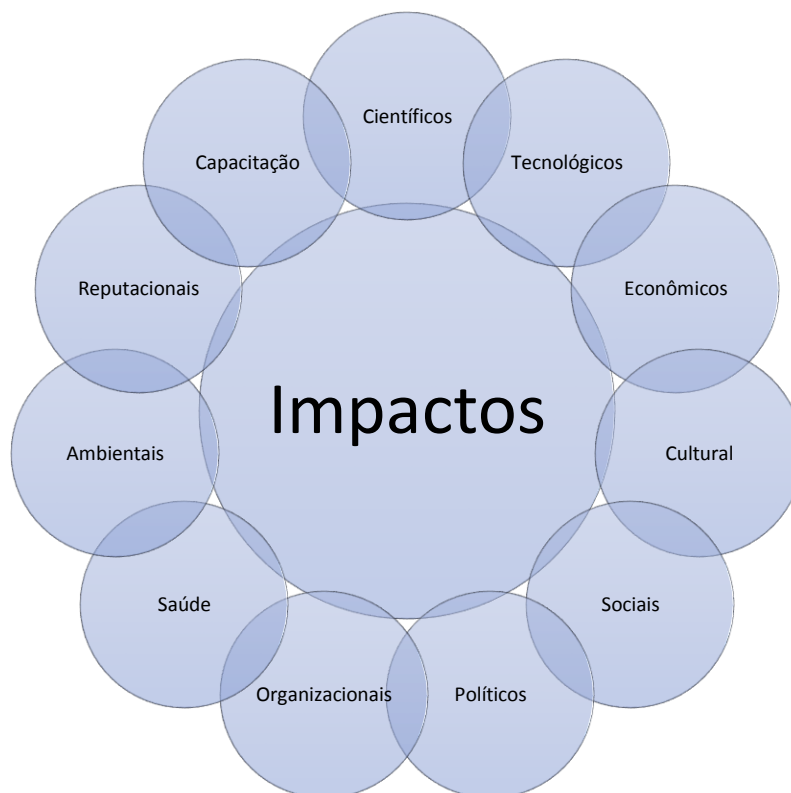
Outro problema relacionado à utilização dos modelos lógicos em avaliações é a tendência do deslocamento da responsabilidade de desenvolvimento deles da equipe gestora da política, a qual tem o dever de comunicar com precisão às partes interessadas os mecanismos de funcionamento do programa, para a equipe de avaliação, de uma forma *ex-post* ao desenho e à utilização dos recursos da política, trazendo uma responsabilidade de reconstrução do programa em uma tentativa de reproduzir o que ele realmente foi (BACH, 2012).

### 2.1.4. TIPOS DE IMPACTO

A literatura que discute avaliações de programas e políticas de CT&I indica que há uma ênfase sobre os impactos econômicos dessas iniciativas, contudo ressalta-se que há

vários outros tipos de efeitos a serem mensurados (GODIN; DORÉ, 2004). A Figura 1 apresenta uma visão sobre essa tipologia de efeitos.

Essa diversidade de efeitos traz dificuldades avaliativas uma vez que diferentes tipos de impactos acabam exigindo maior esforço quanto à compreensão de mecanismos de funcionamento da política, tornando sua teoria de programa mais complexa (ROGERS, 2008). Uma vez que as evidências indicam que os avaliadores definem suas estratégias de acordo com os impactos a serem avaliados (EDLER, 2012), conclui-se que a investigação abrangente dos efeitos das políticas de CT&I tende a ser custosa, tanto em termos de recursos financeiros, quanto em recursos não financeiros. Devido à maturidade de certas áreas – como a econômica – certos tipos de impactos tendem a oferecerem uma maior facilidade técnica para estimação (GODIN; DORÉ, 2004).



*Figura 1 - Tipologia de impactos em Ciência*

*Fonte – GODIN e DORÉ (2004)*

### 2.1.5. RETORNO DO INVESTIMENTO

A análise de custo benefício consiste na quantificação dos custos e dos benefícios produzidos por uma política, incluindo aqueles que o mercado não costuma produzir mensurações de valor econômico (TREASURY, 2003). Muitas vezes empregado como uma forma de avaliação *ex-ante*, pode ser também empregado de forma *ex-post* permitindo uma compreensão absoluta e relativa do valor dos programas de CT&I (OECD WORKSHOP ON ASSESSING THE IMPACTS OF PUBLIC RESEARCH SYSTEMS, 2016a).

No domínio de CT&I, essa técnica apresenta maior número de associação a projetos e programas de pesquisa aplicada e desenvolvimento de tecnologias, enquanto que a sua aplicação para avaliação de programas voltados para pesquisa básica é muito restrita (RUEGG, 2007), devido à dificuldade de monetizar esses achados, o que evidencia a limitação da técnica, não podendo ser entendida como um critério definitivo para decisões sobre investimento. De fato, em CT&I, mais importante que uma visão individual sobre projetos ou programas isolados, é a maximização do retorno do portfólio como um todo, o que pode envolver agrupar algumas iniciativas sem relacionamento direto com benefícios econômicos (LINK; VONORTAS, 2013).

### 2.1.6. *DESIGNS* AVALIATIVOS

Experimentos ou aleatorizações são considerados os “padrões-ouro” em avaliação, uma vez que, ao gerar de forma aleatória grupos de tratamento e de controle, eliminam o problema de viés de autosseleção, produzindo dois conjuntos que, na média, tenham características que tendam a ser estatisticamente equivalentes. Entretanto, o acesso a uma política por somente por questões de sorte, pode levar a debates éticos que inviabilizam a aplicação do desenho (GERTLER et.al 2016).

Quasi-Experimentos também buscam a formação de grupos de tratamento e grupos de controle, contudo não mais a partir de aleatorizações. Embora possam muitas vezes serem mais compatíveis a situações do dia a dia, trazem consigo suposições que muitas vezes trazem críticas e dúvidas sobre seus resultados (GERTLER et.al 2016).

Como *designs* não experimentais, excluindo os quasi-experimentais, pode-se citar os estudos de caso, os quais consistem em estratégias de pesquisa voltadas para o estudo de ciências sociais, quando se busca, principalmente, explicações ou descrições sobre fenômenos, em situações em que não há controle sobre os eventos comportamentais observados (YIN, 2015).

### 2.1.7. COLETA DE DADOS

Uma forma inicial de coleta a ser discutida é o uso de entrevistas. Presenciais ou não presenciais, estruturadas ou não estruturadas, baseiam-se na interação direta entre o entrevistador e o entrevistado. Um desafio dessa modalidade é o não enviesamento das respostas, uma vez que se verifica uma tendência de os entrevistados apresentarem opiniões que acham ser as esperadas pelos avaliadores (“efeito *Hawtorne*”) (WEISS, 1998).

A aplicação de *surveys* tem mostrado vantagens quando a realização de entrevistas tem se mostrado inviável, entretanto traz consigo, como uma das suas principais desvantagens, o fato de que muitos dos potenciais respondentes não retornam as respostas do questionário encaminhado, sendo a baixa representatividade uma ameaça à avaliação. (WEISS, 1998).

Quanto às bases de dados, um desafio relacionado às avaliações de políticas públicas, além de sua disponibilidade, é a sua acurácia e atualidade (WEISS, 1998). No caso do Governo Federal brasileiro, há problemas relacionados a interoperabilidade de dados entre as diferentes organizações<sup>2</sup>.

### 2.1.8. ANÁLISE DE DADOS

Cada método de análise traz consigo vantagens e desvantagens de acordo com o objeto a ser avaliado. Considerando que alguns dos empregados no SIPER já foram discutidos

---

<sup>2</sup> Para maiores informações, consulte: <http://eping.governoeletronico.gov.br/#p1s3.4>. Último acesso: nov. 2018.

no escopo deste texto, pode-se tratar ainda de análises de redes, análises econométricas, análises estatísticas, análises de texto e *roadmaps* tecnológicos, os quais são utilizados para classificação dos artigos identificados.

A análise de redes busca a partir do estudo da evolução temporal dos relacionamentos entre os agentes presentes num sistema de inovação prover conclusões a respeito da adicionalidade comportamental decorrente de uma determinada política (LINK; VONORTAS, 2013). Já a análise econométrica busca prover um arcabouço matemático a fim de testar a validade de hipóteses econômicas sobre o mundo real. Em CT&I, tendem a ser microeconômicas, focando nas diferenças ocorridas entre grupos de tratamento e de controle (OECD WORKSHOP ON ASSESSING THE IMPACTS OF PUBLIC RESEARCH SYSTEMS, 2016a).

O emprego de estatística descritiva, sumarizando informações, analisando tendências nos dados, medidas de dispersão e formatos de distribuição é uma técnica comum de análise (FREY, 2018). Análises de texto, como *text mining*, tornam-se frequentemente encontradas ao longo do ciclo de vida de uma política pública, incluindo sua fase de avaliação (NGAI; LEE, 2016). Já os *roadmaps* tecnológicos tem se apresentado como ferramentas especialmente voltadas para análise prospectiva, com o objetivo de esboçar o futuro de um domínio tecnológico, prevendo como seria com o passar do tempo o desenvolvimento de várias tecnologias correlatas e a interação de atores chaves (GEORGHIOU, 2008).

### 3. METODOLOGIA

A necessidade de comparabilidade de boas práticas entre diferentes comunidades é uma constante presente em vários domínios científicos. No caso das avaliações das políticas de CT&I, com suas particularidades, essa necessidade não foi diferente. Uma iniciativa apresentada em 2012, pela Universidade de Manchester, representou um esforço inédito de meta-análise de 171 avaliações de políticas de inovação realizadas entre 2002 e 2007 no âmbito da União Europeia, permitindo uma comparação entre perfis avaliativos no

âmbito da comunidade de CT&I sob diferentes perspectivas – que variavam da medida mais simples (programa) à medida mais agregada (país ou bloco) (EDLER, 2012).

Desse trabalho, também no âmbito da Universidade de Manchester, originou-se o *Science and Innovation Policy Evaluation Repository* (SIPER), o qual propõe-se a ser um repositório de avaliações de políticas, mas que diferentemente do trabalho de 2012, tem o objetivo de concentrar estudos do mundo inteiro, ainda que o foco inicial sejam os países membros da OECD. Liderado pelo grupo da referenciada Universidade inglesa, apresenta ainda o suporte de equipes brasileiras e francesas, respectivamente da UNICAMP e da IFRIS.

Como apresentado pela equipe do SIPER, o número de avaliações registradas no âmbito desse repositório não reflete o número real de estudos produzidos em cada país, mas sim reflete o resultado do processo de busca realizado pela equipe envolvida no projeto, o qual ainda está em desenvolvimento. Em seu atual estágio, ao final do ano de 2018, há cerca de 540 estudos catalogados, sendo que desses apenas quatro são brasileiros, o que dificulta a obtenção de qualquer diagnóstico sobre as avaliações nacionais. Além disso, há de se considerar que desses trabalhos presentes no SIPER, há uma sobreposição alta das avaliações que estudaram os efeitos (475) e das avaliações que estudaram as falhas de implementação (442). Dessa forma, por uma questão de delimitação de escopo, os estudos nacionais analisados neste trabalho são do primeiro grupo (somativos). Outra restrição estabelecida é que o levantamento fosse realizado somente junto a políticas e programas federais.

Além disso, é importante considerar que o SIPER, por si só, já estabelece requisitos para que uma avaliação seja incluída em seu âmbito. São eles:

1. relacionar-se diretamente com políticas de CT&I
2. apresentar uma sistemática determinação do mérito da avaliação;
3. avaliar um instrumento ou um conjunto de instrumentos;
4. apresentar uma metodologia observável;



5. apresentar evidências do desempenho da política ou programa;
6. ser uma avaliação produzida a partir do ano 2000.

A Figura 2 apresenta uma visão da distribuição atual das avaliações somativas no âmbito do SIPER. Observa-se que os estudos que tiveram por objeto políticas do Reino Unido, da Comissão Europeia e do Canadá correspondem, conjuntamente, em termos atuais, a mais de 55% desse repositório, evidenciando as três principais fontes de práticas avaliativas que atualmente se encontra ao se estabelecer uma análise agregada de como as diferentes equipes de avaliação vem tratando questões particulares às políticas de inovação a partir do SIPER.

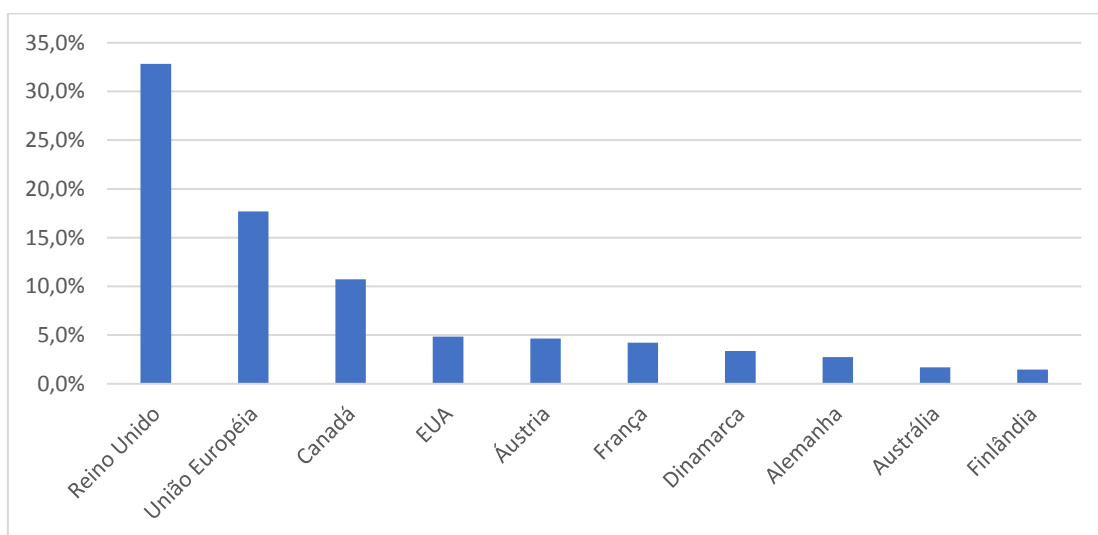


Figura 2 - Dez principais regiões objeto de avaliações catalogadas no âmbito do SIPER

Fonte – SIPER

A base de dados referente as avaliações somativas nacionais foi fruto de uma pesquisa bibliográfica sobre o Google Scholar e sobre os repositórios de instituições<sup>3</sup> selecionadas como mostradas no Quadro 3. De forma complementar, utiliza-se o cruzamento da listagem obtida a partir desse protocolo, com listagem coletada a partir da interação com

<sup>3</sup> Empregou-se como chave de busca sobre esses repositórios: (“avaliação do programa” OR “avaliação da política”) AND (ciência OR inovação) filetype:pdf inurl:<endereçoRepositório>

a equipe brasileira responsável pela manutenção dos dados do SIPER, a fim de se identificar possíveis omissões.

<b>Instituições</b>	<b>Quantidade de Estudos</b>
<b>BID</b>	2
<b>BNDES</b>	3
<b>CGEE</b>	11
<b>CNI</b>	1
<b>FAPEMIG</b>	1
<b>FGV</b>	1
<b>Inter-American Development Bank Washington, D.C</b>	1
<b>IPEA</b>	12
<b>RBI</b>	3
<b>UFRJ</b>	4
<b>Unicamp</b>	4
<b>USP</b>	5

Quadro 3 - distribuição dos estudos entre as instituições selecionadas.

Cabe ressaltar que com a quantidade de avaliações reunidas neste estudo, mesmo restringindo a amostra brasileira apenas a avaliações somativas, caso essas fossem submetidas e indexadas pela equipe do SIPER, o Brasil passaria a ser o terceiro maior país em número de avaliações registradas nesse repositório, sendo superado apenas pelo Reino Unido (164 estudos) e pelo Canadá (62).

Um último ajuste que deve ser explicitado foi a atenção em se identificar possíveis intersecções entre a amostra brasileira analisada e possíveis avaliações brasileiras no âmbito do SIPER. Neste caso, houve a incorporação desses estudos, caso eles tenham sido realizados sobre políticas federais. Ressalta-se que das quatro avaliações identificadas no SIPER, apenas uma atendeu esse critério.

### 3.1. FORMA DE CLASSIFICAÇÃO DAS AVALIAÇÕES NACIONAIS

De forma idêntica ao SIPER, a partir das avaliações coletadas, analisou-se uma a uma, por meio de análise textual, de forma a classificá-las em cada uma das categorias presentes no Quadro 4.

<b>Dimensões</b>	<b>Opções de Classificação</b>
<b>Posicionamento do Avaliador</b>	Interno ao Programa
	Externo ao Programa, porém interno ao governo
	Externo ao Programa e ao Governo
<b>Momento da Avaliação</b>	<i>Ex-Ante</i>
	Acompanhamento
	Interim
	<i>Ex-Post</i>
<b>Utilização da Lógica do Programa</b>	Sim
	Não
<b>Impactos Avaliados</b>	Científicos e Tecnológicos
	Econômicos
	Sociais
	Educação, Habilidade e Capacidades
	Ambiente
	Externalidades
<b>Retorno do Investimento</b>	Sim
	Não
<b>Design</b>	Experimental
	Quasi-Experimental (comparação antes-depois)
	Quasi-experimental (grupos de controle ou de comparação)
	Quasi-Experimental (beneficiários autodeclaração)
	Não-Experimental
<b>Bases Internas</b>	Sim
	Não
<b>Bases Externas</b>	Sim
	Não
<b>Surveys</b>	Sim
	Não
<b>Entrevistas</b>	Sim
	Não
<b>Métodos de Análise de Dados</b>	Estudo de Caso
	Análise de Redes
	Análise Econométrica
	Estatística Descritiva
	Análise do Retorno sobre Investimento
	Análise de Propriedade Intelectual
	Análise de Publicações
	Análise Altimétrica
	Análise de Texto

Quadro 4 - dimensões de análise das avaliações de programas federais de CT&I.

Fonte – SIPER

A partir dessas categorizações, busca-se comparar basicamente dois grupos: um deles composto pelas avaliações presentes no SIPER em extração realizada em julho de 2018, outro pelas avaliações de programas e políticas nacionais, no âmbito federal, coletadas no âmbito deste trabalho segundo a descrição acima. Busca-se assim, para cada uma das dimensões presentes no Quadro 4, permitir comparações entre comportamento avaliativo do SIPER e o comportamento empregado em avaliações de políticas federais, comparações essas que serão realizadas entre as estatísticas geradas por esse repositório e as decorrentes da análise realizada sobre a amostra nacional nesta pesquisa.

Dessas comparações, realiza-se discussões tendo como objetivo o alinhamento das situações identificadas a padrões já discutidos na literatura subjacente ao assunto de avaliação de políticas públicas (YIN, 2015).

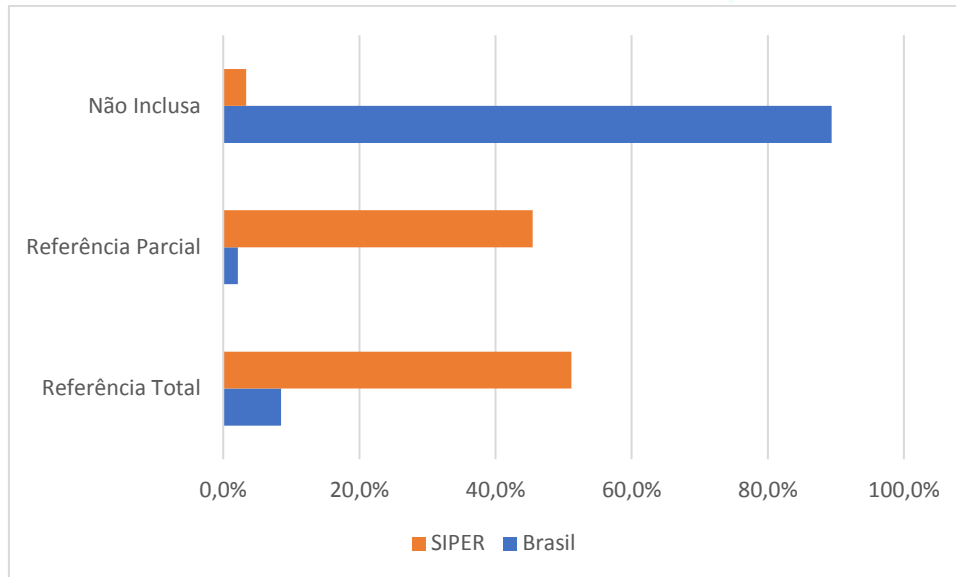
#### 4. RESULTADOS E DISCUSSÃO

Um primeiro fator avaliado foi o posicionamento dos avaliadores. Comparando a amostra brasileira ao SIPER, verificou-se um predomínio das avaliações externas. Contudo, diferentemente do encontrado naquele repositório, observou-se uma importância maior atribuída aos grupos avaliadores internos ao Governo. De fato, na esfera federal há duas instituições diretamente relacionadas à avaliação de políticas públicas, sendo que uma de caráter geral e outra no contexto dos programas de CT&I. Respectivamente trata-se do Instituto de Pesquisa Econômica Aplicada (Ipea) e o Centro de Gestão e Estudos Estratégicos (CGEE). Enquanto apenas 14% das avaliações do SIPER são externas, mas feitas por órgãos da Administração, no Brasil este número está em torno de 40%. Esse dado por si só ressalta o fato de que, embora essas avaliações possam ter sido realizadas por equipes que melhor entendessem os respectivos programas devido ao maior acesso às equipes, existe um maior risco à objetividade e à autonomia ao longo desses estudos,

por esse posicionamento ainda interno ao próprio Governo executor da política (WEISS, 1998).

A comparação entre as avaliações nacionais e o volume de avaliações registradas no SIPER mostra que , diferentemente, do encontrado no âmbito daquele repositório, as equipes desenvolvem seus processos avaliativos sem a preocupação de reconstruir, ao menos parcialmente, a teoria do programa subjacente, sendo as questões avaliativas muito mais decorrentes de hipóteses sobre a entrada de recursos e impactos de longo prazo, do que sobre o conhecimento da cadeia causal que levaria à efetividade da política. A Figura 3 mostra a comparação entre o padrão encontrado no SIPER e o padrão brasileiro de utilização da teoria do programa.

Observa-se que enquanto mais de 50% das avaliações daquele repositório se guiam pela lógica do programa, no caso brasileiro, cerca de 90% das avaliações analisadas neste trabalho não fizeram menção a esses mecanismos. De acordo com recentes painéis da OECD sobre o assunto, razões como busca por simplicidade e por quantificação de efeitos estão entre os motivos que agem como condicionantes para a desconsideração dos relacionamentos causais intermediários que atuam nas políticas de CT&I. A consequência imediata disso é a redução da usabilidade das avaliações no suporte ao redesenho dos mecanismos vulneráveis (OECD WORKSHOP ON ASSESSING THE IMPACTS OF PUBLIC RESEARCH SYSTEMS, 2016a).



*Figura 3 - Comparação entre a referência para teoria do programa entre as avaliações de impacto da amostra nacional e o SIPER*

*Fonte - Elaboração dos autores*

Apesar das avaliações do SIPER apresentarem essa orientação à teoria do programa, não há evidências de que as políticas de inovação de outras regiões, como a União Europeia tenham sido implementadas sem dificuldades quanto ao estabelecimento de modelos de impacto ou do esclarecimento de seus respectivos modelos lógicos. Para o período de catalogação dos estudos desse repositório, é sugerido que a reconstrução de conceitos e objetivos de políticas de CT&I no âmbito de países da Europa Ocidental é uma das principais dificuldades das equipes avaliativas, trazendo-se ambiguidade entre o que deve ser o papel de comunicação e planejamento da equipe responsável pela execução e coordenação do programa e da equipe responsável pela sua avaliação (BACH, 2012).

Sobre os impactos avaliados, apesar da multidimensionalidade das políticas de CT&I, identifica-se um perfil bastante diverso entre a amostra brasileira e o presente no SIPER. A Figura 4 ilustra essa comparação. Observa-se que no âmbito desse repositório, os efeitos de C&T, Econômicos e Capacidade estão presentes em torno de 70% das avaliações. O que reforça junto à comunidade internacional a abordagem de avaliação de múltiplas dimensões de impacto. Contudo, na amostra brasileira analisada, o efeito de maior avaliação – econômico – esteve em cerca de 45% das avaliações e o segundo maior – capacidades – em cerca de 40%, sugerindo um direcionamento para medições



de apenas um único tipo de efeito, desconsiderando os demais possíveis. De fato, da amostra, apenas 12% dos trabalhos avaliaram 2 tipos de impactos. Nenhum avaliou três ou mais.

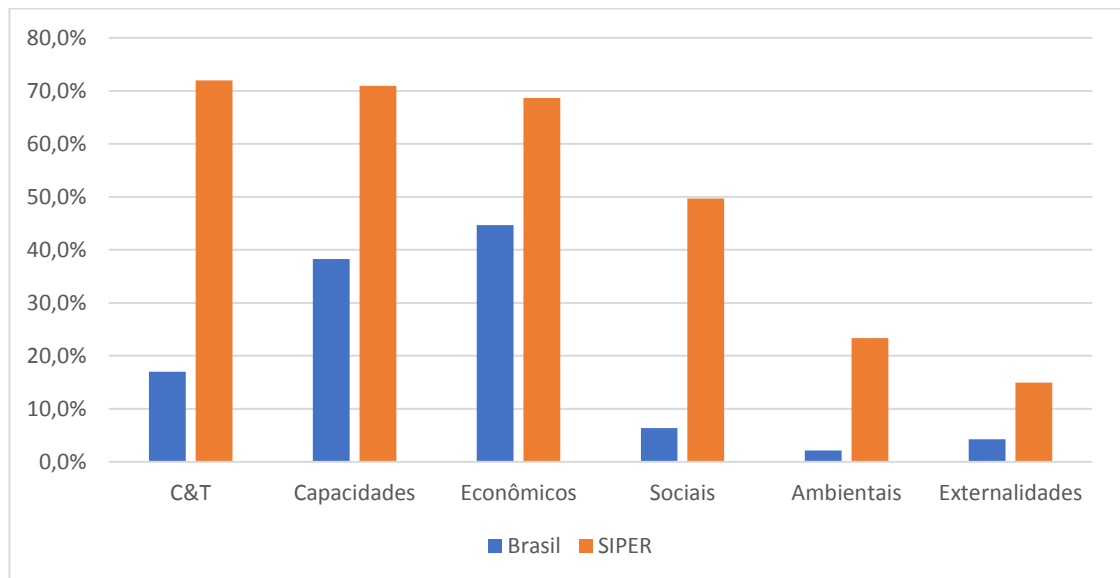


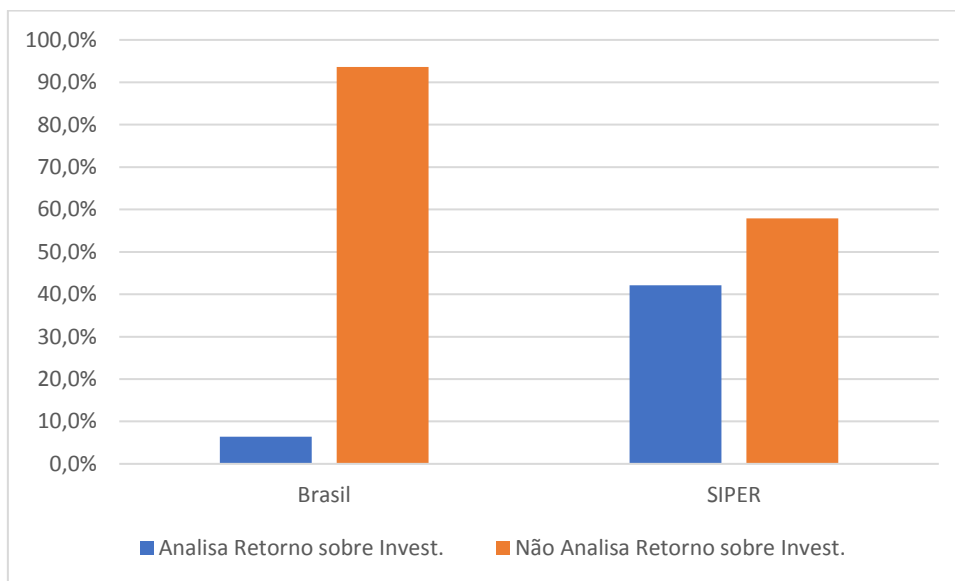
Figura 4 - Distribuição dos estudos avaliativos entre as categorias de impacto do SIPER

Fonte - SIPER

*Nota – A categoria “Capacidades” refere-se à “Capacidade Tecnológica” na definição de ARNOULD e THURIAUX (1997), dessa forma, apresentando alguma sobreposição com as categorias “C&T” (ex: capacidade tecnológica interna) e “Sociais” (ex: capacidade tecnológica externa). A categoria “Education, Skills and Capacity” do SIPER é considerada a correspondente dessa.*

Observa-se que a questão de multiplicidade de causalidade em políticas de CT&I é uma das peculiaridades presentes nesse domínio. Além disso, como apresenta GODIN e DORÉ (2004), há uma vasta tipologia de efeitos relacionados a políticas de inovação, dos quais os econômicos são apenas aqueles cujo o ferramental é o que se encontra em estágio mais consolidado, permitindo assim maior segurança quanto à mensuração. O panorama apresentado na Figura 4 é de uma subavaliação das políticas federais brasileiras, implicando em desconsideração de efeitos positivos ou negativos em dimensões relevantes (SALLES, 2010). Aliado ao desconhecimento dos modelos lógicos desses programas, pode-se identificar o risco de que atores relevantes ao alcance do objetivo de políticas de CT&I possam estar deixando de ser envolvidos devido à desconsideração de perspectivas de impacto relevantes, o que pode revelar não apenas um problema de monitoramento e avaliação, mas de planejamento e coordenação entre

as partes interessadas de um programa de inovação (BRASIL, 2014). Apesar de se observar que, na amostra analisada de avaliações de políticas de CT&I federais, houve um maior foco para se avaliar efeitos econômicos, identifica-se que não houve uma igual ênfase em se medir o retorno sobre o investimento resultante ou esperado dessas políticas. Enquanto no âmbito do SIPER há um relativo equilíbrio entre os estudos, com cerca de 40% deles apresentando essa avaliação, entre as avaliações brasileiras analisadas cerca de 95% delas não se voltaram para a análise desse retorno. A Figura 5 apresenta essa visão comparativa.



*Figura 5 - Comparação entre a amostra brasileira de avaliações e o SIPER no que diz respeito a consideração do retorno sobre investimento*

*Fonte - Elaboração dos autores*

Um efeito disso é a dificuldade de comparação de diferentes intervenções que visem tratar o mesmo problema (em um esforço *ex-ante* de alocação de recursos), deixando às partes interessadas apenas uma visão de diferentes alternativas com impactos positivos (ou negativos) sobre questões similares. O emprego dessas técnicas no âmbito do SIPER mostra que, mesmo na prática, a questão de se empregar técnicas de análise de custo-benefício, mesmo envolvendo resultados e efeitos por vezes intangíveis como no domínio de CT&I não é um obstáculo intransponível. Metodologias para lidar com o problema são várias, como as descritas no *Green Book* do Reino Unido (TREASURY, 2003).

Um perfil avaliativo bastante distinto foi encontrado ao se analisar os desenhos das avaliações. O que se observa é que mais das metades das avaliações da amostra brasileira optaram por um desenho quasi-experimental, sendo que, destas, quase a totalidade é baseada no emprego de grupos de controle (associados à análise com técnicas econométricas). No SIPER, é verificado que cerca de 90% das avaliações então registradas incluíram desenhos não experimentais, desenho este presente apenas em cerca de 40% dos estudos da amostra brasileira. A Figura 6 apresenta essa comparação.

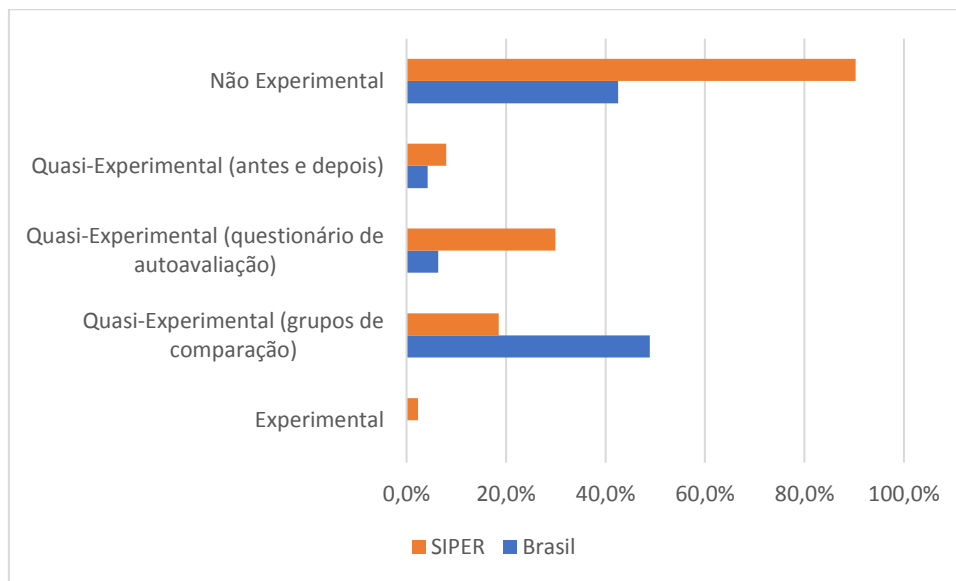


Figura 6 – Comparação entre os desenhos avaliativos empregados pela amostra brasileira de avaliações federais e o SIPER.

Fonte - SIPER

Uma observação decorrente das diferenças de perfis presentes na Figura 6 entre a amostra brasileira e o SIPER é a ênfase que as avaliações presentes neste último repositório apresentaram em *designs* não experimentais, mais qualitativos, compatíveis com uma compreensão mais comportamental dos fenômenos estudados, o que é coerente com o fato de que, ao mesmo tempo que 87% de suas avaliações são classificadas como somativas e 82% como formativas, pela Teoria dos Conjuntos<sup>4</sup> temos que 69% do SIPER é composto por estudos que visam mais do que quantificar impactos, mas também estudar a dinâmica que os provoca.

<sup>4</sup> Cardinalidade (AUB) = Cardinalidade (A) + Cardinalidade (B) – Cardinalidade (A∩B)

Entretanto, o conjunto brasileiro apresenta uma ênfase bem maior sobre o fenômeno de identificação de causalidade, o que explica a maior utilização de *designs* quasi-experimentais. Contudo, diferentemente do observado na amostra brasileira, é observado no âmbito do SIPER que a comunidade avaliadora de CT&I não se tem limitado diante da questão de construção de grupos de controle para mensuração desses impactos, uma vez que o desenho mais comum entre os estudos internacionais tem sido a partir da aplicação de questionários em que os próprios beneficiários avaliam os efeitos dessas políticas. Comparativamente, esta estratégia do SIPER é mais vulnerável principalmente a riscos sobre a validade interna<sup>5</sup> de suas conclusões, contudo, complementada com outros métodos, pode trazer insumos importantes ao processo decisório, melhor do que não ter avaliação alguma sobre o assunto (COOK, CAMPBELL, SHADISH, 2002).

As avaliações federais brasileiras analisadas apresentam um perfil diverso quanto à parte de coleta de dados. Observa-se que o SIPER apresenta um perfil bem mais intenso no que diz respeito à utilização de *surveys* e entrevistas, em que a proporção de utilização chega próximo do dobro da amostra brasileira, ratificando, em parte, a tendência das avaliações desse repositório em complementar suas abordagens quantitativas, muitas vezes cujas informações já estão presentes em repositórios secundários – como bancos de dados - com técnicas para busca de dados primários, muitas vezes para análises qualitativas. Complementando a análise do comportamento de coleta, observa-se que nas avaliações brasileiras houve um uso maior de bases externas para a execução das avaliações, enquanto que, no SIPER, a maior proporção está sobre bases internas. Esse comportamento pode indicar uma dificuldade para acesso aos dados de certas políticas de inovação brasileiras. Uma vez que o acesso a bases de dados da esfera federal está garantido por mecanismos legais, causas como deficiências na preparação, na validação e na modelagem dos dados podem influenciar (Figura 7).

---

<sup>5</sup> Refere-se ao relacionamento entre as proposições iniciais de um avaliador e as suas conclusões finais. Para maiores informações, YIN(2015).

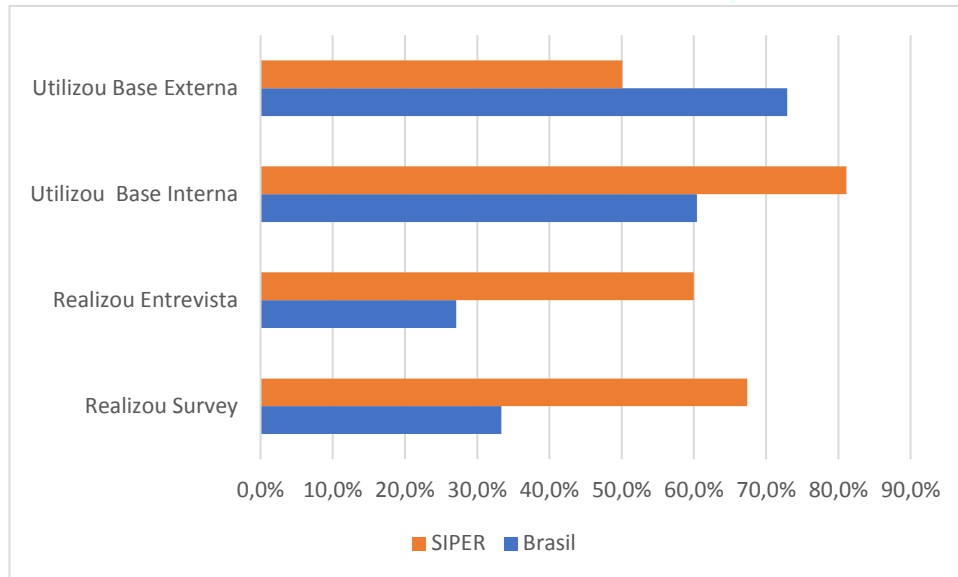


Figura 7 - Métodos de coleta de dados das avaliações analisadas

Fonte - elaboração dos autores

A predominância relatada de *designs* quasi-experimentais veio acompanhada de um alto emprego de técnicas quantitativas de análise. Mais de 70% dos estudos empregam algum nível de análise estatística para reforçar suas afirmações e 50% das avaliações utilizam análises econométricas. Em particular, cabe observar que 100% dos desenhos experimentais que empregaram grupos de controle terminaram utilizando econometria. Entretanto, reforçando uma observação colocada por EDLER (2012), ao analisar as avaliações realizadas no âmbito da União Europeia, não necessariamente o emprego dessas técnicas irá resultar em maior aceitação dos gestores e, conseqüentemente, em utilidade dos estudos. A Figura 8 apresenta uma visão dessa comparação.

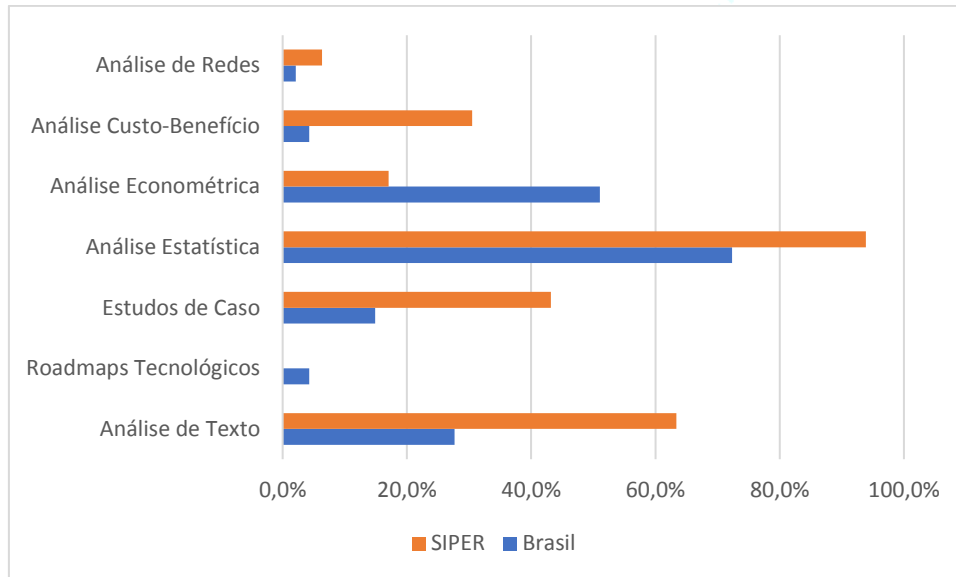


Figura 8 - Comparação entre métodos de análise de dados empregado pela amostra de avaliações federais e o SIPER.

Fonte - elaboração dos autores.

A predominância de métodos quantitativos de análise em proporção bem mais elevada que os métodos qualitativos, na amostra brasileira, sugerem uma falta de integração nas avaliações entre essas duas abordagens e, conseqüentemente, uma escolha pelas equipes avaliadoras de uma abordagem ou outra, em vez da sua conjugação para abordar as diferentes dimensões de análise, conforme sugere BAMBERGER (2012). Essa desconexão já havia sido sugerida pelo perfil dos *designs* avaliativos. Uma possível consequência disso são avaliações robustas quanto ao impacto, principalmente o econômico, mas sem maiores subsídios aos gestores quanto às vulnerabilidades a serem tratadas nos casos em que as intervenções apresentam impactos insatisfatórios, devido à limitação dos desenhos e das técnicas de análise escolhidas.

A questão se agrava à medida que consideramos que a lógica particular dos programas não é considerada pelas equipes avaliadoras, conforme discutido. Fragiliza-se, assim, os estudos produzidos ao colocar as intervenções de CT&I como as principais responsáveis por efeitos de longo prazo que na verdade são condicionados pela ação conjunta de fatores de *framework* como política tributária, política comercial, infraestrutura, entre outros fatores. Como apresenta estudo da OECD sobre o assunto a partir de experiência junto à agência de sueca Vinnova – uma política de inovação deve ser considerada uma condição necessária, mas não suficiente para a ocorrência de um impacto (OECD



WORKSHOP ON ASSESSING THE IMPACTS OF PUBLIC RESEARCH SYSTEMS, 2016b).

## 5. CONCLUSÃO

O SIPER, em seu estágio atual, apresenta um predomínio de avaliações sobre políticas de regiões como União Europeia e Canadá, o que faz com que uma comparação junto a esse repositório tenha como um dos principais benefícios a identificação de práticas avaliativas adotadas em outros sistemas de inovação e que poderiam ser aproveitadas. Este foi o objetivo principal do presente trabalho.

Trata-se de um repositório ainda em construção, mas que pode oferecer reflexões sobre como outras comunidades avaliativas de políticas de CT&I estão tratando problemas semelhantes aos locais. Das principais questões discutidas no âmbito deste trabalho, tornou-se evidente a diferença de ênfase entre a prática da amostra brasileira analisada e as práticas do SIPER no que diz respeito ao emprego da teoria do programa no âmbito das avaliações, à avaliação de múltiplas dimensões, à avaliação do retorno sobre o investimento e à utilização de estratégias de pesquisa não experimentais. Por outro lado, observou-se que, se essa comunidade se expõe mais ao risco de validade interna por meio da atribuição de causalidade por meio de questionários de autoavaliação, as avaliações nacionais apresentaram desenhos quasi-experimentais mais robustos, baseados em grupos de comparação. O custo benefício dessa limitação de desenho sobre o processo decisório de CT&I não foi medido, tornando-se um risco a se tratar para as políticas federais.

Espera-se que a partir da identificação dessas diferenças mais significativas entre esses perfis avaliativos, seja possível para as equipes responsáveis pelo planejamento de avaliações de políticas somativas de CT&I identificarem práticas adotadas internacionalmente que possam potencialmente melhorar os resultados do trabalho – ressalta-se, por exemplo, o emprego no âmbito do Poder Executivo federal brasileiro de um guia de avaliação *ex-ante* de políticas públicas (BRASIL, 2018), o qual traz ênfase na construção de modelos lógicos e na demonstração do retorno econômico, implicando na

necessidade imediata para as políticas de CT&I de melhor aproveitamento de sua multidimensionalidade e de meios de mostrar sua priorização sobre outras políticas concorrentes. a mudança efetiva da política pública e não simplesmente a publicação do estudo. Como trabalho futuro, sugere-se que a utilidade das avaliações junto aos gestores seja estudada, em analogia como o realizado em EDLER (2012).

## REFERÊNCIAS

- BRASIL. Ministério da Transparência e Controladoria-Geral da União. **Relatório de Auditoria nº 201505715 - Estratégia Nacional de Ciência Tecnologia e Inovação – ENCTI**. Brasília, DF, 27 out 2016. Disponível em: <<https://auditoria.cgu.gov.br/download/8602.pdf>>. Acesso em: 16 nov. 2018.
- BRASIL. Instituto de Pesquisa Econômica Aplicada. **Avaliação de Políticas Públicas - Guia Prático de Análise Ex Ante - Volume 1**. Brasília. 2018. Disponível em: <[http://www.ipea.gov.br/portal/images/stories/PDFs/livros/livros/180319\\_avaliacao\\_de\\_politicas\\_publicas.pdf](http://www.ipea.gov.br/portal/images/stories/PDFs/livros/livros/180319_avaliacao_de_politicas_publicas.pdf)>. Acesso em 21 fev. 2019.
- BRASIL. Tribunal de Contas da União (TCU). **Referencial para Avaliação de Governança em Políticas Públicas**. Brasília, 2014.
- BAMBERGER, Michael. **Introduction to mixed methods in impact evaluation**. *Impact Evaluation Notes*, v. 3, n. 3, p. 1-38, 2012.
- BACH, Laurent. **The frontiers of evaluation: some considerations on the European case as Fronteiras da avaliação: algumas considerações sobre o caso europeu**. *RBI-Revista Brasileira de Inovação*, v. 11, p. 67-84, 2012.
- COOK, Thomas D.; CAMPBELL, Donald Thomas; SHADISH, William. **Experimental and quasi-experimental designs for generalized causal inference**. Boston: Houghton Mifflin, 2002.
- CORAZZA, Rosana Icassatti; FRACALANZA, Paulo Sérgio. **Caminhos do pensamento neo-schumpeteriano: para além das analogias biológicas**. *Nova Economia*, v. 14, n. 2, 2009.
- EDLER, Jakob et al. **The practice of evaluation in innovation policy in Europe**. *Research Evaluation*, v. 21, n. 3, p. 167-182, 2012.
- FREY, Bruce B. (Ed.). **The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation**. SAGE Publications, 2018.
- GERTLER, Paul J. et al. **Impact evaluation in practice**. The World Bank, 2016.
- GEORGHIOU, Luke (Ed.). **The handbook of technology foresight: concepts and practice**. Edward Elgar Publishing, 2008.

GODIN, Benoit; DORÉ, Christian. **Measuring the impacts of science: Beyond the economic dimension**. History and sociology of S&T statistics, 2004.

JARAMILLO, I. et al. **Manual de Bogotá: normalización de indicadores de innovación tecnológica en América Latina y el Caribe**. OEA, RICYT, COLCIENCIAS, CYTED, OcyT: Bogotá, Colombia, 2000

LINK, Albert N.; VONORTAS, Nicholas S. (Ed.). **Handbook on the theory and practice of program evaluation**. Edward Elgar Publishing, 2013.

MARCOVITCH, Harvey et al. **Conflict of Interest in Science Communication: More than a Financial Issue Report from Esteve Foundation Discussion Group, April 2009**. Croatian medical journal, v. 51, n. 1, p. 7-14, 2010.

MILES, Ian; SARITAS, Ozcan; SOKOLOV, Alexander. **Foresight for science, technology and innovation**. Switzerland: Springer International Publishing, 2016.

NGAI, E. W. T.; LEE, P. T. Y. **A Review of the literature on Applications of Text Mining in Policy Making**. In: PACIS. 2016. p. 343.

OECD. **Enhancing Research Performance through Evaluation, Impact Assessment and Priority Setting**. Paris, OECD, 2009. Disponível em: <<http://www.oecd.org/sti/inno/Enhancing-Public-Research-Performance.pdf>>. Acesso em: nov. 2018

OECD WORKSHOP ON ASSESSING THE IMPACTS OF PUBLIC RESEARCH SYSTEMS, 2016, Paris. **Approaches to Impact Assessment**, Paris, OECD, 2016a. Disponível em: <<http://www.oecd.org/sti/inno/Mechanisms-OECDImpact.pdf>>. Acesso em: nov. 2018.

OECD WORKSHOP ON ASSESSING THE IMPACTS OF PUBLIC RESEARCH SYSTEMS, 2016, Paris. **Causality Problems**, Paris, OECD, 2016b. Disponível em: <<http://www.oecd.org/sti/inno/Mechanisms-OECDImpact.pdf>>. Acesso em: nov. 2018.

RAUEN, André Tortato. **Avaliação de políticas federais de inovação: desconexão e ausência**. In: COUTINHO, Diogo R; FOSS, Maria Carolina; MOAULLEM, Pedro Salomon B. (Org.). Inovação no Brasil: avanços e desafios jurídicos e institucionais. São Paulo: Blucher, 2017.

ROGERS, Patricia J. **Causal models in program theory evaluation. New directions for evaluation**, v. 2000, n. 87, p. 47-55, 2000.

ROGERS, Patricia J. **Using programme theory to evaluate complicated and complex aspects of interventions**. Evaluation, v. 14, n. 1, p. 29-48, 2008.

SALLES-FILHO, Sergio LM et al. **Multidimensional assessment of technology and innovation programs: the impact evaluation of INCAGRO-Peru**. Research Evaluation, v. 19, n. 5, p. 361-372, 2010.

TREASURY, Her Majesty's. **The green book: appraisal and evaluation in central government**. 2003.

WEGRICH, Kai; JANN, Werner. **Theories of the policy cycle**. In: **Handbook of public policy analysis**. Routledge, 2006. p. 69-88.

---

WEISS, Carol H. **Methods for studying programs and policies**. Weiss, London, Prentice Hall, 1998.

YIN, Robert K. **Estudo de Caso-: Planejamento e Métodos**. Bookman editora, 2015.