

## **An impact of an public investment in science into human resources formation at masters and doctoral levels: The case of Universal 2014**

**Abstract:** In this article we present an assessment of an investment in scientific research (Universal 2014) carried out by the Brazilian government through the National Council for Scientific and Technological Development (CNPq). Its impact on education was estimated by measuring formation of masters and PhDs. We used multivariate regression in a quasi-experiment, measuring the correlation between the amount invested in Universal 2014 and the number of formations. The methodology was recommended by the World Bank and the data taken from the Lattes Curriculum, a public curriculum maintained by the Brazilian government. There was a high impact on human resource formation, indicating the need for greater investment by the government and showing a relationship between investment in science and improvement in education. Finally, a list of the possible impact of other indicators is presented, pointing to an impact far above the amount invested.

**Keywords:** Impact Assessment. Human Resources Training. Quasi-Experiment. Curriculum Lattes. Public Policy for Investment in Science.

## **O Impacto de um investimento público em ciência na formação de recursos humanos em nível de mestrado e doutorado: o caso do Universal 2014**

**Resumo:** Neste artigo apresentamos uma avaliação de um investimento em pesquisa científica (Universal 2014) realizado pelo governo brasileiro por meio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Seu impacto na educação foi estimado pela mensuração da formação de mestres e doutores. Utilizamos a regressão multivariada em um quase-experimento, medindo a correlação entre o valor investido no Universal 2014 e o número de orientações concluídas. A metodologia foi recomendada pelo Banco Mundial e os dados retirados do Currículo Lattes, currículo público mantido pelo governo brasileiro. Houve alto impacto na formação de recursos humanos, indicando a necessidade de maior investimento por parte do governo e mostrando uma relação entre investimento em ciência e a melhoria na educação. Por fim, é apresentada uma lista do possível impacto de outros indicadores, apontando para um impacto muito acima do valor investido.

**Palavras-chave:** Avaliação de Impacto. Formação de Recursos Humanos. Quase-Experimento. Currículo Lattes. Política Pública de Investimento em Ciência.

## **Un impacto de la inversión pública en ciencia en la formación de recursos humanos a nivel de maestría y doctorado: el caso de Universal 2014**

**Resumen:** En este artículo presentamos una evaluación de una inversión en investigación científica (Universal 2014) realizada por el gobierno brasileño a través del Consejo Nacional de Desarrollo Científico y Tecnológico (CNPq). Su impacto en la educación se estimó midiendo la formación de maestrías y doctorados. Utilizamos regresión multivariante en un cuasiexperimento, midiendo la correlación entre el monto invertido en Universal 2014 y el número de formaciones. La metodología fue

recomendada por el Banco Mundial y los datos tomados del Currículo Lattes, un currículo público mantenido por el gobierno brasileño. Hubo un alto impacto en la formación de recursos humanos, lo que indica la necesidad de una mayor inversión por parte del gobierno y muestra la relación entre la inversión en ciencia y la mejora en la educación. Finalmente, se presenta una lista del posible impacto de otros indicadores, apuntando a un impacto muy por encima del monto invertido.

**Palabras clave:** Evaluación de Impacto. Formación de Recursos Humanos. Cuasiexperimento. Curriculum Lattes. Política Pública de Inversión en Ciencia.

## 1 Introdução

O Plano Nacional de Educação (PNE), [Lei nº 13.005/2014](#), traz em suas metas 14 e 16 “Elevar gradualmente o número de matrículas na pós-graduação de modo a atingir a titulação anual de 60.000 (sessenta mil) mestres e 25.000 (vinte e cinco mil) doutores” e “Formar, em nível de pós-graduação, 50% (cinquenta por cento) dos professores da educação básica, até o último ano de vigência deste PNE, e garantir a todos (as) os (as) profissionais da educação básica formação continuada em sua área de atuação, considerando as necessidades, demandas e contextualizações dos sistemas de ensino”. Apesar das metas indicadas, Cabrera e Araújo (2018) verificaram uma diminuição no número de mestrados profissionalizantes, tanto em matrículas quanto em formações, levantando a necessidade de políticas públicas que respondam aos desafios atuais e atendam as crescentes demandas da educação nacional. Essa diminuição também pode estar ocorrendo no mestrado acadêmico e no doutorado.

Uma das formas de atender as metas do PNE é o fomento à formação de recursos humanos, que, de acordo com (WILBERTZ, 2013), pode se dar direta ou indiretamente, pelo setor público ou privado. A formação indireta se dá por meio de investimento em projetos de CT&I, enquanto a direta é o financiamento a universidades e seus programas de PPG. No Brasil, o fomento público indireto é realizado por órgãos tanto estaduais quanto federais. No âmbito federal, o fomento indireto à formação de mestres e doutores é feito por meio do Ministério da Ciência, Tecnologia e Inovação (MCTI), através do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e por meio do Ministério da Educação (ME), através da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Enquanto cabe ao CNPq “promover e fomentar o desenvolvimento e a manutenção da pesquisa científica e tecnológica e a formação de recursos humanos qualificados para a pesquisa, em todas as áreas do conhecimento” (COMPETÊNCIAS

— PORTUGUÊS (BRASIL), [s. d.]), cabe à Capes papel na expansão e consolidação da pós-graduação *stricto sensu* (mestrado e doutorado) em todos os estados da Federação (CAPES, 2012). Assim, enquanto compete ao CNPq, que é ligado ao MCTI, o financiamento à pesquisa, à tecnologia e à inovação, compete à Capes, ligada ao ME, o financiamento para a formação de recursos humanos. No entanto, no que diz respeito a formação de mestres e doutores, essas tarefas se confundem, uma vez que boa parte da pesquisa brasileira se liga a programas de pós-graduação.

O CNPq foi criado em 1951 e tem como missão fomentar e atuar nas políticas públicas de CT&I. Uma forma de compreender o papel do CNPq é analisar as ações de financiamento realizadas. Em 2021 o CNPq executou pagamentos no valor de R\$ 1.131.524.331,05 (um bilhão de reais) <sup>1</sup>. Até o ano de 2016, o CNPq havia lançado 975 ações do tipo “Edital” e 549 do tipo “Bolsas e Auxílios”<sup>2</sup>. Dessas, três ações se destacam, tanto pelo volume total solicitado na demanda quanto no volume de recursos financiados. Os Editais Universais representaram 14,36%<sup>3</sup> do total (R\$ 17.012.482.313,00), os Institutos Nacionais de Ciência e Tecnologia obtiveram 11,65% (R\$ 13.793.092.584,00) e Bolsas Produtividade ficaram com 10,04% (R\$ 11.890.687.902,00). Essas três ações também representam três formas de atuação do CNPq: investimento em projetos de CT&I, investimento direto em grandes centros de tecnologia e concessão de bolsas.

Ocorre que a formação de recursos humanos exigido pelo PNE pode resultar de todas as três formas de investimento acima citados. Políticas públicas trazem efeitos tanto diretos quanto indiretos. Os efeitos indiretos estão de acordo com a teoria da Hélice Tripla em que, de acordo com Leydesdirff e Ivanova (2016), há três eixos, ortogonais, (1) relativo à produção do conhecimento, (2) relativo à geração de riquezas (comércio e indústria) e (3) relativo ao controle normativo (governância), de modo que cada vértice formado pode tanto inibir quanto catalisar alterações nos outros dois vértices, em sistema dinâmico e autogerido. Não é somente a educação que afeta a ciência e o desenvolvimento, mas a ciência também afeta a educação, tanto diretamente, quanto indiretamente, sendo que a governância ainda é afetada e afeta os outros eixos. Assim, tanto o investimento em projetos de CT&I, quanto o

<sup>1</sup> <https://www.portaltransparencia.gov.br/orgaos/20501-conselho-nacional-de-desenvolvimento-cientifico-e-tecnologico>

<sup>2</sup> Relatório tirado por meio de uma ferramenta interna ao CNPq que foi descontinuada em 2016.

<sup>3</sup> Os valores apresentados são a demanda bruta, ou seja, o total de recursos solicitados. O total aprovado é muito inferior, mas segue aproximadamente a mesma proporção.

investimento direto em grandes centros de tecnologia, podem trazer resultados indiretos em termos de formação de recursos humanos, tanto mestres quanto doutores.

O efeito de investimentos em projetos de CT&I na formação humana é conhecido e esperado pelo CNPq. No questionário de Prestação de Contas Final<sup>4</sup> das chamadas Universal, um dos itens de preenchimento obrigatório é justamente a formação realizada, em termos de iniciação científica, mestres, doutores. O CNPq espera o mesmo resultado indireto com as outras duas ações anteriormente citadas: investimentos a grandes centros de tecnologia e concessão de bolsas de Produtividade em Pesquisa. Assim, o CNPq espera como parte dos resultados da política pública de investimento em CT&I a formação de mestres e doutores.

No Brasil, a avaliação das políticas públicas, inclusive com divulgação do objeto a ser avaliado e dos resultados alcançados, é obrigação imposta pela Constituição Brasileira (art. 37, § 3º-I, § 16, art. 74, I e II, art. 165, § 16, art. 193). Na educação, a Lei de Diretrizes e Bases da Educação Nacional (LDB) assegura a avaliação nacional da educação em todos os seus níveis. No investimento em CT&I, o Novo Marco Legal (lei nº 13.243/2016) prevê prestação de contas simplificada por parte dos pesquisadores e o decreto nº 9.283, que a regulamenta, institui avaliações de resultados como um de seus princípios básicos.

A legislação brasileira não só exige que avaliações sejam realizadas para todas as políticas públicas de investimento, mas também exige que essas sejam realizadas com transparência e objetividade. A transparência é princípio da Constituição Federal e também regulamentado pela Lei 12.527, a Lei da Transparência. Nela, é garantido o direito de acesso à informação de forma objetiva, transparente, clara e em linguagem de fácil compreensão.

Desta forma, esta pesquisa busca uma avaliação do efeito indireto do investimento em CT&I, particularmente da Chamada Universal 2014 do CNPq, na formação de mestres e doutores. Mais especificamente, busca-se medir a correlação entre o valor investido em um projeto de pesquisa da Chamada Universal 2014 e a formação humana resultante, ou seja, uma forma de impacto desse valor investido. Secundariamente, busca-se comparar a correlação medida com os investimentos

---

<sup>4</sup> Os questionários não estão publicamente disponíveis, mas podem ser adquiridos por meio da Lei de Acesso à Informação.

diretos realizados na forma de bolsas, de modo a avançar no tema sobre práticas de governança na avaliação do investimento realizado na pesquisa científica, atendendo a necessidade legal de avaliação com transparência. Finalmente, pretende-se indicar se há outras correlações, indicando outros possíveis impactos e futuras pesquisas no tema a realizar.

Na introdução deste artigo apresentou-se a contextualização do tema estudado, a motivação do estudo e os objetivos da pesquisa. A seção 2 apresenta a metodologia. A seção 3 apresenta o Currículo Lattes, de onde a maior parte dos dados foram retirados, e a 4 detalha a forma de obtenção dos indicadores necessários e os procedimentos adotados para a extração dos dados. A seção 5 dá os resultados da pesquisa e, por fim, na 6 temos as considerações finais e sugestões para trabalhos futuros.

## 2 Metodologia

Segundo (BORNMANN, 2013), o impacto da ciência é a avaliação do retorno social, cultural, ambiental e econômico, de forma que o retorno social representa a contribuição ao capital social da nação (estimulando, por exemplo, novas abordagens a problemas sociais, debates públicos ou melhorando a política pública), o retorno cultural é a contribuição para a preservação ou ganho do patrimônio cultural, como, por exemplo, o entendimento de como as culturas ou sociedade se relacionam, o retorno ambiental adiciona ao capital natural da nação (redução de lixo ou poluição, reciclagem, redução do desmatamento ou produção de energia limpa) e o retorno econômico estaria relacionado a, por exemplo, capital humano (habilidades, técnicas, produtividade e conhecimento). Dessa forma, eventual correlação entre uma política pública de investimento em CT&I e a formação de recursos humanos pode ser considerado como um dos impactos dessa política.

Para medir esse impacto, seguiu-se a metodologia indicada pelo Banco Mundial, em seu manual intitulado Avaliação de Impacto na Prática (GERTLER *et al.*, 2018). Nele, o impacto de uma política pública é medido pela diferença entre o estado atual daquele que recebeu o tratamento (neste caso, a política pública de investimento em ciência, ou seja, o financiamento) e aquele estado de como estaria esse mesmo indivíduo caso essa política nunca tivesse existido (o contrafactual).

No entanto, o contrafactual é apenas teórico, uma vez que é impossível afirmar

qual seria o estado do indivíduo sem a política e, por isso, precisa ser estimado (GERTLER *et al.*, 2018). A forma mais básica de se estimar o contrafactual é supor que, sem o tratamento, não haveria alterações a serem medidas. Ou seja, supõe-se que o estado se manteria no tempo e, para calcular o impacto, basta verificar o estado atual daquele que recebeu o tratamento e comparar com o seu estado anterior. Assim, mede-se a mudança que o tratamento causou no tratado, supondo que o contrafactual seria estado idêntico ao do beneficiado antes de eventual financiamento. Ocorre que, ao longo do tempo, outros fatores externos ao tratamento podem afetar a pessoa tratada, fazendo com que a diferença entre os estados atual e inicial não seja devido apenas ao efeito do tratamento. Assim, o contrafactual apresenta viés, que é a diferença entre o estado estimado e o estado que seria o real do tratado (GERTLER *et al.*, 2018).

Dessa forma, para aumentar a precisão da medida do impacto, utiliza-se um grupo de controle, um grupo que não recebeu o financiamento e que permite que se estime outros fatores externos. A escolha de quem pertencerá a esse grupo, no entanto, é complexa e pode definir o estudo realizado. De acordo com o Banco Mundial, a melhor metodologia é a experimental, em que escolher-se-ia todos os elegíveis para o tratamento e aplicar-se-ia o mesmo apenas em alguns indivíduos aleatoriamente selecionados. A seleção aleatória tem a função de garantir que, na média, estima-se corretamente todas as mudanças externas ao tratamento realizado. Subtraindo esses efeitos ao grupo de tratamento, temos o impacto da política pública. Essa é a metodologia experimental.

No entanto, como se trata de análise ex-post, não é possível a seleção aleatória e o grupo de controle precisa ser substituído por uma estimativa estatística. Uma opção seria calcular a média dos indicadores dos grupos de tratamento e controle, no período inicial e final indicados, para depois apenas se subtrair a média desses indicadores. O método tem cálculo simples e dá o impacto médio de forma direta, mas tem a desvantagem de não indicar o grau de confiança estatística, ou seja, possui uma maior chance de nos dar indicador de impacto superestimado ou irreal.

É por esse motivo que se adotou a metodologia indicada pelo Banco Mundial (GERTLER *et al.*, 2018) da regressão multivariada. Nessa metodologia, utiliza-se de ferramentas estatísticas para se criar um grupo de controle artificial. O método supõe que todos os fatores externos à política pública que afetaram o grupo de tratamento

também afetaram o grupo de controle, e vice-versa. É, portanto, um método quase-experimental.

Na regressão multivariada, uma variável busca medir a diferença, no tempo, entre o grupo que recebeu e o que não recebeu o financiamento. Uma vez que essa diferença é tratada por uma variável independente, a segunda variável passa a medir o efeito do tratamento.

A regressão linear multivariada possui diversos benefícios sobre a diferença simples, por usar de métodos estatísticos que dão a indicação da real representatividade do indicador, eliminando resultados que poderiam não ser reais. Dessa forma, este estudo seguirá o modelo adotado por Milesi et al. (2014), que usaram de regressão linear multivariada para comparar a produtividade de pesquisadores, antes e depois de um financiamento (REESE Award). Milesi et al. (2014) deixam claro que, quando possível, a melhor metodologia compara quem concorreu à mesma política pública mas não foi contemplado (grupo de controle) com quem concorreu e foi selecionado (grupo de tratamento). O Banco Mundial propõe metodologia semelhante, assim, este estudo adota essa comparação.

Milesi et al. (2014) se utilizam da seguinte equação:

$$y_{ti} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 after + \beta_4 after^2 + \sum_c \beta_c x_{ci} + u_i + r_{ti}$$

Equação 1 – Modelo de Regressão Linear para comparar a Produtividade no Tempo por meio de um Estudo Bibliométrico (MILESI *et al.*, 2014).

Na Equação 1,  $i$  se refere a um índice de beneficiados,  $t$  é o tempo,  $after$  é o tempo após a política pública,  $\sum_c \beta_c x_{ci}$  são covariáveis incluídas no modelo,  $u_i$  é variação aleatória entre beneficiados que não é uma função do tempo e  $r_{ti}$  é variação aleatória ano a ano dos pesquisadores financiados ( $u_i + r_{ti}$  estimam o erro da equação).

Ocorre que Milesi et al. (2014) se utilizam de uma série temporal, enquanto este estudo compara apenas dois pontos no tempo. A análise temporal não é necessária, uma vez que são usados apenas dois pontos temporais de referência necessários para a análise do impacto (o inicial e o final). Dessa forma, as variáveis independentes quadráticas e as constantes que refletem seus erros, ou seja, as variáveis que buscam estimativa de progressão temporal e que refletem as alterações temporais não

lineares, não serão aqui necessárias.

Além disso, Milesi et al. (2014) adotam soma de variáveis correlatas que ajustam para ano de doutorado, financiamento por outras agências de fomento, nota da instituição de vínculo, área de estudo e o sexo. Muitos desses dados não estão disponíveis e, portanto, não foram diretamente utilizados. No entanto, entende-se que todas essas variáveis estão altamente correlacionadas com o nível do indicador antes do financiamento. Dessa forma, a equação será alterada para:

$$y_{ti} = \beta_0 + \beta_3 after + \beta_c x_{ci} + u_i$$

Equação 2 – Modelo de Regressão Linear sem variáveis temporais.

Finalmente,  $\beta_0$  e  $u_i$  são constantes, que representam não só o valor inicial do estado, mas também o eventual erro. Dessa forma, as duas podem ser unidas em apenas uma, uma vez que buscamos apenas o  $\beta_3$  (o impacto) na equação acima. Simplificando a equação, temos:

$$y_{ti} = \beta_i + \beta_1 before_i + \beta_3 after_i$$

Equação 3 – Modelo de Regressão Linear simplificada.

Em que  $\beta_i + \beta_1 before_i$  mede a correlação de fatores externos ao financiamento e substitui  $\beta_0 + \sum_c \beta_c x_{ci} + u_i + r_{ti}$ .

Rearranjando a equação é alterando os nomes das variáveis, no intuito de facilitar a compreensão da fórmula utilizada, a regressão linear multivariada tem a seguinte equação:

$$Indicador_{atual} = C + I * Indicador_{anterior} + P * Financiamento$$

Equação 4 - Regressão Linear entre índice atual, índice anterior e financiamento.

Na equação acima, *Indicador* representa a variável estudada, como por exemplo, orientação de doutorado. *C* é uma constante que reflete o estado inicial e também inclui o erro, *I* dá a correlação de fatores externos não temporais, entre o *Indicador<sub>anterior</sub>* e o *Indicador<sub>atual</sub>* e *P* mede a correlação do financiamento com o *Indicador<sub>atual</sub>*.

Como *Indicador* é, neste estudo, o número de mestres e doutores orientados, tomando o número de doutores orientados como indicador de exemplo, *Doutores<sub>2014</sub>*

é o número de doutores formados pelo pesquisador até o final do ano de 2014 e  $Doutores_{2019}$  esse número até o final de 2019.

$C$  representa uma constante e estima o contrafactual quando se supõe que não houve fatores externos ou temporais. Dessa forma, estimar  $C$  representaria estimar a média de doutores formados no ano de 2014, ou seja, o estado inicial do indicador, mais o erro.

$I$  representa a taxa constante de progressão no tempo independentemente do financiamento. Ou seja, representa a tendência que se tem de crescimento espontâneo do número de doutores formados pelos pesquisadores, ou a medida de efeitos externos. Assim,  $I * Indicador_{anterior}$  estima os efeitos externos e temporais no contrafactual.

$P$  mede o efeito do financiamento, ou seja, a correlação entre o valor investido e o número de doutores formados. Como o impacto é a medida da diferença entre os estados dos grupos de tratamento e controle,  $P * Financiamento$  nos dá justamente a estimativa do impacto da política pública no indicador em questão.

### 3 Currículo Lattes

Para a realização deste estudo, precisou-se extrair informações sobre a formação humana realizada de cada pesquisador. Esses foram retirados dos Currículos Lattes, que é um currículo informatizado, padronizado, público e mantido pelo governo brasileiro, mais especificamente pelo CNPq. Ele é de uso obrigatório, no Brasil, para todos aqueles que queiram concorrer a financiamentos públicos para uma pesquisa científica e se tornou um padrão no registro da vida dos estudantes de pós-graduação e pesquisadores do país. Além disso, ele é hoje adotado pela maioria das instituições de fomento, universidades e institutos de pesquisa do País, sendo adotado, inclusive, por instituições particulares que o usam como padrão para vagas de emprego, financiamentos ou outras vantagens. Ele é de acesso público, tanto para o cadastro quanto para a consulta e permite a consulta por área, formação, atuação profissional, palavras chaves, entre outros.

Ainda, por ter uma estrutura padronizada, permite que se realize mineração de dados para que os indicadores sejam automaticamente recuperados. Para que se tenha uma compreensão da dimensão de informações presentes no Currículo Lattes, o arquivo que representa sua estrutura possui 6.574 linhas, ou seja, são mais de

5.000 diferentes campos, padronizados, que podem ser preenchidos e recuperados. Observa-se que, no momento da escrita deste artigo, a Plataforma Lattes tem 7.427.277 currículos cadastrados, que após importados geram cerca de 1 Tb de dados.

Como vamos medir a mudança ao longo do tempo do grupo de controle e do grupo de tratamento, procuramos apenas aqueles campos que podem ser diretamente relacionados com o ano. Sob essas regras, extrai-se do Currículo Lattes 140 indicadores diferentes. Para que se entenda melhor, damos o exemplo de um doutorado orientado. São mais de 30 diferentes informações que podem ser inseridas e catalogadas, como nome do doutorando, instituição, título da tese, coorientadores etc. No entanto, cada doutorado orientado possui apenas um ano de titulação. Portanto, de todas as informações que se pode extrair de uma entrada dessas no Currículo Lattes, tira-se apenas um indicador, que foi nomeado de “Orientação de Doutorado”.

Ressalta-se que, sob estas regras, pode-se extrair 140 indicadores do Currículo Lattes. No entanto, apenas quatro indicadores foram utilizados inicialmente: Doutorados Orientados, Mestrados Orientados, Mestrados Profissionalizantes Orientados e Pós-Doutorados Orientados.

Pelo volume de informações que devem ser analisadas, este estudo se baseia no uso de BigData e de acordo com Guthrie et al (2013) pertencem ao que ele chama de ferramentas de análise do Grupo 2: bibliometria, mineração de dados e análise econométrica. O Grupo 2 constitui de Ferramentas Quantitativas e são normalmente escaláveis, transparentes, permitem comparações entre avaliações diferentes e são adequadas a usos frequentes e longitudinais. São normalmente livres de julgamento, mas a implementação da avaliação costuma ter alto custo inicial. Assim, de acordo com Guthrie et al (2013), o estudo pode ser facilmente expandido a outros indicadores ou outros financiamentos. Assim, como parte dos objetivos secundários, o estudo foi expandido a outros indicadores, revelando diversos indicadores que possuem correlação com o investimento realizado. Essa análise é apresentada nas considerações finais.

#### **4 Obtenção e tratamento dos dados**

Antes de se extrair quaisquer dados, foi necessário definir os anos de início e

de fim a serem considerados. Stern (2014) demonstra que a correlação entre o quantitativo de citações ao longo do tempo já possui forte correlação ao número de citações em anos anteriores a partir do terceiro ano de publicação (janela de dois anos após o ano de publicação do artigo), explicando cerca de 70% (69,2% para economia e 71,8% para ciência política) dos dados (STERN, 2014). A correlação aumenta significativamente nos anos posteriores, para cerca de 85% com uma janela de 3 anos, e cerca de 95% em janela de 4 anos. Como o Universal tem uma duração de dois anos, foi adotada a janela de seis anos para a medida do impacto. Ressalta-se, no entanto, que essa janela foi escolhida com base na correlação de um único indicador e que um maior estudo seria necessário para definir se essa está adequada.

No entanto, no ano de 2020 houve fato singular que pode representar um viés capaz de prejudicar os resultados encontrados, a pandemia COVID-19, muito embora a metodologia busque justamente eliminar o efeito de fatores externos como esse. Mesmo assim, para evitar que a pandemia interfira na aplicação do método, foi adotado como o tempo final do estudo o dia 01/01/2020, de forma que todos os Currículos Lattes que não foram atualizados dessa data a janeiro de 2022, data de aquisição dos currículos, foram desconsiderados.

Com uma janela de seis anos, o tempo inicial é o ano de 2014, assim, o Edital Universal 2014 foi escolhido como objeto de estudo. Observa-se que a escolha do ano final de 2019 (ou seja, todos os indicadores anteriores a 01/01/2020) como objeto de estudo também se deu para que se tenha um maior número de Currículos Lattes atualizados, uma vez que apenas esses podem ser usados. Assim, quanto mais recente é o estudo, menos currículos estão atualizados e menor é o poder estatístico.

Uma vez definido a chamada a ser estudada e os anos a serem considerados, passou-se à coleta de dados. Foram coletadas três origens de dados diferentes: O Currículo Lattes de todos os pesquisadores envolvidos, a Planilha de Demanda Bruta, que lista todos os pesquisadores que concorreram à chamada estudada e a listagem de todos os pagamentos realizados pelo órgão.

Os Currículos Lattes são publicamente acessíveis. No entanto, ferramentas de proteção impedem a extração automática por meio de *web scraping* (programa de computador que baixa os currículos automaticamente). Dessa forma, os currículos foram obtidos por meio da ferramenta Extrator Lattes, disponibilizado pelo CNPq (EXTRAÇÃO DE DADOS - PORTAL MEMÓRIA, [s. d.]). Para a extração, foi necessária

a criação de um programa em Python que verificou os dados necessários para a extração, realizou a extração e a coleta dos indicadores, em cada um dos currículos baixados.

A Planilha de Demanda Bruta foi retirada por meio de um sistema interno ao CNPq de relatórios. No entanto, a planilha também pode ser obtida por meio de solicitação ao canal da Lei de Acesso à Informação. Nela, constam o nome completo de cada pesquisador que concorreu ao Universal 2014, bem como diversos outros dados relativos à proposta submetida, incluindo-se o número do processo, realizado para identificar se houve ou não pagamentos em relação ao processo submetido.

A listagem de todos os pagamentos realizados está publicamente disponível em [http://dadosabertos.cnpq.br/pt\\_BR](http://dadosabertos.cnpq.br/pt_BR). A planilha contempla o número do projeto, o valor pago e o ano de pagamento, além de outras informações. Dessa forma, liga-se o pagamento ao pesquisador pelo número do projeto, e o pesquisador a seu Lattes por meio do nome completo. Finalmente, para evitar a possibilidade de homônimos, foi realizada verificação por meio de outra ferramenta internamente disponível de relatórios, que liga por meio de chave única o nome do pesquisador a seu Lattes. Dessa forma, garantiu-se que não houve casos de homônimos no estudo realizado.

Para a realização da regressão e uma vez que se adotou o limite imposto pela pandemia como aquele de observação final dos dados, dos 7.427.277 currículos cadastrados na Plataforma Lattes, foram importados 2.651.713 que são aqueles atualizados a partir de 01/01/2020 (a listagem pode ser obtida no portal do Extrator Lattes). Observa-se que a importação e o tratamento desses currículos geraram mais de 700 gigabytes de dados.

Todos os 2.651.713 foram minerados em busca dos dados, para que fossem retirados os dados básicos que permitissem sua correta identificação, bem como seus respectivos indicadores. A importação foi necessária uma vez que a planilha fornecida pelo órgão não possui o número identificador do currículo, mas apenas o nome do pesquisador. Como resultado, foi construída tabela contendo 253.506.545 de linhas, cada linha indicando o ano em que foi registrado cada um dos 140 diferentes indicadores de cada currículo.

No Edital Universal 2014 houve 16.902 pesquisadores participantes, sendo que desses 15.907 tinham o currículo atualizado em 01/01/2020 e apenas 5.222 tiveram financiamento realizado. Todos os 15.907 currículos foram convertidos nos 140

diversos indicadores, calculados tanto para o ano de 2014 quanto para o final do ano de 2019, gerando, portanto, uma tabela de cinco colunas: identificador do pesquisador, indicador, pagamento, indicador em 2014 e indicador em 2019.

O resultado foi novamente trabalho em outra tabela que somou os indicadores anteriores aos anos 2014 e 2019, de modo a possibilitar a comparação no tempo, e agrupados em faixa e área de pesquisa, resultando em tabela com 3.028.200 linhas com indicadores pertencentes a 15.907 pesquisadores.

O Universal 2014 é separado em três distintas faixas de concorrência, a depender do valor solicitado: Faixa A, Faixa B e Faixa C. Dessa forma, pesquisadores que submeteram propostas para a Faixa A não concorrem com aqueles que submeteram à Faixa B, e assim por diante. Assim, os dados foram agrupados por essas faixas e também pelas 80 áreas de pesquisa em que foi permitida a concorrência, resultando em 33.600 tabelas (3 modalidades x 80 áreas x 140 indicadores).

## 5 Resultados e discussões

Uma vez levantados os dados e definidas as 33.600 tabelas, aplicou-se, então, em cada uma dessas tabelas, a regressão linear da Equação 4:

$$\text{Indicador}_{2020} = C + I * \text{Indicador}_{2014} + P * \text{Financiamento}$$

Uma das vantagens de usar a regressão multivariada é que a estatística nos dá a medida da hipótese nula “ $P > |t|$ ”, ou seja, a probabilidade estatística de que, considerados fatores aleatórios e não observados, considere-se que o indicador resultou num impacto quando isso não é verdadeiro. Em outras palavras, o teste da hipótese nula nos dá a probabilidade de que consideremos um impacto que, na verdade, não foi causado pelo financiamento e sim por outros fatores externos e/ou aleatórios. Com o teste da hipótese nula “ $P > |t|$ ”, pode-se realizar um estudo exploratório em todos os indicadores e daí se retirar quais foram aqueles com alta probabilidade estatística de representar um impacto real, além da medida desse impacto. Para facilitar a compreensão, adotaremos o conceito de Grau de Confiança, definido como  $1 - “P > |t|”$  e expresso em porcentagem. Assim, o Grau de Confiança nos daria as chances da estatística ser relevante ou real.

Neste estudo, foram listados apenas os impactos com Grau de Confiança de 99% (em que o teste da hipótese nula “ $P > |t|$ ” é inferior a 0,01). Simplificadamente,

tem-se assim 99% de certeza de que o impacto medido é real. Também, tem-se alta chance probabilística de que o impacto medido está subvalorizado, uma vez que muitos impactos reais não atingirão o nível de confiança proposto, isso porque, nos indicadores que não satisfizerem o grau de confiança de 99%, teremos apenas 1% de certeza de que não houve impacto nessa área, com 99% de erro. Ou seja, temos a certeza de que estamos subestimando o impacto encontrado, mas, ao mesmo tempo, estamos garantindo uma alta probabilidade estatística de que esse impacto é real. A medida foi necessária para que se reduza a possibilidade de se medir impacto não condizente com a realidade, pois o impacto é indireto e resultante de milhares de regressões.

Assim, chega-se ao impacto do programa, aqui definido e calculado como  $P * \text{Financiamento}$ . Ou seja, para se chegar no impacto, cada parâmetro  $P$  considerado significativo (99% de Grau de Confiança) foi multiplicado pelo valor total financiado naquela faixa, programa e indicador, para se chegar ao Impacto Médio. O resultado foi novamente filtrado de modo a eliminar os impactos menores que uma unidade, uma vez que impactos muito pequenos também podem ser considerados como nulos. Finalmente, todos os impactos foram agrupados por indicador e somados, chegando-se ao impacto final da Política Pública:

Tabela 1 – Impacto da Chamada Universal 2014 em Formação Humana a nível de Pós Graduação

<b>Indicador</b>	<b>Impacto</b>
Orientação Concluída de Doutorado	604
Orientação Concluída de Mestrado	294
Orientação Concluída de Pós-Doutorado	193
Orientação em Andamento de Doutorado	91
Orientação em Andamento de Pós-Doutorado	9
Orientação em Andamento de Mestrado	-27

Fonte: Produção própria.

Chama atenção a existência de impacto negativo. No entanto, pode-se interpretar o impacto negativo no número de orientações em andamento de mestrado como, na verdade, um impacto positivo na medida em que houve migração de orientações de mestrado para o doutorado. Ou seja, enquanto o número de orientações em andamento de mestrado cai, o número de orientações em andamento de doutorado sobe significativamente, indicando que o financiamento levou os pesquisadores a darem preferência a uma formação de maior nível.

Assim, o impacto indireto do investimento realizado na educação é claro. De acordo com os dados abertos do CNPq, foi financiado nessa política pública o valor total de R\$ 194.101.318,26. Com esse investimento, notamos impacto indireto na formação de 294 mestres, 604 doutores e 193 pós-doutores. Reforça-se que esse é apenas uma pequena parcela do impacto total do financiamento, que objetivou uma pesquisa científica e não a formação de recursos humanos e que ainda teremos um aumento de 91 doutores (se todos conseguirem completar seus doutorados).

A título de comparação, o CNPq financiava diretamente a formação de recursos humanos, na forma de bolsas para Mestrado e Doutorado<sup>5</sup>. Cursos de mestrado e doutorado são essenciais para assegurar a oferta de mão de obra qualificada, tanto para a indústria, quanto para as universidades, afetando o crescimento socioeconômico do estado (OLIVEIRA; AMARAL, 2017) e faz parte do programa educativo do Brasil.

O valor do subsídio para bolsas de mestrado é de R\$ 1.350,00, sendo a duração padrão para o financiamento de 48 meses. Assim, cada mestre formado é subsidiado em aproximadamente R\$ 32.400,00. Cada doutor recebe bolsa no valor de R\$ 2.000,00 e o subsídio tem duração padrão de 48 meses, de modo que para cada doutor formado é subsidiado o valor de R\$ 96.000,00. Já o Pós-Doutor recebia bolsa de R\$ 4.100,00 por dois anos. Assim, o valor financiado por Pós-Doutor era de R\$ 98.400,00.

No Universal 2014, temos um impacto de 267 mestres, 695 doutores e 202 pós-doutores (somando-se aqueles concluídos com os ainda em andamento). Para se obter o mesmo número de mestres, doutores e pós-doutores com bolsas, seria necessário um investimento total de R\$ 95.247.600,00 reais, aproximadamente 49% do valor financiado. Ou seja, o impacto indireto na educação foi de no mínimo o equivalente a 49% do valor financiado. Vale lembrar que esse valor ainda está subestimado, uma vez em que há Grau de Confiança de 99% em sua existência e que esse é apenas o impacto indireto.

A formação humana de mestres e doutores se relaciona diretamente com a qualidade da educação. Morais et al. (2022) estudaram o Programa de Pós-Graduação em Ensino de Ciências Naturais e Matemática da Universidade Federal

---

<sup>5</sup> [http://www.cnpq.br/web/guest/view/-/journal\\_content/56\\_INSTANCE\\_0oED/10157/25366](http://www.cnpq.br/web/guest/view/-/journal_content/56_INSTANCE_0oED/10157/25366)

do Rio Grande do Norte (PPGECNM-UFRN) e indicaram o impacto do mestrado profissionalizante na educação, mostrando que materiais didáticos desenvolvidos são efetivamente usados em sala de aula em 70% dos casos, que em 55% dos casos os produtos educacionais foram publicados, que 71% propuseram atividades inovadoras, que 60% tiveram melhoria salarial, mantendo o emprego anterior em 70% dos casos. Mesmo em programas de outras áreas, espera-se efeito semelhante em que o conhecimento adquirido pelos mestres e doutores contribui para a melhoria da qualidade da educação no Brasil.

## 6 Considerações finais

A mesma metodologia foi aplicada a todos os indicadores passíveis de extração do Currículo Lattes, chegando-se a um possível impacto do Universal 2014 passível de ser medido pelo Currículo Lattes. Observa-se que, dos 140 diferentes indicadores, 26 representam o fim ou conclusão de uma atividade, com 26 respectivos indicadores que representam o início dessas atividades e que são, portanto, altamente correlacionados aos primeiros. Assim, os 11 indicadores que representam o fim de uma atividade foram agrupados com os 11 indicadores representando seu início para resultar em novas atividades, eliminando a alta correlação. Por exemplo, “Fim de Vínculos” foi subtraído de “Início de Vínculos” resultando no indicador “Novos Vínculos”.

Além disso, os 15 indicadores que representavam o início de uma atividade e que se correlacionam fortemente com seus respectivos pares que representam a conclusão dessa mesma atividade foram ignorados. Por exemplo, “Conclusão de Doutorado” é um indicador presente no estudo, mas “Início de Doutorado” foi retirado do mesmo. Restaram, portanto, 114 indicadores:

Tabela 2 – Impacto da Chamada Universal

<b>Indicador</b>	<b>Impacto</b>
Artigo Publicado	6071
Novos Vínculos Empregatícios ou Funcionais	1236
Participação em Banca de Doutorado	1119
Participação em Banca de Mestrado	899
Orientação Concluída de Doutorado	604
Trabalhos em Eventos	562
Produção de Trabalho Técnico	462
Desenvolvimento de Material Didático ou Instrucional	413
Publicação de Capítulo de Livro	321

Prêmio ou Título Recebido	316
Apresentação de Trabalho Técnico	309
Orientação Concluída de Mestrado	294
Participação em Congresso	202
Orientação Concluída de Pós-Doutorado	193
Participação em Banca de Exame de Qualificação	162
Texto em Jornal ou Revista	103
Produção de Software	96
Orientação em Andamento de Doutorado	91
Livro Publicado ou Organizado	78
Participação em Simpósio	73
Outras Participações de Bancas Julgadoras	71
Produção Bibliográfica	55
Outras Participações em Congressos	49
Patente Registrada	44
Editoração	44
Organização de Evento	38
Participação em Encontro	36
Participação em Oficina	32
Ministério de Curso de Curta Duração	32
Novas Extensões Universitárias	29
Conclusão de Extensão Universitária	26
Produção de Artes Cênicas	25
Outra produção Técnica	25
Publicação de Prefácio ou Posfácio	25
Participação em Exposição	21
Programa de Rádio ou TV	21
Participação em Banca Julgadora para Professor Titular	20
Treinamento Ministrado	14
Produção Técnica - Processos ou Técnicas	11
Produção Artística ou Cultural	11
Participação em Seminário	11
Orientação em Andamento de Pós-Doutorado	9
Tradução	8
Orientação em Andamento de Aperfeiçoamento ou Especialização	7
Novas Atividades Técnico/Científicas	6
Produção de Carta, Mapa ou Similar	4
Conclusão de Outra Formação Complementar	3
Participação em Banca Julgadora para Concurso Público	3
Produção de Artes Visuais	3
Produção - Demais Trabalhos	3
Novas Atividades de Conselho, Comissão ou Consultoria	1
Participação em Banca Julgadora para Avaliação de Cursos	1
Desenvolvimento de Desenho Industrial	-2
Conclusão de Especialização	-2

Mídia Social, website ou Blog	-3
Produção Técnica - Produto Tecnológico	-3
Participação em Banca Julgadora para Livre Docência	-5
Novos Estágios	-7
Novas Atividades de Pesquisa e Desenvolvimento	-8
Orientação em Andamento de Graduação	-10
Orientação em Andamento de Iniciação Científica	-11
Artigo Aceito para Publicação	-18
Novas Atividades de Ensino	-23
Participação em Direção ou Administração	-23
Orientação em Andamento de Mestrado	-27
Outras Orientações em Andamento	-28
Novas Atividades de Serviço Técnico-Especializado	-53
Conclusão de Pós-Doutorado	-54
Conclusão de Aperfeiçoamento	-67
Conclusão de Curso de Curta Duração	-82
Novos Projetos de Pesquisa	-155
Outras Orientações Concluídas	-331
Participação em Banca de Graduação	-716

Fonte: Produção própria.

A primeira coisa que se observa é que o impacto pode ser muito maior que o esperado. Para se ter uma compreensão de sua dimensão, somente na área da educação ainda seria necessário examinar, por exemplo, os 1.236 novos vínculos, em sua maioria em atividades de professorado, 413 materiais didáticos produzidos, cada um com seu valor intrínseco que poderia ser calculado, além da enorme quantidade de materiais, inclusive editoriais, produzidos e sua relação com a educação. O impacto indireto na educação pode facilmente ultrapassar o próprio valor do financiamento e esse ainda se somaria ao impacto direto, na forma de produção de conhecimento e que pode ser verificado na forma de 6071 artigos publicados.

Uma análise rápida também pode indicar que a política pública de financiamento à pesquisa levou os pesquisadores a se concentrar em atividades mais especializadas e menos em atividades mais rápidas ou de menor “nível”. Um exemplo é o impacto negativo de 716 participações em banca de graduação, com o aumento de 1119 participações em banca de doutorado.

Um indicador que pode ser analisado com especial atenção é o impacto negativo em atividades de direção e administração, que representa um importante eixo da Tripla Hélice. Embora o impacto negativo tenha sido, na verdade, pequeno (23 pesquisadores, dos 13.104 financiados, teriam deixado alguma atividade de

direção e administração), mostra-se que cabe estudo para verificar se ele é real e seu motivo.

Precisa-se lembrar que, como foram realizadas 33.600 regressões lineares, não é possível afirmar que o impacto representado na Tabela 2 está correto. Isso porque, assim como o impacto possui altas chances de estar seriamente subestimado, há uma alta chance estatística de ter sido listado como impacto algum erro estatístico, uma vez que o Grau de Confiança é de 99%. Assim, reforçamos a necessidade de um estudo detalhado para confirmar a veracidade do impacto listado antes de qualquer decisão a ser tomada com base nos achados.

Apesar disso, podemos inferir que o impacto do investimento realizado é alto, superando em muitas ordens o financiamento realizado, devendo, portanto, ser uma Política Pública não só a ser mantida, mas incentivada e expandida.

### Agradecimentos

Em especial agradecimento aos professores Clovis Zapata e Livio Amaral, sem os quais nunca teria realizado este estudo.

### Referências

BORNMANN, L. What is societal impact of research and how can it be assessed? A literature survey. *JOUR. Journal of the American Society for Information Science and Technology*, [s. l.], v. 64, n. 2, p. 217–233, 2013.

CABRERA, M. R.; ARAÚJO, M. S. T. de. Análise das políticas educacionais na Pós-graduação Stricto-Sensu no âmbito da formação continuada de professores. *Revista de Ensino de Ciências e Matemática*, [s. l.], v. 9, n. 3, 2018.

CAPEL. **Competências**. [S. l.], 2012. Disponível em: <http://www.capes.gov.br/acessoainformacao/80-conteudo-estatico/acesso-a-informacao/5418-competencias>. Acesso: 8 Aug. 2022.

COMPETÊNCIAS — PORTUGUÊS (BRASIL). [S. l.], [s. d.]. Disponível em: <https://www.gov.br/cnpq/pt-br/acesso-a-informacao/institucional/competencias>. Acesso: 22 Jun. 2022.

EXTRAÇÃO DE DADOS - PORTAL MEMÓRIA. [S. l.], [s. d.]. Disponível em: <http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados>. Acesso: 5 Jul. 2022.

GERTLER, P. J. *et al.* **Avaliação de impacto na prática**. Segunda Eded. [S. l.]: World Bank Publications, 2018.

MILESI, C. *et al.* Charting the Impact of Federal Spending for Education Research: A Bibliometric Approach. *Educational researcher*, Los Angeles, CA, v. 43, n. 7, p. 361–370, 2014.

MORAIS, E. *et al.* O Mestrado Profissional em Ensino de Ciências Naturais e Matemática da UFRN: uma análise das dissertações e acompanhamento de egressos de 2012 a 2020. **Revista de Ensino de Ciências e Matemática**, [s. l.], v. 13, n. 3, 2022.

OLIVEIRA, T.; AMARAL, L. Public Policies in Science and Technology in Brazil: challenges and proposals for the use of indicators in evaluation. **Mugnaini, R**, [s. l.], p. 189–217, 2017.

WILBERTZ, J. Evaluating the societal relevance of research. [s. l.], n. april, 2013.

### **Publicação Bilíngue**

X Manifestamos a disponibilidade em realizar a publicação bilíngue. Estamos cientes que a tradução poderá ser feita por equipe indicada pela *REnCiMa* ou por nós contratada. Estamos cientes, também, que os custos serão de nossa responsabilidade, autores do artigo submetido.