




<https://doi.org/10.18222/eae.v35.10777>

INDICADORES QUALITATIVOS NA AVALIAÇÃO DA PÓS-GRADUAÇÃO EM EDUCAÇÃO

 ROBERT EVAN VERHINE^I
 LYS MARIA VINHAES DANTAS^{II}
 ÂNGELO RICARDO DE SOUZA^{III}

^I Universidade Federal da Bahia (UFBA), Salvador-BA, Brasil; rverhine@gmail.com

^{II} Universidade Federal do Recôncavo da Bahia (UFRB), Cachoeira-BA, Brasil; lys@ufrb.edu.br

^{III} Universidade Federal do Paraná (UFPR), Curitiba-PR, Brasil; angelosou@gmail.com

RESUMO

Analisa-se a avaliação da pós-graduação na Área de Educação da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, entre 2017 e 2020, destacando as mudanças decorrentes da ampliação do uso de indicadores qualitativos. Após a contextualização da avaliação e a discussão de aspectos teóricos sobre o tema, descreve-se o processo adotado para a objetivação da avaliação e os resultados obtidos, por análise de regressão, do comportamento dos indicadores quanti e qualitativos na definição das notas dos programas da área. Propõem-se redução do número de indicadores, valorização de indicadores qualitativos como fontes para o *feedback* necessário à melhoria do programa e adoção de processos de objetivação derivada da construção de critérios avaliativos que parametrizem o trabalho da avaliação.

PALAVRAS-CHAVE AVALIAÇÃO DA EDUCAÇÃO • PÓS-GRADUAÇÃO • POLÍTICAS EDUCACIONAIS • AVALIAÇÃO EM LARGA ESCALA.

COMO CITAR:

Verhine, R. E., Dantas, L. M. V., & Souza, A. R. de. (2024). Indicadores qualitativos na avaliação da pós-graduação em educação. *Estudos em Avaliação Educacional*, 35, Artigo e10777. <https://doi.org/10.18222/eae.v35.10777>

INDICADORES CUALITATIVOS EN LA EVALUACIÓN DE POSGRADO EN EDUCACIÓN

RESUMEN

En este artículo se analiza la evaluación del posgrado en el Área de Educación de la Coordenação de Aperfeiçoamento de Pessoal de Nível Superior [Coordinación de Mejora de Personal de Nivel Superior] entre 2017 y 2020, destacando los cambios resultantes de la ampliación del uso de indicadores cualitativos. Luego de la contextualización de la evaluación y la discusión de los aspectos teóricos sobre el tema, se describe el proceso adoptado para la objetivación de la evaluación y los resultados obtenidos mediante análisis de regresión del comportamiento de los indicadores cuantitativos y cualitativos en la definición de las notas de los programas del área. Se propone la reducción del número de indicadores, la valorización de los indicadores cualitativos como fuentes para el *feedback* necesario para la mejoría del programa, y la adopción de procesos de objetivación derivados de la construcción de criterios evaluativos que parametren el trabajo de la evaluación.

PALAVRAS CLAVE EVALUACIÓN DE LA EDUCACIÓN • POSTGRADO • POLÍTICAS EDUCATIVAS • EVALUACIÓN A GRAN ESCALA.

QUALITATIVE INDICATORS IN THE EVALUATION OF GRADUATE PROGRAMS IN EDUCATION

ABSTRACT

This article analyzes the evaluation of graduate studies in the Education Area by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior [Brazilian Federal Agency for Support and Evaluation of Graduate Education] between 2017-2020 highlighting the changes resulting from the expanded use of qualitative indicators. Following a discussion on the evaluation and theoretical aspects of the subject, the process adopted for the objectification of the evaluation and the results obtained is described. Regression analysis is used to analyze the behavior of both quantitative and qualitative indicators in defining the grades of the programs in the area. The article proposes a reduction in the total number of indicators, emphasizes the importance of qualitative indicators as valuable sources of the feedback needed to improve the program, and advocates for the adoption of objectification processes derived from the construction of evaluative criteria that parameterize the evaluation work.

KEYWORDS EDUCATIONAL EVALUATION • POSTGRADUATES • EDUCATIONAL POLICIES • LARGE-SCALE EVALUATION.

Recebido em: 6 DEZEMBRO 2023

Aprovado para publicação em: 27 FEVEREIRO 2024



Este é um artigo de acesso aberto distribuído nos termos da licença Creative Commons do tipo BY-NC.

INTRODUÇÃO

No final de 2018, duas importantes decisões foram tomadas pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) em relação à avaliação de programas de pós-graduação e ao ciclo 2017-2020. Em outubro, o Conselho Superior da Capes acatou um documento, preparado pela Comissão para o Acompanhamento do Plano Nacional de Pós-Graduação, que incluía, entre suas várias recomendações, o “balanço entre indicadores quantitativos e qualitativos”. Em dezembro do mesmo ano, o Conselho Técnico-Científico (CTC-ES) da Capes aprovou a reestruturação da Ficha de Avaliação que, de maneira significativa, aumentava o papel dos indicadores qualitativos no processo geral de avaliação dos programas. Pela primeira vez, os insumos fornecidos pelos programas, que tradicionalmente eram tratados qualitativamente apenas para sua contextualização, passaram a receber peso na avaliação. Outros fatores, também de natureza qualitativa e que lidavam com os processos formativos e os impactos do programa, receberam maior importância. Por outro lado, os resultados quantitativos dos programas, vinculados a professores e a discentes, tiveram seu peso diminuído. Para a Capes, a ênfase se volta para o objetivo-chave da pós-graduação: a formação discente. Entende-se que essa formação demanda envolvimento na produção acadêmica (na forma de publicações e de produtos tecnológicos, por exemplo), mas vai além de tal produção como um objetivo programático central.

Como consequência dessas decisões, cada uma das 49 áreas de avaliação da Capes se dedicou a customizar a Ficha de Avaliação, especificando os indicadores para os 12 itens preestabelecidos, organizados em três quesitos. Com essa alteração, os indicadores qualitativos, que requerem interpretação subjetiva, se tornaram mais prevalentes que no passado, para todas as áreas. Em algumas delas, os indicadores quantitativos permaneceram predominantes, mas todas as áreas se moveram na direção de uma avaliação que valorizava aspectos menos imediatamente tangíveis, substituindo, portanto, algo mais facilmente mensurável por um processo complexo, exigindo um esforço na definição de critérios que diminuíssem a subjetividade inerente às avaliações qualitativas.

Tal movimento foi especialmente evidente na Área de Educação, na qual o percentual de indicadores qualitativos adotados subiu de 31,2%, em 2017, para 68,1%, em 2021. Ao elevar drasticamente esse percentual, a Área não só acompanhou a tendência geral da Capes, mas também respondeu a uma demanda de sua comunidade acadêmica, apresentada e constantemente reforçada desde o estabelecimento do sistema de avaliação, em 1980. Inúmeras publicações, como Gatti et al. (2003), Moreira et al. (2004) e Sguissardi (2006), argumentavam que o aumento no uso de indicadores qualitativos favoreceria uma avaliação que entendesse a qualidade do programa de forma mais holística e contextualizada, considerando, de maneira

apropriada, a complexidade dos processos formativos. A confiança excessiva nos indicadores quantitativos observada no modelo de avaliação da Capes também foi criticada¹ por um grupo de estrangeiros que avaliaram o sistema no início deste século (Spagnolo & Calhau, 2002) e, mais recentemente, pelos membros do Fórum de Coordenadores de Programas de Pós-Graduação em Educação, em documento elaborado em Fortaleza (Forpred, 2017).

Apesar disso, por muitos anos a demanda por aumento do uso de indicadores qualitativos não foi ouvida. A resistência à mudança pode ser atrelada à natureza do modelo de avaliação da Capes. Originalmente implementado em 1980, o modelo havia sido formulado para permitir a distribuição de cotas de bolsas aos programas de pós-graduação de modo a favorecer programas com níveis de qualidade mais altos (Castro, 2006). A lógica era que programas com maior qualidade utilizariam as bolsas de maneira mais efetiva, e que a vinculação dos resultados da avaliação com o financiamento federal conduziria ao desenvolvimento dos programas, de modo geral. De acordo com esse raciocínio, era necessário diferenciar os níveis de qualidade dos programas, o que significou, desde o início, a adoção de uma escala de qualidade na qual os programas seriam comparativamente alocados.

Inicialmente, a escala hierárquica adotada pela Capes era composta de cinco níveis, rotulados como A, B, C, D e E, do mais alto para o mais baixo. Em 1998, a escala foi expandida para sete níveis, que passam a ser rotulados de 1 a 7, de maneira que os dois níveis superiores (6 e 7) seriam altamente exclusivos, reservados para os programas considerados excelentes por padrões internacionais (Capes, 2021). Ao CTC-ES foi atribuída a responsabilidade da preservação da integridade da escala, assegurando que as diferenças entre os níveis fossem consistentes e comparáveis, tanto dentro de cada área de avaliação como entre as diferentes áreas. Considerando esse contexto e o fato de a avaliação da Capes ocorrer em larga escala e estar associada ao financiamento e à regulação, a avaliação era considerada de alto impacto (*high stakes*), de forma que o uso de indicadores quantitativos (ou numéricos) foi percebido como a estratégia apropriada, consistente com uma vasta literatura que divulgava a quantificação como um mecanismo que assegurava que os resultados da avaliação alcançariam padrões desejáveis de objetividade, reprodutibilidade, transparência, economia de tempo, além de permitir a adoção de abordagens analíticas sofisticadas (Biagioli, 2018; Khandker et al., 2009; Todeschini & Baccini, 2016). Em um clássico tratado sobre a busca de objetividade na ciência e na vida pública, Porter (1995) argumenta que a quantificação, nas ciências, é uma “estratégia para

1 Esses mesmos autores apresentaram outras críticas ao modelo de avaliação adotado pela Capes, que não foram abordadas neste artigo. Para um resumo de tais críticas, inclusive aquelas do Ministério Público Federal, que resultaram na paralisação de 2021 por ordem judicial, ver Brasil (2023).

superar a distância e a desconfiança” e que é “crucial para transformar habilidades experimentais locais em conhecimento público” (p. 47, tradução nossa).

Por outro lado, a literatura em defesa de uma abordagem qualitativa para a avaliação, mesmo quando em larga escala e *high stakes*, é também ampla. De acordo com essa literatura (Patton, 1999; Richie & Lewis, 2003; Usaid, 2005), os indicadores qualitativos envolvem percepções e julgamentos, além de organizar os dados de acordo com temas e conceitos em lugar de fazê-lo com base em enumeração, sendo, dessa forma, capazes de capturar um cenário mais abrangente e matizado da realidade do que sua contraparte quantitativa. Contudo, como enfatizado por TolaData (2021), ao desenvolver tais indicadores, é importante garantir que eles sejam “realistas, inteligíveis, inequívocos, comparáveis, fáceis para monitorar, registrar e reportar, que produzam dados de alta qualidade e que sejam contexto-específicos” (p. 1, tradução nossa). Esses requisitos são pertinentes no contexto da avaliação da Capes visto que, como já mencionado, os resultados têm implicações na regulação e no financiamento e devem ser suficientemente claros e precisos para orientar decisões governamentais e dar *feedback* aos programas, para o aprimoramento da qualidade.

Uma vez que o quadriênio avaliativo 2017-2020 se encerrou, tendo sido o primeiro com ênfase nos dados qualitativos, o propósito deste artigo é relatar a experiência da Área de Educação e seus esforços de objetivação dos 32 indicadores qualitativos utilizados para determinar as notas obtidas pelos programas. As decisões da Área para operacionalizar a análise dos referidos indicadores foram baseadas em fundamentos teóricos. O sucesso de tal esforço, por outro lado, pode ser avaliado, pelo menos em parte, pela análise quantitativa dos dados produzidos, examinando hipóteses que buscam captar o comportamento dos indicadores em relação às notas atribuídas aos programas. Assim, na próxima seção são apresentados e discutidos aspectos teóricos da avaliação em larga escala e a adoção de indicadores qualitativos. Além disso, é descrito o processo adotado pela área para sua objetivação. Na sequência, são brevemente apresentados os resultados estatísticos que oferecem um *insight* sobre a utilidade da empreitada. Por fim, nas conclusões, são endereçadas as implicações da experiência, não apenas em termos do futuro da avaliação da Capes, mas especificamente a avaliação do próximo quadriênio (2025-2028).

INDICADORES QUALITATIVOS EM AVALIAÇÕES EM LARGA ESCALA: SUA OBJETIVAÇÃO NO CENÁRIO *HIGH STAKES* DA CAPES

Considerações teóricas

O debate sobre abordagens quantitativas e qualitativas nas avaliações é antigo e, assim como nas discussões sobre métodos em pesquisa científica, tende a

reforçar uma dicotomia inexistente. Ao longo do tempo, tem sido possível a adoção de abordagens mais objetivas e restritas, com construção/aplicação de instrumentos padronizados e análises estatísticas, combinadas com outras mais subjetivas, que incluem, por exemplo, visitas técnicas e revisões por pares (Stufflebeam & Shinkfield, 2007). O fundamental é aplicar o método mais adequado para responder ao que foi perguntado (Lamont, 2012; Worthen et al., 2004) e para propiciar, de maneira justa e ética, o julgamento de valor que é parte de qualquer processo avaliativo. No caso da Capes, seu modelo de avaliação é híbrido, combinando análises estatísticas com análises mais qualitativas de informações que são fornecidas pelos programas sob avaliação, em relatórios anuais.

Em um primeiro momento, é fundamental diferenciar as avaliações internas, para fins específicos (como no caso da sala de aula ou de uma ação em uma instituição), de avaliações externas em larga escala, implementadas como políticas governamentais, muitas vezes compulsórias, cujos objetivos dizem respeito à avaliação de outras políticas públicas. Vale registrar que, para o ciclo 2017-2020 da Capes, a Área de Educação precisou avaliar 188 programas, o que demandou 60 avaliadores especialistas *ad hoc*, além de outros integrantes das comissões constituídas ao longo do quadriênio. Considerando que a Capes contava, no ciclo 2017-2020, com 49 áreas avaliando um total de mais de 4.500 programas, sua avaliação é nitidamente realizada em grande escala.

As avaliações externas, incluindo aquelas implementadas em larga escala como política pública, são uma ferramenta de gestão cuja finalidade maior é promover a qualidade do objeto avaliado ao favorecer a comparação entre ele e os padrões de qualidade definidos, permitindo um julgamento de valor sobre seu atingimento ou não (Minayo, 2009). Vista por esse prisma, toda avaliação é qualitativa; os dados coletados é que têm natureza qualitativa ou quantitativa, com o uso explícito (ou não) de números diferenciando os dois enfoques (Barreto, 1993). Worthen et al. (2004) consideram que as avaliações precisam ter um recorte qualitativo, nem que seja apenas para a descrição do programa sob avaliação, de maneira que sem isso o processo avaliativo fica incompleto.

No cenário Capes, é objeto da avaliação a qualidade da pós-graduação, observando que, no Brasil, esses programas têm grande diversidade, considerando-se pontos de amadurecimento e a proposição da missão das instituições que oferecem os cursos, assim como contexto geográfico (capital ou interior, regiões brasileiras), contexto social e características identitárias de cada área de conhecimento em dado momento histórico. Muitas das críticas feitas ao processo de avaliação em larga escala dizem respeito à crença de que diferentes não devem ser avaliados a partir de uma base única, como se fosse possível comparar, pelos mesmos indicadores, laranjas e morangos. Contudo, o que está sendo avaliado é a qualidade da prestação

de um determinado serviço – no caso, de oferta de pós-graduação *stricto sensu* – dentro de um contexto geral de um sistema único sob regulação pelo governo federal (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior [Capes], 2022). É, portanto, possível e desejável que o governo federal preste contas de suas políticas para a sociedade brasileira e que as avaliações sejam ferramentas apropriadas, produzindo resultados que possam vir a ser corretamente interpretados.

Tal prestação de contas inclui a avaliação de elementos como “impacto na sociedade”, que pressupõe a coleta de atributos intangíveis a serem “captados indiretamente por meio de suas formas de manifestação” (Minayo, 2009, p. 85). Nesse caso, é necessário objetivar os níveis de avaliação para os indicadores qualitativos, não como *ranking* (valor hierarquizador), mas como *grading*, que permite a atribuição de um valor qualificativo ao que está sendo avaliado (Gomez, 2014). Com isso, a discriminação se dá na comparação entre o padrão de qualidade e o que está sendo avaliado, e não em uma comparação entre programas. Contudo, embora os critérios para avaliação dos indicadores qualitativos tenham sido adotados como *grading* (avaliação a critério), no sentido global, a Capes utiliza o modelo de avaliação como *ranking*.

A definição da matriz de referência é o primeiro passo no ciclo avaliativo, como ilustrado na Figura 1, e ela deve expressar o que aquela instituição governamental ou política pública entende como qualidade. Ou seja, assim como em pesquisa se operacionalizam os conceitos, nos processos avaliativos a operacionalização do que seja “qualidade” também deve ocorrer, traduzida na matriz de referência, nas dimensões a serem avaliadas, nos descritores que darão lastro para a criação de questões avaliativas (em casos de testes e provas, por exemplo) ou de indicadores (para uma avaliação que adote outros tipos de coleta de dados) (Lamont, 2012; Stufflebeam & Shinkfield, 2007).

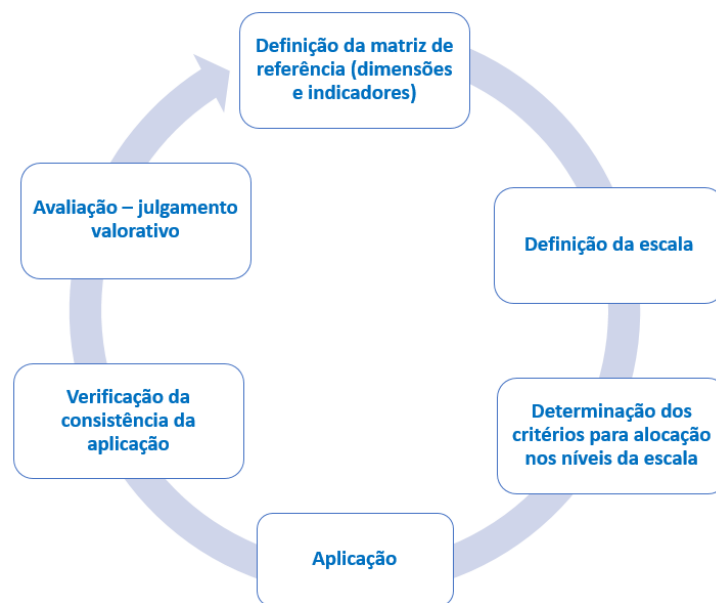
Sobre indicadores, Minayo (2009, pp. 84-85) considera que há um consenso sobre sua definição: “parâmetros quantificados ou qualitativos que servem para detalhar se os objetivos de uma proposta estão sendo bem conduzidos (avaliação de processo) ou foram alcançados (avaliação de resultados)”. A mesma autora argumenta que, para serem úteis, os indicadores precisam ter temporalidade, normalização, atenção à mesma especificação ou forma de medida, além de produção regular ao longo do tempo. A adoção de indicadores qualitativos e/ou quantitativos, portanto, é decorrente do que se quer avaliar e do como avaliar.

Nesse contexto, uma matriz é um instrumento de política pública (Secchi, 2014, p. 103)² e reflete o posicionamento de determinada corrente ideológica e go-

2 “[O]s instrumentos de política pública são meios disponíveis para a transformação das intenções contidas na política pública em ações para o enfrentamento do problema público”.

vernamental. A matriz também tem um aspecto educativo ao informar, de maneira sistematizada, sobre elementos considerados essenciais (embora não exclusivos ou restritivos) para a definição da qualidade do objeto avaliado. Por outro lado, restrições técnicas, financeiras, de recursos humanos, entre outras, resultam na redução da matriz àquilo que pode ser avaliado, e não ao escopo abrangente que deveria ser avaliado. Por isso, é fundamental que, quando da análise, interpretação e divulgação dos resultados, esses limites sejam considerados.

FIGURA 1
Ciclo de avaliação em larga escala



Fonte: Elaboração dos autores (2023).

No ciclo, as etapas das definições de escopo, escala e critérios de julgamento, como posto na Figura 1, são eminentemente políticas e demandam, idealmente, especialistas de orientações filosóficas diversas (Barreto, 1993). Se nos comitês avaliativos de proposição do escopo e dos critérios forem assegurados espaços para a participação dos avaliadores e avaliados, a matriz final, ainda que com limitações, contribuirá para uma avaliação que faça sentido aos avaliados. Em adição, essa definição *a priori* colaborará para diminuir vieses e personalismos durante a análise dos indicadores e a atribuição dos resultados nos níveis da escala.

Os critérios de julgamento devem ser definidos atentamente, observando que cada nível da escala inclui o que é esperado nos níveis inferiores (em vários casos, escalas são representadas como degraus em uma escada) e é caracterizado pelos elementos identitários próprios. Critérios são também chamados padrões, ou *standards*, a depender da literatura consultada, e são aplicados tanto nas abordagens

quanti como qualitativas. Há escalas que permitem alocação de valores, atitudes, sentimentos, interesses, dor, entre inúmeras possibilidades (Lamont, 2012).

Para ser confiável, o critério deve ser definido de modo a haver evidência clara do significado esperado pela entidade avaliadora (Reckase, 2023) e que seja compreensível para o avaliado. Para defini-lo, é necessária a etapa de *standard setting*, processo por meio do qual se determina o valor operacional consistente com o critério conceitual definido pela entidade avaliadora. Esse valor é usualmente um número que permite a alocação dos resultados na escala (Reckase, 2023). Ressalta-se, contudo, que esse número corresponde a determinado comportamento esperado. Por essa razão, o processo de construção da escala e do *standard setting* pressupõe que cada nível tenha uma descrição e que esteja justificado à luz do objetivo da avaliação e dos objetivos da política que está sendo avaliada.

Processos operacionais

Conforme a visão teórica apresentada, a definição de critérios e a posterior moderação dos avaliadores são fundamentais para minimizar o risco de assimetrias na avaliação, pelos mais diversos vieses. Fóruns consultivos e/ou deliberativos devem ser constituídos para proposição de: matrizes; escala; padrões (critérios) e seus pesos na escala; análise dos indicadores; verificação da consistência; e avaliação propriamente dita. Essas comissões devem ser formadas de maneira a assegurar múltiplas vozes desde o delineamento da matriz de referência até a aplicação de seus indicadores e a avaliação final.

No quadriênio 2017-2020, a Área de Educação da Capes se voltou para a definição dos critérios e a posterior moderação com os avaliadores, de modo a buscar a objetivação no processo avaliativo. A matriz de referência adotada pela agência já estava definida pela Portaria n. 122, de 5 de agosto de 2021, Art. 24, segundo a qual os resultados deveriam ser apresentados em uma escala de cinco níveis, chamados “conceitos”, a saber: muito bom (MB); bom (B); regular (R); fraco (F); insuficiente (I). Havia ainda a possibilidade de alocação do resultado como não aplicável (NA). Faltava, portanto, a definição dos critérios para cada nível dessa escala, considerando os 32 indicadores qualitativos da Ficha de Avaliação da Área. Assim, inicialmente foi constituída uma comissão para a proposição de critérios para os indicadores qualitativos, que foram a base do processo de moderação realizado posteriormente, que permitiu que a avaliação dos indicadores qualitativos ocorresse de maneira a assegurar comparabilidade e discriminação, de forma transparente.

Os critérios foram classificados como numéricos, implicitamente numéricos ou conceituais (Capes, 2022), e cada nível foi descrito de maneira objetiva e comparável, com atributos consistentes para cada descritor. Assim, embora o critério para cada nível da escala seja diferente dos demais, a mudança diz respeito à variação

da qualidade de determinado atributo fixo (clareza, grau de detalhamento, presença ou ausência de determinado aspecto, entre outros). Como critérios analíticos de comportamento, foram adotados elementos amplos, de modo que pudessem ser utilizados para avaliar o mesmo objeto em contextos diversos, como visto na literatura (Tierney & Simon, 2004).

Para exemplificar os critérios adotados, o indicador 1.1.5, Infraestrutura,³ foi analisado com base na quantidade de ocorrência do comportamento esperado (critério numérico). Esse indicador diz respeito à

. . . análise das condições estruturais adequadas para o alcance dos objetivos do programa, considerando disponibilidade de: a) salas de aula; b) laboratórios; c) amplo acesso à Internet; d) espaços multiusuários; e) biblioteca com acervo compatível com as necessidades do Programa; f) espaço próprio para a administração do programa (coordenação e secretaria); g) espaço próprio para grupos ou linhas de pesquisa; h) espaço próprio para os docentes e/ou para orientação dos discentes. (Capes, 2022, p. 33).

Sua avaliação foi objetivada nos cinco níveis da escala, conforme observado a seguir.

- Muito bom: há evidências claras de que todos os oito elementos relacionados estão disponíveis e são plenamente adequados para o funcionamento do programa.
- Bom: há evidências claras de que seis ou sete dos elementos relacionados estão disponíveis e são adequados para o funcionamento, com qualidade, do programa. Desses, é fundamental a adequação das salas de aula e o amplo acesso à internet.
- Regular: há evidências claras de que quatro ou cinco dos elementos relacionados estão disponíveis e são adequados para o funcionamento, com qualidade, do programa, desde que, entre eles, haja adequação das salas de aula e o amplo acesso à internet.
- Fraco: há evidências claras de que dois ou três dos elementos relacionados estão disponíveis e são adequados para o funcionamento do programa.
- Insuficiente: há evidências claras de que um ou nenhum dos elementos relacionados está disponível e é adequado para o funcionamento do programa.

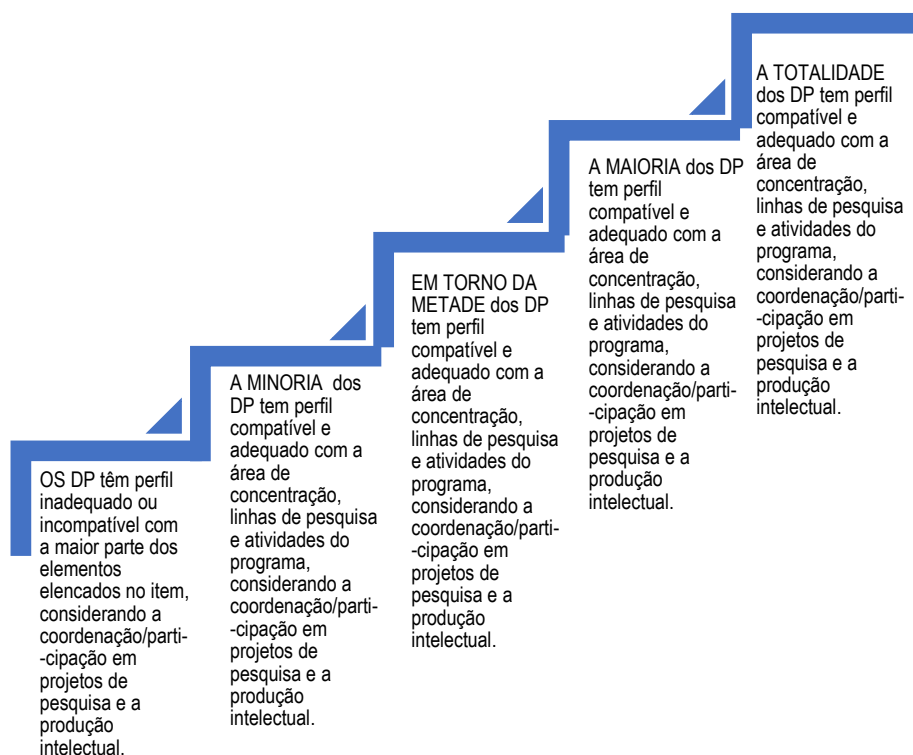
Representando a escala como uma escada, a Figura 2 apresenta outro exemplo, dessa vez para o indicador 1.2.1, cujos critérios foram definidos a partir da tendência da frequência de ocorrência do comportamento-foco (critério implicitamente

3 Os indicadores são identificados por números na Ficha de Avaliação, que pode ser acessada no Relatório de avaliação da área de educação (Capes, 2022).

numérico). O referido indicador diz respeito a “compatibilidade e adequação do perfil dos docentes permanentes (DP) em relação às áreas de concentração, linhas, projetos de pesquisa e atividades didáticas do programa” (Capes, 2022, p. 34). Observando a escala, é possível identificar a mesma qualidade de atributo em todos os níveis (adequação e compatibilidade), ainda que com frequências diversas (totalidade, maioria, em torno da metade, minoria e inadequação/incompatibilidade).

FIGURA 2

Representação gráfica da escala e critérios (baseados na frequência de ocorrência do aspecto em análise) utilizados no indicador 1.2.1 – Área de Educação, ciclo 2017-2020

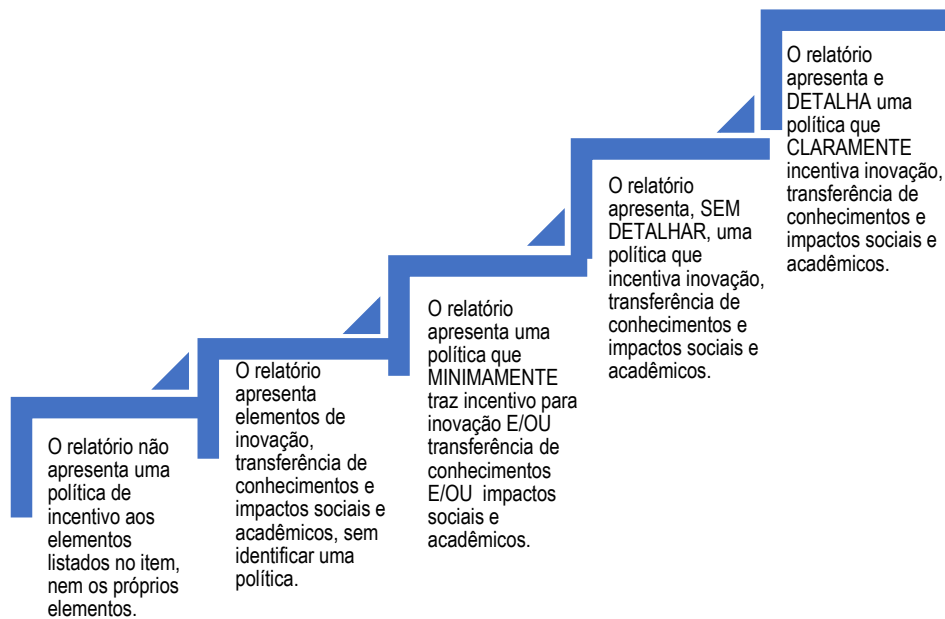


Fonte: Elaboração dos autores (2023).

Para ilustrar a proposição de critérios categoriais, o terceiro exemplo diz respeito ao indicador 1.3.2. “Desenvolvimento de política de incentivo à inovação, transferência de conhecimentos e impactos sociais e acadêmicos” (Figura 3). Nesse caso, o foco foi dado em dois aspectos da política: a presença ou não de uma política de incentivo a três elementos (inovação, transferência de conhecimentos e impactos sociais e acadêmicos) e seu grau de detalhamento e a existência ou não dos elementos e o quanto eles são incentivados por essa política.

FIGURA 3

Representação gráfica da escala e critérios (baseados na existência e outras características dos aspectos em análise) utilizados no indicador 1.3.2 – Área de Educação, ciclo 2017-2020



Fonte: Elaboração dos autores (2023).

Definidos os critérios, novas comissões foram formadas, dessa vez para a avaliação propriamente dita. Um primeiro momento foi dedicado à moderação (treinamento) dos avaliadores, de acordo com o quesito e a natureza dos programas (se acadêmico ou profissional). Nesse momento, os subgrupos de trabalho receberam a ficha e um tutorial para realizar a avaliação-piloto em dois programas, comuns a todos. Os resultados foram inseridos em formulários do Google Forms. Após um tempo predeterminado, cada subgrupo foi reunido pela coordenação da área para discussão das discrepâncias observadas, alinhamento do entendimento sobre os indicadores e seus critérios, bem como para ajustes necessários na própria base de critérios. Essas avaliações foram consideradas um exercício, sem serem posteriormente adotadas. Após esse encontro, cada avaliador recebeu a listagem de programas a serem por ele avaliados, sem saber, *a priori*, quem seriam os outros avaliadores dedicados a eles (foram formadas duplas sempre diferentes por programa).

Depois de uma semana, um segundo momento de moderação foi conduzido, para esclarecimento de dúvidas e novos alinhamentos. A partir de então, os avaliadores tiveram um prazo um pouco maior para realizar as análises e avaliações de seus quesitos, ainda individualmente. Todos esses resultados foram colocados em formulários próprios para cada quesito. Coube à coordenação da área fazer a sistematização dos resultados e registrar as inconsistências.

O terceiro momento da dinâmica envolveu a reunião (remota) de cada subgrupo, durante a qual as discrepâncias maiores (ou seja, as de dois ou mais graus),

identificadas na sistematização dos resultados obtidos por quesito, foram discutidas entre os pares, buscando diminuir o tamanho de cada discrepância até, no máximo, um grau. Em boa parte dos casos, a discrepância foi decorrente da não localização, pelo avaliador, da informação no relatório do programa (ausência de anexos, informações dispersas em todo o relatório e não no item específico, entre outras deficiências). Uma posição consensual, embora ideal, não foi exigida. Para cada um dos seis grupos⁴ de trabalho, um turno específico, tipicamente de 4 horas, foi reservado durante um período de 10 dias. A moderação das negociações entre os pares foi conduzida pela coordenação da área. Após a conclusão, cada avaliador, individualmente ou em conjunto com seu par, encaminhou seu parecer (conceito + justificativa) para a coordenação, para a sistematização final. Tal sistematização foi repassada para os avaliadores que compuseram a Comissão da Avaliação Quadrienal, última etapa de avaliação no âmbito da Área de Educação.⁵

Os resultados obtidos em cada grupo foram sistematizados, para análise da eficácia dos processos de moderação, e estão apresentados nas tabelas 1 e 2. Como se pode ver da Tabela 1, a grande maioria das avaliações (aproximadamente 80%) foi inicialmente convergente, revelando que os critérios preestabelecidos serviram para alinhar, até certo ponto, os diferentes olhares individuais.

TABELA 1

Distribuição de frequência das discrepâncias de avaliação dos indicadores qualitativos antes da moderação e ajustes - Área de Educação, ciclo 2017-2020

PROGRAMAS ACADÊMICOS	QUESITO 1		QUESITO 2		QUESITO 3	
	N	%	N	%	N	%
Total de Indicadores x PPG	2430	100,0	552	100,0	1380	100,0
Discrepâncias de 2 graus	271	11,2	73	13,2	216	15,7
Discrepâncias de 3 ou 4 graus	121	5,0			61	4,4
PROGRAMAS PROFISSIONAIS	QUESITO 1		QUESITO 2		QUESITO 3	
	N	%	N	%	N	%
Total de Indicadores x PPG	900	100,0	200	100,0	500	100,0
Discrepâncias de 2 graus	130	14,4	47	23,5	61	12,2
Discrepâncias de 3 ou 4 graus	78	8,7			32	6,4

Fonte: Elaboração dos autores com base em dados do *Relatório da comissão de análise dos indicadores qualitativos* (Capes, 2022).

- 4 Foi composto um grupo de trabalho para cada quesito (3), separados em acadêmico e profissional. Daí os seis grupos.
- 5 Os resultados apresentados nesta seção são oriundos do Relatório preliminar de avaliação (Capes, 2022).

Após o segundo encontro de moderação e o movimento de ajustes entre os pares avaliadores, praticamente foram eliminadas as discrepâncias de 2, 3 e 4 graus (Tabela 2).

TABELA 2

Distribuição de frequência das discrepâncias de avaliação dos indicadores qualitativos após a moderação e ajustes – Área de Educação, ciclo 2017-2020

PROGRAMAS ACADÊMICOS	QUESITO 1		QUESITO 2		QUESITO 3	
	N	%	N	%	N	%
Total de Indicadores x PPG	2430	100,0	552	100,0	1380	100,0
Discrepâncias de 2 graus	24	1,0	2	0,4	3	0,2
Discrepâncias de 3 ou 4 graus	2	0,1			0	0,0
PROGRAMAS PROFISSIONAIS	QUESITO 1		QUESITO 2		QUESITO 3	
	N	%	N	%	N	%
Total de Indicadores x PPG	900	100,0	200	100,0	500	100,0
Discrepâncias de 2 graus	8	0,9	4	2,0	8	1,6
Discrepâncias de 3 ou 4 graus	0	0,0			2	0,4

Fonte: Elaboração dos autores com base em dados do *Relatório da comissão de análise dos indicadores qualitativos* (Capes, 2022).

Com a avaliação dos indicadores qualitativos objetivada – via definição de critérios e moderação dos avaliadores, como mostram os dados das tabelas 1 e 2 –, na próxima seção são apresentados os resultados da análise desses indicadores para a composição da nota final de cada programa, observados o grupo de programas acadêmicos, o grupo de programas profissionais e, por fim, o conjunto de todos os programas (análise global).

ANÁLISE DO COMPORTAMENTO DOS INDICADORES QUALITATIVOS

Os procedimentos metodológicos utilizados no sentido de objetivar indicadores qualitativos, como os discutidos anteriormente, são relevantes devido à ausência de literatura sobre o tema no contexto do processo de avaliação de alto impacto da Capes. No entanto, conforme observado por Borsboom et al. (2004), a fonte primária para compreender como uma medida funciona deve ser substantiva, e não metodológica. Para fornecer tal análise substantiva, o artigo concentra-se agora nos resultados da análise do comportamento de cada um dos 47 indicadores (32 qualitativos e 15 quantitativos), relacionando-os entre si e com o resultado final da avaliação – a nota do programa. A análise foi norteada por três perguntas:

1. Quais indicadores afetam significativamente a nota do programa?
2. Quais indicadores têm o maior efeito na nota do programa?

3. Quais são as diferenças nos efeitos provocados na nota do programa considerando-se indicadores quantitativos e qualitativos?

Para examinar melhor a terceira pergunta, foram formuladas três hipóteses que focalizam a diferença do comportamento entre os dois tipos de indicador – quantitativo e qualitativo. As hipóteses supõem que o componente subjetivo dos indicadores qualitativos faz com que eles funcionem de maneira sistemática e previsivelmente diferente dos indicadores quantitativos, quando ambos os tipos buscam o mesmo objetivo, que nesse caso é dimensionar a qualidade do programa de acordo com uma escala vertical quantitativa. As três hipóteses, com uma breve justificção para cada uma, são as seguintes:

1. Devido ao fato de os indicadores qualitativos serem geralmente subjetivos e porque os avaliadores tendem a preferir fornecer resultados de avaliação positivos, a média dos indicadores qualitativos tende a ser mais alta do que a média dos indicadores quantitativos.
2. Uma vez que os membros de uma equipe de avaliação da Capes costumam ter formações e áreas de estudo semelhantes, suas percepções tendem a ser homogêneas, o que significa que os desvios-padrão dos indicadores quantitativos tendem a ser maiores do que os desvios-padrão dos indicadores qualitativos.
3. Como os indicadores quantitativos tendem a ser mais precisos do que os indicadores qualitativos, e também porque o resultado final (nota do programa) da avaliação da Capes é quantitativo, os indicadores quantitativos funcionam melhor do que os qualitativos como preditores de resultados.

As análises consideraram os 47 indicadores (ou variáveis independentes), categorizados de acordo com uma codificação de cinco pontos (5 = muito bom; 4 = bom; 3 = regular; 2 = fraco; 1 = insuficiente). Essas variáveis independentes foram examinadas em dois formatos, sem pesos (atribuindo igual valor a cada uma) e com pesos (conforme estabelecido pela Área de Educação⁶). O primeiro formato se concentra em como cada indicador funciona individualmente, e o segundo se concentra em sua contribuição para a nota final dada aos programas na Área de Educação. No entanto, as análises revelaram que os resultados com e sem ponderação eram essencialmente idênticos, uma vez que as diferenças relativas nos pesos gerais entre os indicadores eram muito pequenas, sem dúvida devido ao fato de que os pesos para os três quesitos – programa, formação e impacto na sociedade – eram os mesmos (33,3%). Portanto, na análise apresentada a seguir, apenas o formato não ponderado foi utilizado.

6 Cada indicador, assim como cada item, tem um peso específico no cálculo da nota do programa de pós-graduação.

No que diz respeito à variável dependente, a distribuição da nota final, foram considerados dois formatos, um com notas variando de 1 a 7 e outro com notas variando de 1 a 5. As duas distribuições representam o formato de duas etapas do processo de avaliação da Capes, em que inicialmente os programas recebem notas de 1 a 5 e, no segundo momento, aqueles com nota 5 são separados em notas 5, 6 e 7. Na segunda etapa, indicadores com peso relativamente baixo na primeira etapa, como internacionalização (3.3.2a) e liderança (3.1.5), assumem uma importância muito maior. Uma vez que os 47 indicadores que compunham o formulário de avaliação foram projetados para uso apenas na primeira etapa, o formato de cinco pontos gerou resultados que faziam mais sentido do ponto de vista de validade aparente, e, portanto, foi o formato adotado para as análises de regressão do estudo.

Para examinar as três hipóteses, apenas programas com dados completos foram considerados, o que resultou na retirada de programas criados recentemente, ainda sem produtos finais e egressos. Dessa forma, dos 188 programas avaliados pela área, foram retirados 15 casos, sendo 9 de natureza acadêmica e 6 da modalidade profissional. Além disso, para as análises, foram utilizados três agrupamentos de programas: (1) todos os programas ($n = 173$); (2) programas acadêmicos ($n = 128$); e (3) programas profissionais ($n = 45$). Ao abordar as hipóteses 1 e 2, foram comparadas as médias e desvios-padrão dos indicadores. Para avaliar a hipótese 3, foi utilizada a regressão linear múltipla *stepwise*. Para todas as hipóteses, os 15 indicadores quantitativos e os 32 indicadores qualitativos foram reduzidos a duas variáveis compostas, denominadas aqui de QUANTI e QUALI.⁷ Como será visto, indicadores individuais também foram utilizados no contexto das análises de regressão. A seguir são apresentados os resultados e a discussão das hipóteses do estudo.

Hipóteses 1 e 2

Conforme previsto, de acordo com uma lógica baseada em diferenças nos graus de subjetividade, a média para QUALI foi substancialmente maior do que a média para QUANTI em todos os três grupos analisados (global, acadêmico e profissional). Da mesma forma, os resultados de desvio-padrão também estiveram de acordo com as expectativas para os três segmentos analisados.

O fato de as médias dos programas acadêmicos terem sido mais altas em relação aos profissionais não é surpreendente considerando que os primeiros são, de modo geral, mais maduros que os últimos. Observando o panorama da Área de Educação, seu primeiro programa acadêmico foi criado em 1965, sendo 45 anos mais velho que a primeira oferta de pós-graduação profissional. A diferença na

7 As variáveis QUANTI e QUALI foram calculadas por meio da soma dos resultados de todas as variáveis quantitativas (no caso de QUANTI) e qualitativas (QUALI) dividido pelo número de variáveis contempladas (32 QUALI e 15 QUANTI).

maturidade dos programas pode também explicar por que as dispersões, observadas as duas modalidades, foram substancialmente mais altas no grupo dos profissionais, uma vez que a homogeneidade da qualidade, que tende a se dar entre os programas com o passar do tempo, ainda não aconteceu. As medidas descritivas para as variáveis resumo dos dois tipos de indicadores são apresentadas na Tabela 3. Essas informações serão utilizadas na discussão das hipóteses 1 e 2.

TABELA 3

Estatísticas descritivas das variáveis resumo QUALI e QUANTI por tipo de programa e também para o grupo geral - Área de Educação, ciclo 2017-2020

NATUREZA DO PROGRAMA	VARIÁVEL	N	MÍNIMO	MÁXIMO	MÉDIA	DESVIO-PADRÃO
Acadêmico	QUANTI	128	2,40	5,00	4,12	0,48
	QUALI	128	3,31	5,00	4,39	0,43
Profissional	QUANTI	45	2,40	4,60	3,61	0,58
	QUALI	45	2,31	4,81	4,02	0,52
Global	QUANTI	173	2,40	5,00	3,99	0,55
	QUALI	173	2,31	5,00	4,29	0,48

Fonte: Elaboração dos autores com base nos resultados da avaliação da Capes para a Área de Educação no ciclo 2017-2020.

Hipótese 3

A análise de regressão múltipla torna possível medir a contribuição de cada variável independente para o comportamento de uma variável dependente escalonada, controlando as outras variáveis independentes na equação linear. A análise de regressão é um instrumento útil para a avaliação de indicadores, pois, ao revelar o poder de explicação de cada um em relação à variável dependente, também capta seu poder de discriminação em relação aos outros incluídos no mesmo modelo.⁸ Assim, usando os procedimentos de *best fit*, a referida análise serve para estimar tanto a funcionalidade quanto o poder de discriminação dos 47 indicadores adotados pela Área de Educação.

Na discussão da hipótese 3, vamos avaliar duas medidas derivadas dos modelos de análise de regressão. Uma delas se refere a cada uma das variáveis independentes no modelo, estabelecendo o efeito único da variável independente na variável dependente (coeficiente de regressão) e a outra revelando o nível descritivo

8 Os autores reconhecem que, em razão do número relativamente pequeno de casos analisados frente ao número de variáveis explicativas, os resultados devem ser considerados com cautela, em especial com relação ao grupo de programas profissionalizantes. O uso desse tipo de análise se justifica, no entanto, como instrumento exploratório para detectar possíveis tendências e relações, conforme defendido por Tufte (1974), quanto à análise de fenômeno de cunho social.

desse efeito (p-valor do indicador). Consideramos os coeficientes de regressão padronizados em termos da relação entre desvios-padrão e, portanto, sua magnitude pode ser comparada com a das outras variáveis independentes, permitindo, assim, uma classificação dos efeitos dos indicadores do maior para o menor. Apresentamos a seguir os resultados obtidos a partir do ajuste dos modelos de interesse.

Como primeiro passo, foram estimados os efeitos relativos das variáveis resumo QUANTI e QUALI para cada um dos três grupos em estudo, utilizando a nota do programa resultante da primeira rodada de avaliação como variável dependente, conforme pode ser observado na Tabelas 4, que trata respectivamente dos grupos geral (n = 173), acadêmico (n = 128) e profissional (n = 45).

TABELA 4

Estimativas dos efeitos (e respectivos níveis descritivos) dos modelos com as variáveis resumo Quali e Quanti, ajustados para os três grupos considerados – Área de Educação, ciclo 2017-2020

NATUREZA DO PROGRAMA	VARIÁVEL	COEFICIENTE	P-VALOR
Acadêmico	QUANTI	0,525	0,000
	QUALI	0,370	0,000
Profissional	QUANTI	0,481	0,000
	QUALI	0,400	0,001
Geral	QUANTI	0,490	0,000
	QUALI	0,355	0,000

Fonte: Elaboração dos autores com base nos resultados da Avaliação da Capes para a Área de Educação no ciclo 2017-2020.

Ao comparar os efeitos relativos de QUANTI e QUALI, pode-se observar que, para todos os grupos de programas, o poder de explicação é maior para o primeiro do que para o segundo. No entanto, ambas as variáveis independentes são estatisticamente significativas, e a diferença entre o coeficiente de regressão das duas variáveis é relativamente pequeno. As mesmas tendências se repetem ao considerar os dois grupos – acadêmicos e profissionais – separadamente, sendo a diferença entre os coeficientes para o conjunto dos programas profissionais especialmente pequeno (0.481 x 0.400). Conclui-se, assim, que embora o efeito dos indicadores quantitativos tenda a ser maior, os indicadores qualitativos também contribuem para a nota do programa de maneira significativa.

Como segundo passo, cada um dos indicadores individuais – quantitativos e qualitativos – foi introduzido no modelo de regressão e, usando uma abordagem *stepwise*, aqueles que contribuíram melhor para explicar a variável de saída foram identificados. Para o grupo composto por todos os programas, o procedimento

estatístico reduziu os 47 indicadores para apenas sete, que, em seu conjunto, explicaram aproximadamente 70% da variância nas notas atribuídas (Tabela 5). Dos sete estatisticamente significativos, quatro diziam respeito ao quesito “Formação” (publicação de estudantes derivada do trabalho de conclusão, produção docente de alta qualidade, alunos que são autores e docentes envolvidos em pesquisa e orientação) e três foram relacionados ao quesito “Impacto na sociedade”. Enquanto os indicadores que trataram da formação foram todos de natureza quantitativa, os outros três foram qualitativos, tratando, respectivamente, da liderança acadêmica dos docentes, dos melhores produtos do programa e da relevância das atividades sociais, econômicas e culturais do programa.

TABELA 5

Estimativas dos coeficientes (e respectivos níveis descritivos) dos indicadores que se mostraram estatisticamente significantes no modelo ajustado para a nota final do grupo geral de programas – Área de Educação, ciclo 2017-2020

INDICADOR	TIPO	COEFICIENTE	P-VALOR
2.1.2 – Formação – % Publicação discente resultante de TCC	QUANTI	0,317	0,000
2.4.2 – Formação – % Produção docente de alta qualidade	QUANTI	0,247	0,000
2.2.1 – Formação – % Discentes que são autores	QUANTI	0,188	0,001
3.1.5 – Impacto – Relevância da participação acadêmica docente	QUALI	0,149	0,008
3.1.3 – Impacto – Qualidade das produções do PPG	QUALI	0,144	0,003
2.5.3 – Formação – % Docentes envolvidos em pesquisa e orientação	QUANTI	0,132	0,007
3.2.1 – Impacto – Relevância (econômica, social e cultural) das atividades do PPG	QUALI	0,129	0,014

Fonte: Elaboração dos autores com base nos resultados da Avaliação da Capes para a Área de Educação no ciclo 2017-2020.

Nota: Os coeficientes estão ordenados de acordo com o efeito na variação da nota média do programa.

Quando os outros dois grupos, programas acadêmicos e programas profissionais, são considerados separadamente, os resultados indicam perfis distintos, tanto em relação aos resultados do grupo global quanto em respeito à comparação entre os dois grupos em si. Para o conjunto de programas acadêmicos (Tabela 6), o número de variáveis significativas aumenta de sete para oito, inclusive, agora, quatro indicadores qualitativos. É interessante observar que, desses quatro, dois integram o quesito 1 “Programa”, sendo um sobre relações com a graduação e outro focando a qualidade dos dados utilizados no planejamento estratégico.

TABELA 6

Estimativas dos coeficientes (e respectivos níveis descritivos) dos indicadores que se mostraram estatisticamente significantes no modelo ajustado para a nota final do grupo de programas Acadêmicos – Área de Educação, ciclo 2017-2020

INDICADOR	TIPO	COEFICIENTE	P-VALOR
2.1.2 – Formação – % Publicação discente resultante de TCC	QUANTI	0,274	0,000
2.2.4 – Formação – % Discente com artigos Qualis \geq A4	QUANTI	0,218	0,000
3.1.5 – Impacto – Relevância da participação acadêmica docente	QUALI	0,212	0,000
3.1.2 – Impacto – Razão artigos (A1+A2) do PPG/ N docentes	QUANTI	0,180	0,009
3.1.3 – Impacto – Qualidade das produções do PPG	QUALI	0,159	0,002
2.4.1 – Formação – Média da produção docente	QUANTI	0,149	0,034
1.3.4 – Programa – Pertinência dos dados e participação de pessoas no planejamento	QUALI	0,117	0,019
1.2.2 – Programa – Interação do PPG com a graduação	QUALI	0,116	0,019

Fonte: Elaboração dos autores com base nos resultados da Avaliação da Capes para a Área de Educação no ciclo 2017-2020.

Nota: Os coeficientes estão ordenados de acordo com o efeito na variação da nota média do programa.

Por outro lado, no caso dos programas profissionais (Tabela 7), o número de variáveis independentes estatisticamente significativas se reduz para apenas quatro, incluindo duas de natureza qualitativa – liderança acadêmica dos docentes e adequação temática dos trabalhos de conclusão. O fato de que o último tem coeficiente negativo é surpreendente, sugerindo uma falta de clareza sobre o significado da palavra “adequação”, algo aparentemente não resolvido pelas orientações fornecidas.

Vale mencionar que o peso para o indicador 2.2.6, que trata de egressos, tem o maior coeficiente entre todos os indicadores analisados, o que é relevante, especialmente no contexto dos profissionais, pois sua finalidade expressa é a preparação para o mercado de trabalho.

TABELA 7

Estimativas dos coeficientes (e respectivos níveis descritivos) dos indicadores que se mostraram estatisticamente significantes no modelo ajustado para a nota final do grupo de programas profissionalizantes- Área de Educação, ciclo 2017-2020

INDICADOR	TIPO	COEFICIENTE	P-VALOR
2.2.6 – Formação – % Egressos com artigos Qualis \geq A4	QUANTI	0,552	0,000
3.1.5 – Impacto – Relevância da participação acadêmica docente	QUALI	0,389	0,000
2.1.2 – Formação – % Publicação discente resultante de TCC	QUANTI	0,258	0,011
2.1.1 – Formação – Adequação temática dos TCC com as linhas de pesquisa	QUALI	-0,202	0,038

Fonte: Elaboração dos autores com base nos resultados da Avaliação da Capes para a Área de Educação no ciclo 2017-2020.

Nota: Os coeficientes estão ordenados de acordo com o efeito na variação da nota média do programa.

Como visto, os resultados da análise dos indicadores desagregados são coerentes com os dos indicadores agregados (QUANTI *versus* QUALI), evidenciando um peso relativamente maior para os quantitativos, porém indicadores qualitativos são também estatisticamente relevantes, sugerindo eficácia por parte dos dois tipos de indicador como componentes da avaliação de programas da Área de Educação.⁹

Cinco descobertas produzidas pelas análises de regressão são particularmente notáveis. A primeira é que o desejo da Capes de aumentar a ênfase dos fatores de “Formação” e “Impacto na sociedade” parece ter sido alcançado, em todos os grupos investigados, sendo identificados indicadores significativos dentro de cada um dos dois quesitos de interesse específico. Indicadores do quesito “Programa” também foram representados, pelo menos no contexto do grupo acadêmico, algo relevante considerando sua natureza eminentemente qualitativa. A segunda é a relativa ausência do indicador que dimensiona a média ponderada das publicações dos professores. Esse indicador não é significativo na análise geral, nem na análise dos profissionais, e, na análise dos acadêmicos, embora significativo, tem peso limitado, sendo superado por cinco outras variáveis na equação. O referido indicador tem sido, historicamente, o determinante mais forte da nota final do programa, uma tendência amplamente criticada no passado e intencionalmente diluída na avaliação mais recente, sendo colocada como uma subunidade de “Formação”, em vez de compreender um quesito separado dedicado à “produção intelectual”.

Uma terceira consideração trata do fato de que, para o grupo de programas profissionais, apenas um dos indicadores de impacto na sociedade tenha sido significativo. Os programas profissionais são idealmente projetados para gerar alterações positivas e melhorias concretas na sociedade, supostamente compensando o foco excessivamente acadêmico dos programas de pós-graduação mais tradicionais do Brasil. É provável que os indicadores que trataram do impacto social do programa, ou seja, os contidos nos itens 3.2 e 3.3, não tenham influenciado a nota estatisticamente, não por terem ausência de impacto no seu meio, mas por terem impacto que varia pouco entre os programas, pelo menos no olhar dos avaliadores, algo observado pelos autores deste trabalho ao conduzir a avaliação. Ademais, como mencionado anteriormente, os programas profissionais na área tendem a ser relativamente novos, com um tempo de funcionamento de menos de 10 anos, então talvez a falta de maturidade do programa no grupo profissional explique esse achado inesperado.

9 Vale registrar que a apresentação dos resultados referentes às três hipóteses neste artigo é propositalmente resumida, focalizando apenas o que é essencial para discussão sobre a pertinência dos indicadores qualitativos *versus* quantitativos. Estatísticas adicionais podem ser obtidas via contato com os autores deste artigo.

Outra observação relevante é a notável diferença entre os resultados dos dois grupos – acadêmicos e profissionais – analisados, tendo indicadores significativos bem diferenciados, em termos de número e também de conteúdo. O argumento, frequentemente promovido, de que não existe distinção entre as duas modalidades não se sustenta diante dos achados apresentados, aplicados, é claro, apenas para a Área de Educação.

Por fim, como a quinta descoberta, vale ressaltar que o total dos indicadores destacados pela análise, considerando-se os três grupos focalizados, é bem menor do que o total de indicadores utilizados na avaliação. Em outras palavras, embora todos os 47 indicadores tenham sido considerados na avaliação realizada pela Área de Educação, são 14 aqueles com maior importância na explicação das diferenças entre as notas atribuídas. A relevância deste ponto será retomada no tópico final deste artigo.

CONSIDERAÇÕES FINAIS

Em livro recentemente publicado, baseado em sua tese de doutoramento aprovada na Universidade Leiden, na Holanda, André Brasil argumentou sobre a necessidade de uma melhor contextualização do programa para a avaliação da Capes, de forma a considerar a imensa diversidade regional e institucional que caracteriza a realidade da pós-graduação no Brasil (Brasil, 2023). Para que se atinja tal objetivo, o autor defende maior ênfase tanto na autoavaliação do programa como na utilização de indicadores qualitativos na avaliação externa. É interessante perceber que esse ponto de vista, justificado de maneira detalhada pelo autor, é semelhante àquele que lastra a decisão, pelo CTC-ES, em 2018, de dar maior ênfase à autoavaliação (Leite et al., 2020) e aos indicadores qualitativos no contexto da avaliação dos programas de pós-graduação referente ao período de 2017-2020, tratados neste trabalho.

Em 2018, o CTC-ES resolveu dar maior ênfase aos indicadores qualitativos no contexto da avaliação dos programas de pós-graduação referente ao período de 2017-2020. Como já observado, essa decisão foi fortalecida por recomendações, feitas no passado por membros da comunidade acadêmica, que argumentaram que uma abordagem mais subjetiva no processo de avaliação permitiria uma visão holística de cada programa, considerando o contexto e a complexidade do processo de formação de pós-graduação. Contudo, mesmo que altamente justificado, o foco qualitativo não é fácil de implementar, especialmente considerando a demanda da Capes pelo *ranking* de programas com base em uma escala quantitativa.

De fato, a utilização predominante de indicadores qualitativos é contrária à ênfase quantitativa praticada desde a implementação do modelo de avaliação da Capes, introduzido em 1980 e ajustado em 1998, a partir do qual os programas vêm

sendo classificados em uma escala de qualidade de sete pontos. Nesse cenário, a abordagem quantitativa era considerada a mais apropriada, pelas supostas vantagens em termos de objetividade, reprodutibilidade, transparência e eficiência, exigindo menos tempo e permitindo a aplicação de técnicas estatísticas sofisticadas, elementos valorizados considerando o caráter *high stakes* e de larga escala da avaliação.

Por outro lado, o valor dos indicadores qualitativos não pode ser negado. Para que fossem utilizados no contexto Capes, portanto, precisaram ser objetivados de maneira a minimizar a subjetividade individual e promover transparência e comparabilidade, sem, por outro lado, perder a sensibilidade para a complexidade de formação em seus contextos singulares.

Dentro da Capes, a Área de Educação abraçou a recomendação de dar maior ênfase aos indicadores qualitativos, criando uma matriz de avaliação na qual mais de dois terços de todos os indicadores (ou seja, 32 de 47) eram de natureza qualitativa. No entanto, ao fazer isso, a área reconheceu a importância da objetivação e da transparência, considerando a classificação necessária e seu impacto em relação ao financiamento e à regulamentação do programa. Assim, após uma discussão teórica de conceitos-chave, o artigo descreveu dois processos de objetivação: o estabelecimento de critérios para a avaliação de cada indicador e a moderação dos julgamentos feitos pelos avaliadores. Os dois processos, segundo os autores, oferecem modelos que podem ser utilizados por outras áreas de avaliação, muitas das quais não utilizam métodos qualitativos em suas pesquisas e, portanto, têm pouca experiência com o uso de indicadores qualitativos para fins de avaliação de programas de pós-graduação.

Este artigo relata os esforços dos autores para examinar o comportamento dos indicadores no contexto da avaliação da Capes, de maneira exploratória, analisando os dados produzidos pela Área de Educação, referentes ao quadriênio 2017-2020, em termos das três hipóteses que, ao serem testadas, indicariam que os dois tipos de indicadores – quantitativos e qualitativos – se comportam de maneira similar ou diferente. Como relatado, as análises revelaram apenas um suporte parcial para as hipóteses, pois alguns indicadores qualitativos foram preditores significativos das notas finais do programa. No entanto, os indicadores quantitativos tenderam a ser mais robustos. Outros achados gerados pelas análises também merecem atenção. Embora o poder explicativo dos indicadores tenha variado entre os grupos analisados (global, acadêmico ou profissional), ficou evidente que, no sentido geral, os indicadores que trataram do quesito “Formação” tendem a ser os predominantes, o que está de acordo com o entendimento de que a finalidade mais importante da pós-graduação é a preparação dos seus alunos. Especialmente relevante é o fato de que, em todos os três grupos, o indicador “2.1.2 – Formação – % Publicação

discente resultante de TCC” se destacou, mostrando o valor de considerar a formação em termos dos seus resultados concretos. Por outro lado, entre os indicadores relacionados ao impacto do programa na sociedade, observa-se o peso do indicador que aborda a participação acadêmica dos docentes permanentes, evidenciando a importância da atuação do professor na determinação da nota final do programa, não tanto por meio de sua produção bibliográfica (conforme imaginado no passado), mas pelas atividades de liderança no contexto da comunidade acadêmica.

Além disso, vale enfatizar que o número total de indicadores significativos, quantitativos e qualitativos foi relativamente pequeno dado o seu número total, sugerindo que a área poderia ter obtido os mesmos resultados de classificação usando um número muito menor de preditores. O processo de redução pode ser iniciado com a consideração dos 14 indicadores identificados nesse estudo como sendo estatisticamente significativos, mas uma definição deve aguardar uma análise mais sofisticada do poder de discriminação, como visto, por exemplo, em Klecka (1980). Ao avaliar este último ponto, vale ressaltar que um número menor de indicadores simplificaria a avaliação, tornando-a mais fácil tanto para os programas na produção de seus relatórios anuais para a Capes quanto para os avaliadores na formulação de seus julgamentos coletivos. No entanto, deve-se entender que a avaliação da Capes tem dois objetivos, um somativo e outro formativo. A classificação de programas de acordo com o desempenho ao longo de um período de quatro anos é somativa, enquanto o *feedback* fornecido aos programas, por meio de seminários, interação com a coordenação e, especialmente, seu relatório final de avaliação, é formativo. Enquanto muitos dos 32 indicadores qualitativos contribuem apenas de forma limitada para a dimensão somativa, há razão para acreditar que sua importância para o *feedback* do programa é significativa, uma vez que indicam tendências relacionadas a processos complexos de maneira contextualizada.

Assim, ao concluir este artigo, os autores fazem três recomendações-chave: (1) buscar reduzir o número total de indicadores para simplificar o processo de avaliação; (2) continuar a valorizar os indicadores qualitativos, especialmente por seu valor como fonte importante para o *feedback* necessário para a melhoria do programa; e (3) submeter os indicadores qualitativos à objetivação, preestabelecendo critérios e moderando os julgamentos dos avaliadores. Essas recomendações se aplicam à avaliação da Capes, mas também a outras avaliações de larga escala que busquem determinar níveis de qualidade de um objeto. Também, em referência à Capes, é recomendado que a agência promova interações entre as áreas para fornecer orientações comuns e compartilhar procedimentos para lidar com o uso de indicadores qualitativos, uma vez que, para muitas áreas, sua utilização não tem sido prática histórica.

REFERÊNCIAS

- Barreto, J. A. E. (1993). Avaliação: Mitos e armadilhas. *Ensaio: Avaliação e Políticas Públicas em Educação*, 1, 45-54.
- Biagioli, M. (2018). Quality to impact, text to metadata: Publication and evaluation in the age of metrics. *KNOW: A Journal on the Formation of Knowledge*, 2(2), 249-275. <https://doi.org/10.1086/699152>
- Borsboom, D., Mellenbergh, G. J., & Heerden, J. van. (2004). The concept of validity. *Psychological Review*, 111(4), 1061-1071. <https://doi.org/10.1037/0033-295x.111.4.1061>
- Brasil, A. (2023). *Advancing the evaluation of graduate education: Towards a multidimensional model in Brazil*. Universiteit Leiden.
- Castro, C. de M. (2006). A Capes na visão de um ex-diretor. *Análise*, 17(2), 360-376.
- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). (2021). *Equivalência entre conceitos e notas*. https://www.gov.br/capes/pt-br/centrais-de-conteudo/documentos/avaliacao/2021.04.05_RelatrioGTequivalenciaconceitosnotasversoformatadaV3.pdf
- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). (2022). *Relatório de avaliação de Área de Educação*. https://www.gov.br/capes/pt-br/centrais-de-conteudo/documentos/avaliacao/19122022_RELATORIO_AVALIACAO_QUADRIENAL_comnotaEducao.pdf
- Fórum de Coordenadores de Programas de Pós-Graduação em Educação (Forpred). (2017). *Carta de Fortaleza*. (Documento não publicado).
- Gatti, B., Andre, M., Favero, O., & Candau, V. M. F. (2003). O modelo de avaliação da Capes. *Revista Brasileira de Educação*, 22, 137-144. <https://doi.org/10.1590/S1413-24782003000100012>
- Gomez, M. N. G. de. (2014). Da questão da validade ao julgamento de valor: Mediação informacional da avaliação científica. In *Anais do 15. Encontro Nacional de Pesquisadores em Ciência da Informação*.
- Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2009). *Handbook on impact evaluation: Quantitative methods and practices*. World Bank Publications.
- Klecka, W. R. (1980). *Discriminant analysis*. Sage.
- Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, 38(1), 201-221. <https://doi.org/10.1146/annurev-soc-070308-120022>
- Leite, D., Verhine, R. E., Dantas, L. M. V., & Bertolin, J. C. D. (2020). A autoavaliação na pós-graduação (PG) como componente do processo avaliativo Capes. *Avaliação: Revista da Avaliação da Educação Superior*, 25(2), 339-353. <https://doi.org/10.1590/S1414-4077/S1414-40772020000200006>
- Minayo, M. C. de S. (2009). Construção de indicadores qualitativos para avaliação de mudanças. *Revista Brasileira de Educação Médica*, 33(Supl. 1), 83-91. <http://educa.fcc.org.br/pdf/rbem/v33s01/v33s01a09.pdf>
- Moreira, C. O. F., Hortale, V. A., & Hartz, Z. M. A. (2004). Avaliação da pós-graduação: Buscando consenso. *Revista Brasileira de Pós-Graduação*, 1(1), 26-40. <https://www.arca.fiocruz.br/handle/icict/60206>
- Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, 34(5 Pt 2), 1189-1208.
- Portaria n. 122, de 5 de agosto de 2021. (2021). Consolida os parâmetros e os procedimentos gerais da Avaliação Quadrienal de Permanência da pós-graduação stricto sensu no Brasil. Brasília, DF. <http://cad.capes.gov.br/ato-administrativo-detallar?idAtoAdmElastic=6742>

- Porter, T. M. (1995). Quantification and the accounting ideal in science. *Social Studies in Science*, 22(4), 633-652. <https://www.jstor.org/stable/285458>
- Reckase, M. D. (2023). *The psychometrics of standard setting: Connecting policy and test scores*. CRC Press.
- Richie, J., & Lewis, J. (Eds.) (2003). *Qualitative research practice: A guide to social science students and researchers*. Sage.
- Secchi, L. (2014). Instrumentos de políticas públicas. In R. F. Boullosa (Org.), *Dicionário para a formação em gestão social* (pp. 103-105). CIAGS/UFBA.
- Sguissardi, V. (2006). A avaliação defensiva no “modelo Capes de avaliação”: É possível conciliar avaliação educativa com processos de regulação e controle do Estado? *Perspectiva*, 24(1), 49-88. <http://educa.fcc.org.br/pdf/rp/v24n01/v24n01a03.pdf>
- Spagnolo, F., & Calhau, M. G. (2002). Observadores internacionais avaliam a avaliação da Capes. *Infocapes*, 10(1), 7-34.
- Stufflebeam, D., & Shinkfield, A. J. (2007). *Evaluation theory, models and applications*. HB Printing.
- Tierney, R., & Simon, M. (2004). What’s still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research, and Evaluation*, 9(1), Article 2. <https://doi.org/10.7275/jtvt-wg68>
- Todeschini, R., & Baccini, A. (2016). *Handbook of bibliometric indicators: Quantitative tools for studying and evaluating research*. John Wiley & Sons.
- ToloData. (2021). *Qualitative indicators and their relevance in monitoring and evaluation*. <https://www.toladata.com/blog/qualitative-indicators-and-their-relevance-in-monitoring-and-evaluation/>
- Tufte, E. R. (1974). *Data analysis for politics and policy*. Prentice-Hall.
- United States Agency for International Development (Usaid). (2005). *Handbook on qualitative indicators*. https://pdf.usaid.gov/pdf_docs/PNAEB361.pdf
- Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (2004). *Avaliação de programas: Concepções e práticas*. Gente.