

**QUANDO OS RESULTADOS
PERDEM SUA IDENTIDADE:
SEDIMENTAÇÃO
NORMATIVA E
ENCAPSULAMENTO NA
ARQUITETURA DE
MENSURAÇÃO DO
DESEMPENHO DAS
UNIVERSIDADES FEDERAIS
BRASILEIRAS**

**WHEN RESULTS LOSE THEIR IDENTITY: NORMATIVE LAYERING AND
ENCAPSULATION IN THE PERFORMANCE MEASUREMENT ARCHITECTURE
OF BRAZILIAN FEDERAL UNIVERSITIES**

Ciências Sociais Aplicadas • 12/09/2026

REGISTRO DOI: [10.70773/revistatopicos/789157609](https://doi.org/10.70773/revistatopicos/789157609)

Carlos David Santos Branco¹

RESUMO

Sistemas públicos de mensuração podem produzir informação de resultado sem disponibilizá-la como categoria autônoma no nível em que a accountability se exerce. O artigo denomina essa falha encapsulamento de resultados: a informação de resultado passa a integrar um composto de nível superior e deixa de estar disponível isoladamente no nível da prestação de contas. Com base em pesquisa documental, 23 indicadores públicos aplicáveis às universidades federais brasileiras, provenientes do Censo da Educação Superior, da Decisão TCU nº 408/2002, do Sinaes e dos indicadores de fluxo do Inep, foram classificados por protocolo replicável ancorado na definição operacional oficial de cada indicador. Doze medem atividade, três medem produto, três medem resultado e cinco são híbridos. Não foi identificado, no corpus analisado, indicador oficial de resultado puro no nível institucional. A informação é produzida no nível do curso e chega aos níveis agregados fundida a insumos e processos dentro de índices compostos. A decomposição ponderal do Conceito Preliminar de Curso ilustra o fenômeno: 55% da ponderação corresponde a componentes de resultado, 30% a insumo docente e 15% a percepção de processo, fração ainda mais diluída na agregação que produz o Índice Geral de Cursos. O contraste entre as camadas normativas de 2008 e 2017, reconstruído por rastreamento de processo sobre documentos datados, distingue o mecanismo de sedimentação normativa de explicações rivais como falha de produção, ambiguidade estratégica e isomorfismo. A contribuição é um elo intermediário na cadeia de gestão do desempenho: mensurar, preservar a legibilidade, incorporar e usar.

Palavras-chave: indicadores de desempenho; universidades federais; mensuração de resultados; accountability; administração pública.

ABSTRACT

Measurement systems can produce outcome information without making it available as an autonomous category at the level where accountability is exercised. The article calls this gap outcome encapsulation: outcome information is absorbed into a higher-level composite and ceases to be available on its own at the accountability level. Based on documentary research, 23 public indicators applicable to Brazilian federal universities, drawn from the Higher Education Census, TCU Decision 408/2002, Sinaes, and Inep flow indicators, were classified through a replicable protocol anchored in each indicator's official operational definition. Twelve measure activity, three measure outputs, three measure outcomes, and five are hybrid. No official pure-outcome indicator was identified, in the corpus analyzed, at the institutional level. Outcome information is produced at the course level and reaches aggregate levels merged with inputs and processes inside composite indices. The weight structure of the Preliminary Course Concept (CPC) illustrates the pattern: 55% of the weighting corresponds to outcome components, 30% to faculty inputs, and 15% to process perceptions, a share diluted further once it feeds into the General Course Index (IGC). Contrasting the normative layers of 2008 and 2017, reconstructed through process tracing across dated documents, distinguishes the mechanism of normative layering from rival explanations such as production failure, strategic ambiguity, and isomorphism. The contribution is an intermediate link in the performance management chain: measure, preserve legibility, incorporate, use.

Keywords: performance indicators; federal universities; outcome measurement; accountability; public administration.

1. INTRODUÇÃO

A mensuração do desempenho tornou-se um dos traços mais visíveis da administração pública desde as reformas gerenciais dos anos 1980 e 1990, quando governos passaram a produzir volumes crescentes de indicadores para informar decisões, orientar o financiamento e prestar contas à sociedade (Hood, 1991; Osborne; Gaebler, 1992; Pollitt; Bouckaert, 2011). No Brasil, a reforma gerencial de 1995 incorporou a orientação para resultados como princípio organizador do aparelho do Estado (Bresser-Pereira, 1996), em trajetória posterior marcada por avanços e descontinuidades (Abrucio, 2007; Secchi, 2009).

Poucos campos absorveram essa cultura de mensuração com a intensidade da educação superior (Cave *et al.*, 1997; Espeland; Sauder, 2007; Hazelkorn, 2015). Sobre as universidades federais brasileiras sobrepõem-se, no período analisado, pelo menos quatro instrumentos públicos de abrangência nacional: as estatísticas do Censo da Educação Superior, os indicadores de gestão exigidos pela Decisão nº 408/2002 do Tribunal de Contas da União (TCU), os indicadores de qualidade do Sistema Nacional de Avaliação da Educação Superior (Sinaes) e os indicadores de fluxo e trajetória estudantil calculados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). A expansão da mensuração, contudo, não garante que as dimensões relevantes do desempenho permaneçam distinguíveis quando a informação atravessa níveis institucionais. Esse é o problema deste artigo.

1.1. O Problema

O que está em exame não é a quantidade de indicadores, tampouco a escassez de medidas de resultado em si mesma. O problema é o processo pelo qual a informação de resultado, depois de produzida

pelo Estado, perde legibilidade ao atravessar os níveis da arquitetura institucional: do curso, onde é gerada, até a instituição, onde a prestação de contas se exerce.

A literatura distingue o que uma organização faz (atividade), o que entrega (produto) e o que muda em razão de sua ação (resultado) (Hatry, 2006; Bouckaert; Halligan, 2008; Van Dooren; Bouckaert; Halligan, 2015), categorias detalhadas na seção 2.3. Um sistema pode conter dezenas de indicadores e ainda assim medir predominantemente uma única dessas dimensões. Quando isso ocorre, a interpretação do desempenho fica sistematicamente enviesada. Volumes de atividade são lidos como se fossem efeitos, e a ausência de informação legível sobre resultados passa despercebida sob a aparência de abundância informacional (Radin, 2006; Van Dooren; Bouckaert; Halligan, 2015). A prática administrativa agrava o quadro: diplomas emitidos são tratados como sinônimo de formação de qualidade, e índices compostos misturam insumos, processos e efeitos em um único número. Essas confusões condicionam o que gestores priorizam, o que órgãos de controle cobram e o que a sociedade consegue enxergar (Behn, 2003; Bruijn, 2002).

1.2. O Que Sabemos, o Que Não Sabemos e por Que o Caso Importa

Três correntes de literatura cercam o problema, cada qual explicando um aspecto dele. A literatura de produção mostra que resultados são tecnicamente caros de medir, manifestam-se com defasagem e têm atribuição causal apenas parcial, de modo que sistemas tendem a produzi-los menos do que medidas de atividade (Hatry, 2006; Van Dooren; Bouckaert; Halligan, 2015). A literatura de

uso mostra que medidas de resultado atreladas a consequências geram distorções comportamentais (Smith, 1995; Bevan; Hood, 2006; Propper; Wilson, 2003) e que o uso efetivo da informação depende de condições organizacionais bem documentadas (Moynihan; Pandey, 2010; Kroll, 2015; Gerrish, 2016). A literatura de índices compostos, por fim, estabelece que a agregação sacrifica capacidade de discriminação analítica (Jannuzzi, 2005; Dill; Soo, 2005). Combinadas, as três correntes deixam sem resposta por que uma categoria de informação oficialmente produzida deixa de estar disponível como unidade autônoma no nível em que a accountability se exerce.

A cadeia canônica mensurar, incorporar e usar (Van Dooren; Bouckaert; Halligan, 2015) trata como não problemático o intervalo entre produzir a informação e preservar sua identidade categorial ao longo do trajeto. Falta nessa cadeia um elo que garanta legibilidade à categoria mensurada quando a informação atravessa níveis de agregação. Este estudo toma esse elo ausente como objeto.

O caso brasileiro é adequado para revelar a lacuna porque nele os dois diagnósticos disponíveis falham ao mesmo tempo. O Estado produz informação genuína de resultado sobre suas universidades (aprendizagem aferida por exame nacional, valor agregado, desfechos de trajetória), o que torna a falha de produção insuficiente para explicar o padrão observado. Essa informação, porém, não chega de forma legível ao nível institucional da prestação de contas, o que torna prematura qualquer discussão sobre falha de uso: não se usa bem o que não se consegue ler. A literatura brasileira examinou cada instrumento separadamente, dos indicadores de gestão do TCU e sua relação com o desempenho discente (Barbosa; Freire; Crisóstomo, 2011) à arquitetura avaliativa do Sinaes (Polidori;

Marinho-Araujo; Barreyro, 2006; Verhine; Dantas; Soares, 2006) e à avaliação de políticas públicas em geral (Trevisan; Van Bellen, 2008; Ceneviva; Farah, 2012). Não identificamos, na literatura examinada, exame do conjunto da arquitetura classificado segundo a distinção entre atividade, produto e resultado, nem teste do pressuposto de que produzir informação equivale a disponibilizá-la.

1.3. Pergunta, Objetivos e Natureza da Contribuição

O artigo pergunta por quais processos institucionais a arquitetura pública de mensuração das universidades federais brasileiras preserva ou torna menos legível a identidade categorial dos resultados quando novas camadas normativas se sobrepõem às anteriores. Responder a essa pergunta exige uma operação empírica auxiliar: classificar os indicadores públicos segundo a distinção entre atividade, produto e resultado, por protocolo ancorado nas fórmulas e ponderações oficiais. O protocolo permite ao leitor auditar cada classificação e sua distribuição por categoria, nível e camada temporal. Com base nessa classificação, o mecanismo que explica a distribuição observada é reconstruído por rastreamento de processo, com condições de contorno especificadas, confronto com explicações rivais e critérios de falseabilidade.

O estudo busca uma explicação processual: reconstruir documentalmente o mecanismo por trás de uma trajetória institucional singular, sem estimar efeitos médios nem testar covariação entre casos. A validação segue o padrão do rastreamento de processo (George; Bennett, 2005), detalhado na seção 3.

A contribuição principal é teórica. O artigo identifica uma falha intermediária entre a produção e o uso da informação de

desempenho, a perda de legibilidade categorial, e propõe um mecanismo testável para explicá-la. As evidências mostram que o Estado produz informação de resultado, mas que, no nível institucional, onde a prestação de contas formalmente se exerce, nenhum indicador oficial de resultado puro chega a existir: o resultado comparece apenas fundido a insumos e processos dentro de índices compostos. Esse estado estrutural é aqui denominado encapsulamento dos resultados, formalmente definido na seção 2.5. O Brasil é utilizado como caso revelador. A contribuição consiste na formulação do mecanismo, de modo a permitir teste em outros setores cujas arquiteturas também se formam por camadas normativas sobrepostas, sob demanda regulatória por um número síntese. Sustentam essa contribuição um diagnóstico categorial integrado da arquitetura brasileira e um protocolo de classificação auditável e aplicável a outros setores. O objeto do artigo é a mensuração e o processo institucional que a produziu, não o desempenho das universidades em si.

2. FUNDAMENTAÇÃO TEÓRICA E CONSTRUÇÃO DO GAP

2.1. O Que Sabemos: Gerencialismo, Auditoria e o Caso Brasileiro

O movimento gerencialista dos anos 1980, sintetizado por Hood (1991) em sete componentes doutrinários, incluía a definição de padrões explícitos de desempenho e a ênfase no controle por resultados, em substituição ao controle por procedimentos. Osborne e Gaebler (1992) popularizaram a fórmula segundo a qual o que não é medido não é gerenciado, e governos de matizes distintos incorporaram baterias de indicadores a seus instrumentos de gestão. Pollitt e Bouckaert (2011) mostram que a intensidade da mensuração variou entre países, mas nenhuma administração

relevante ficou imune ao movimento. Pollitt (2013) acrescenta que as lógicas de mensuração raramente são coerentes entre si dentro de um mesmo sistema, observação pertinente ao caso aqui analisado.

Power (1997) ofereceu a leitura crítica mais influente do processo. A explosão de auditorias decorre menos de demanda técnica por informação do que de uma transformação nas relações de confiança, e a auditoria verifica, com frequência, aquilo que é auditável, não aquilo que importa. Uma arquitetura de indicadores pode ser densa, regular e formalmente rigorosa e, ainda assim, deixar invisível a dimensão que justificaria sua existência.

No Brasil, dois marcos delimitam o terreno. O primeiro é constitucional. A Constituição de 1988 atribuiu ao controle externo, exercido pelo Congresso Nacional com auxílio do TCU, a fiscalização contábil, financeira, orçamentária, operacional e patrimonial da administração, quanto à legalidade, à legitimidade e à economicidade (Brasil, 1988, arts. 70 e 71). Esse mandato favorecia indicadores associados a recursos, operações e economicidade, e será decisivo para compreender a fisionomia dos indicadores que o TCU viria a exigir das universidades. O segundo é a orientação para resultados do Plano Diretor da Reforma do Aparelho do Estado, de 1995 (Bresser-Pereira, 1996), cuja trajetória foi menos linear do que o desenho original sugeria: Abrucio (2007) descreve avanços localizados convivendo com fragmentação, e Secchi (2009) mostra que os modelos gerencial e de governança se sobrepuseram ao paradigma burocrático sem substituí-lo, padrão de acumulação sem substituição que reaparecerá na seção 5.

A literatura nacional de avaliação registra padrão análogo. A avaliação foi incorporada mais como exigência formal do que como

prática integrada à decisão (Trevisan; Van Bellen, 2008), e o elo entre avaliação e responsabilização depende de condições institucionais raramente completas (Ceneviva; Farah, 2012). Este artigo acrescenta um elo anterior a essa cadeia: mesmo quando a informação existe, nada garante que ela permaneça legível como categoria própria ao longo do trajeto. Essa literatura ajuda a entender por que mensuração e auditabilidade se expandiram no país, mas não explica por que diferentes lógicas de mensuração seguem coexistindo sem integração dentro de um mesmo sistema.

2.2. Desempenho Como Construto e a Cadeia Mensurar, Incorporar e Usar

Desempenho, no setor público, é um construto multidimensional e disputado, que varia em profundidade (de insumos a efeitos finais) e em amplitude (dos microdesempenhos individuais aos macrodesempenhos sistêmicos) (Bouckaert; Halligan, 2008). A gestão do desempenho supõe três operações: mensurar, incorporar e usar (Van Dooren; Bouckaert; Halligan, 2015). Behn (2003) demonstra que gestores mensuram para finalidades distintas (avaliar, controlar, orçar, motivar, aprender) e que nenhum indicador serve igualmente a todas. Diante de um sistema de mensuração, o número de indicadores importa menos do que a pergunta sobre a que dimensões do desempenho eles efetivamente respondem.

Essa cadeia de operações incorpora um pressuposto. A informação mensurada chegaria íntegra às etapas de incorporação e uso, e a transição entre mensurar e usar seria problema de vontade política ou de capacidade organizacional, raramente de transformação da própria informação. Moynihan (2008) avança ao mostrar que a informação de desempenho é objeto interpretado e disputado pelos

atores; mesmo em sua formulação, porém, ela permanece reconhecível enquanto tal. De Bruijn (2002) acrescenta que sistemas maduros exigem multiplicidade de medidas e espaço interpretativo. Nenhuma dessas formulações oferece um mecanismo para a possibilidade de que, em arquiteturas construídas por camadas normativas sobrepostas, a transição entre mensurar e usar degrade a informação categorialmente, a ponto de um resultado deixar de ser legível como resultado. Essa degradação constitui o objeto analítico próprio deste estudo, distinto tanto da não produção quanto do mau uso.

A literatura documenta ainda os efeitos perversos da mensuração: miopia, subotimização, ossificação, manipulação (Smith, 1995; Propper; Wilson, 2003; Bevan; Hood, 2006). Esses riscos tornam a composição categorial de um sistema questão substantiva. Se o sistema mede sobretudo atividade, é para a atividade que o comportamento organizacional tenderá a se orientar (Bruijn, 2002; Moynihan, 2008).

2.3. Atividade, Produto e Resultado

A lente analítica do artigo é a cadeia lógica que estrutura os modelos de produção do valor público (McLaughlin; Jordan, 1999; Hatry, 2006; Talbot, 2010): insumos são mobilizados em atividades, atividades geram produtos, produtos contribuem para resultados.

Atividade designa a mobilização de insumos e a realização de operações. Trata-se do que a organização faz e com quais recursos, medido por matrículas, docentes, vagas e razões entre insumos, como custo por aluno ou alunos por professor. Seguindo Bouckaert e Halligan (2008), insumos e processos são tratados conjuntamente

sob essa categoria, pois ambos descrevem o funcionamento da organização, não suas consequências.

Produto (output) designa a entrega concluída, diretamente controlável pela organização. Na universidade, o produto paradigmático do ensino é o concluinte diplomado, contável no momento em que ocorre, sem defasagem, e sob controle quase integral da organização, que define os critérios de integralização e emite o diploma.

Resultado (outcome) designa a mudança de estado no público-alvo ou no ambiente, atribuível ao menos em parte à ação organizacional. Incluem-se aí aprendizagem demonstrada, valor agregado, permanência ou evasão como desfechos de trajetória e inserção profissional. Resultados são parcialmente controláveis, sofrem influência externa, manifestam-se com defasagem e comportam gradações entre intermediários e finais (Hatry, 2006; Van Dooren; Bouckaert; Halligan, 2015). A atribuição parcial não desqualifica a categoria, pois nenhum resultado de política pública é integralmente atribuível a uma única organização.

A distinção tem consequência prática direta. Um diploma emitido nada garante sobre a aprendizagem que certifica, e sistemas que colapsam essas categorias produzem leituras enganosas do desempenho (Boyne, 2002; Radin, 2006).

2.4. Indicadores, Índices Compostos e o Problema Conhecido da Agregação

Um indicador é uma representação simplificada e mensurável de uma dimensão do fenômeno de interesse, não o fenômeno em si (Jannuzzi, 2005). Dessa natureza decorrem duas propriedades: todo

indicador seleciona o que se torna visível, e indicadores são reativos, pois os atores respondem ao que é medido (Espeland; Sauder, 2007; Bevan; Hood, 2006).

Para evitar ambiguidade nas seções empíricas, adota-se uma taxonomia de quatro termos: indicador elementar (fórmula que computa uma única dimensão, como número de concluintes), indicador composto (agrega componentes heterogêneos mediante ponderações do produtor, como o Conceito Preliminar de Curso, CPC, e o Índice Geral de Cursos, IGC), componente (cada elemento ponderado dentro de um composto) e nível de agregação (a unidade em que o indicador é oficialmente calculado e divulgado: curso ou coorte, instituição, programa).

Que índices compostos percam informação é achado bem estabelecido. Eles ganham em síntese e comunicabilidade (Jannuzzi, 2005), mas perdem em discriminação: ao combinar insumos, processos e resultados, o leitor não sabe qual dimensão varia quando o número muda (Jannuzzi, 2005; Dill; Soo, 2005). As ponderações raramente têm justificativa teórica, e pequenas alterações de peso reordenam posições (Dill; Soo, 2005). O índice sintético serve à comunicação; o diagnóstico exige o retorno aos indicadores analíticos (Jannuzzi, 2005). Essa literatura trata o custo analítico da composição como propriedade do instrumento em si. Falta explicar sob que condições essa perda de discriminação passa a ser traço estrutural da própria arquitetura de accountability, isto é, quando o composto se torna a via oficial pela qual uma categoria inteira alcança determinado nível de prestação de contas. É nesse deslocamento do instrumento para a arquitetura que este artigo se diferencia dessa literatura.

2.5. O Que Não Sabemos: Legibilidade Categorical e Encapsulamento

Duas definições formais ajudam a preencher essa lacuna teórica.

Legibilidade categorial é uma propriedade observável da arquitetura de mensuração: a capacidade de identificar, no nível decisório relevante, a magnitude correspondente a uma categoria (aqui, resultado) por meio de indicador oficial, sem necessidade de reconstrução do composto. Legibilidade não se confunde com disponibilidade de dados. Os microdados abertos do Inep permitem que um analista habilitado reconstrua médias institucionais de Enade (Exame Nacional de Desempenho dos Estudantes) ou do Indicador de Diferença entre os Desempenhos Observado e Esperado (IDD); isso, porém, não integra a arquitetura oficial de accountability, composta pelos indicadores que constam do relatório de gestão ao TCU e dos atos regulatórios do MEC. A prestação de contas formal opera sobre indicadores oficiais, não sobre agregações *ad hoc* de terceiros. O sentido aqui adotado é distinto do de Scott (1998), para quem legibilidade designa a capacidade do Estado de ler a sociedade; aqui, trata-se da capacidade de quem exerce o controle de ler, na informação oficial, a categoria que ela mensura.

Encapsulamento de resultados é o estado da arquitetura de mensuração em que três condições se verificam simultaneamente: (1) a informação de resultado existe e é oficialmente produzida em um nível inferior de agregação; (2) essa informação é incorporada a um composto que combina componentes de categorias heterogêneas; (3) o composto constitui o veículo oficial pelo qual a categoria alcança o nível em que a accountability se exerce, sem que exista, nesse nível, indicador autônomo equivalente. Quando as três

condições se cumprem, a categoria deixa de funcionar como unidade independente de interpretação institucional, ainda que permaneça presente como componente interno do composto.

Três delimitações precisam o alcance do conceito. A primeira diz respeito à fronteira com a agregação, que admite um teste operacional: se a transformação muda apenas a escala, a categoria permanece identificável (médias institucionais de Enade continuariam a ser resultado); se a informação precisa atravessar categorias heterogêneas para chegar ao nível decisório, há encapsulamento. A segunda diz respeito ao que efetivamente se perde: não a recuperabilidade matemática dos componentes, formalmente possível por decomposição, mas a disponibilidade da categoria como unidade oficial de interpretação. A terceira afasta uma redução indevida. O conceito não se resume ao custo analítico já conhecido dos índices compostos, propriedade do instrumento; o encapsulamento é uma relação entre instrumento, categoria e nível de accountability. Dela decorre o critério de reversão: o encapsulamento cessa quando o nível de accountability passa a dispor de indicador autônomo da categoria, ainda que o composto continue a existir.

Dessas definições decorre a reformulação teórica proposta. A cadeia canônica carece de um elo intermediário: mensurar, preservar a legibilidade, incorporar e usar. O elo de preservação pode falhar de modo independente dos demais, e o caso brasileiro documenta um padrão consistente com essa falha.

Uma última precisão conceitual. Das múltiplas dimensões do termo accountability (answerability, enforcement, reporting), este artigo trata da prestação de contas formal e da avaliação pública no nível

institucional, os instrumentos pelos quais a universidade federal, como organização, responde ao controle externo e à regulação.

2.6. Educação Superior: O Caso Brasileiro Como Caso Revelador

Universidades são organizações de objetivos múltiplos, tecnologias incertas e resultados de longa maturação, e o "resultado" do ensino superior é plural (aprendizagem, diplomação, permanência, inserção, mobilidade), cada dimensão com custos, defasagens e problemas de atribuição próprios. Cave *et al.* (1997) documentam três décadas de tentativas britânicas de equilibrar indicadores de insumo, produto e resultado, concluindo que a tensão entre comparabilidade e validade nunca foi resolvida. Jongbloed e Vossensteyn (2001) mostram, em onze países da OCDE, que fórmulas de financiamento por desempenho pagam majoritariamente por produto, por ser mais contável, tempestivo e atribuível. Os rankings agravam o quadro ao comprimir dimensões heterogêneas em posições ordinais, com reatividade documentada (Hazelkorn, 2015; Espeland; Sauder, 2007).

No Brasil, o Exame Nacional de Cursos, o Provão, aplicado entre 1996 e 2003, foi o primeiro instrumento nacional de aferição direta do desempenho de concluintes. O Sinaes, instituído em 2004, substituiu-o por um sistema tripartite que articula avaliação institucional, avaliação de cursos e exame de desempenho estudantil, o Enade (Brasil, 2004b; Polidori; Marinho-Araujo; Barreyro, 2006; Verhine; Dantas; Soares, 2006). Essa trajetória revela uma assimetria que a seção 5 explora: dentro do fluxo avaliativo do MEC, um instrumento pôde substituir o outro, porque a competência normativa para revisar era do mesmo titular; entre fluxos de

competências distintas, como o controle externo do TCU e a avaliação do MEC, nenhuma substituição ocorreu.

O Plano Nacional de Educação ilustra o mesmo problema em outro plano. A Meta 12 combina metas de atividade (elevar as taxas de matrícula) com estratégias de ampliação de concluintes e elevação de qualidade (Brasil, 2014), cobradas por indicadores de natureza única, o que estende a pergunta deste artigo à política educacional.

2.7. Modelo Analítico

O modelo analítico é a cadeia atividade, produto e resultado, aplicada à definição operacional oficial de cada indicador. A unidade de análise na etapa classificatória é o indicador; na etapa explicativa, o objeto é a arquitetura institucional que organiza sua produção, agregação e publicação entre níveis. O atributo classificado é a dimensão que a fórmula efetivamente mensura. Admite-se uma quarta categoria, híbrido/ambíguo, para indicadores cuja fórmula combina componentes de mais de uma categoria. Forçar a classificação de casos genuinamente ambíguos destruiria a informação mais interessante: a existência e a estrutura dos híbridos são, elas próprias, um achado sobre a arquitetura e a chave do mecanismo que a produziu.

3. METODOLOGIA

3.1. Desenho e Enquadramento Epistemológico

Trata-se de pesquisa empírica, documental, de abordagem quantitativa-descritiva na etapa classificatória e explicativo-processual na etapa causal. O desenho foi escolhido porque a pergunta exige a reconstrução temporal de mecanismos

documentados: normas datadas, competências legais e fórmulas oficiais constituem exatamente o tipo de evidência de que o rastreamento de processo se alimenta. O material empírico é constituído exclusivamente por documentos e dados públicos: normas legais, decisões de órgãos de controle, notas técnicas e metodologias oficiais do Inep e do TCU. Os elementos empíricos utilizados (fórmulas, ponderações, datas, competências legais) são rastreáveis às fontes oficiais identificadas, condição necessária para que o protocolo seja auditável por terceiros.

3.2. Objeto e Universo

O objeto são os indicadores públicos de desempenho aplicáveis às universidades federais brasileiras. Os Institutos Federais não foram incluídos, pois seu sistema próprio (Plataforma Nilo Peçanha) responde a marco normativo distinto, o que comprometeria a comparabilidade do corpus. O universo é o conjunto de indicadores que satisfazem, cumulativamente, os critérios da seção 3.4.

3.3. Fontes de Dados

Quatro famílias de fontes compõem o corpus: as estatísticas do Censo da Educação Superior, coletadas anualmente pelo Inep e divulgadas em sinopses estatísticas (Brasil, 1996; Inep, 2023); os indicadores de gestão exigidos das Instituições Federais de Ensino Superior pela Decisão nº 408/2002 do Plenário do TCU, com metodologia consolidada em documento orientador conjunto TCU/Sesu/SFC (TCU, 2002; Brasil, 2004a), base formal da prestação de contas ao controle externo no período analisado; os indicadores de qualidade do Sinaes, instituído pela Lei nº 10.861/2004 (Conceito Enade, IDD, CPC e IGC, os dois últimos por atos normativos do MEC

em 2008, com fórmulas detalhadas pelo Inep) (Brasil, 2004b, 2008a, 2008b; Inep, 2020); e os indicadores de fluxo TAP, TDA e TCA, calculados pelo Inep a partir do acompanhamento longitudinal de coortes, com metodologia publicada em 2017 (Inep, 2017).

3.4. Critérios de Seleção e Completude do Corpus

Foram incluídos os indicadores que atendem simultaneamente a quatro critérios: publicidade (divulgação regular em fonte oficial de acesso aberto); abrangência nacional (cálculo padronizado para todas as universidades federais); recorrência (produção periódica, não pontual); e documentação metodológica (definição operacional oficial publicada).

Os critérios excluem, por construção: indicadores internos de planejamento (PDI), que violam publicidade e padronização; rankings privados e métricas bibliométricas sem padronização oficial nacional; os conceitos derivados de avaliação *in loco* (Conceito de Curso e Conceito Institucional), vinculados ao calendário de atos autorizativos de cada instituição, não a ciclo periódico comparável; as variantes anuais dos indicadores de fluxo, desagregações temporais das mesmas medidas acumuladas; e as desagregações das variáveis-síntese do Censo, que não constituem indicadores distintos. Aplicados os critérios, o corpus compreende todos os indicadores identificados que os satisfazem simultaneamente: não há, nesse recorte, indicador elegível fora da Tabela 1.

3.5. Protocolo de Classificação

O protocolo segue cinco regras.

1. Regra da definição operacional. A classificação considera a fórmula oficial do indicador, nunca seu nome. Um indicador denominado "taxa de sucesso" não é classificado como resultado por força do rótulo, mas pelo que sua fórmula computa.
2. Regra de atividade. Classifica-se como atividade o indicador que mensura volume de insumos (pessoas, recursos, custos) ou de operações (matrículas, vagas, cursos), inclusive razões entre componentes dessa natureza.
3. Regra de produto. Classifica-se como produto o indicador que mensura entregas concluídas sob controle direto da instituição (concluintes), inclusive quando normalizadas por denominador de atividade.
4. Regra de resultado. Classifica-se como resultado o indicador que mensura mudança de estado no público-alvo (aprendizagem demonstrada, valor agregado, desfecho definitivo de trajetória), ainda que a atribuição causal à instituição seja parcial, sem implicar atribuição causal integral nem validade ou suficiência como medida de efetividade institucional.
5. Regra de hibridismo. Classifica-se como híbrido/ambíguo o indicador cuja fórmula combina, com pesos relevantes, componentes de mais de uma categoria, ou cuja dimensão não se reduz a nenhuma delas. Quando um único componente domina inequivocamente a fórmula, prevalece a categoria desse componente.

Duas decisões de fronteira merecem justificativa. A primeira é o par conclusão/desistência. A conclusão é produto, pois constitui entrega institucional sob controle direto da organização (a diplomação); a desistência definitiva é resultado, pois constitui mudança de estado do estudante, parcialmente determinada por fatores externos. Os dois eventos pertencem à mesma coorte, mas ocupam posições distintas da cadeia lógica, e a seção 4.7 mostra que reclassificá-los não altera o achado central. A segunda envolve o Conceito Enade e o IDD. Sua classificação como resultado é ontológico-operacional, referida à dimensão que a fórmula visa mensurar, e não constitui validação psicométrica das medidas. As limitações conhecidas do Enade, como o caráter transversal do exame e o baixo engajamento de estudantes sem consequências individuais em jogo (Verhine; Dantas; Soares, 2006), afetam a validade da mensuração, não sua categoria; como mostra a seção 4.7, o eventual rebaixamento dessas medidas agravaria o diagnóstico em vez de enfraquecê-lo.

3.6. Tratamento e Análise

Os indicadores foram tabulados e analisados por contagens e proporções em três cortes: composição geral por categoria, distribuição por nível de análise e distribuição por camada temporal de criação. O nível de análise segue regra única: considera-se o nível de agregação em que o indicador é oficialmente calculado e divulgado pela fonte responsável (curso ou coorte; instituição; programa). Dado o tamanho do universo ($n = 23$), as contagens absolutas são a medida primária e os percentuais têm função ilustrativa. Não se empregam técnicas inferenciais, pois o universo é censitário em relação aos critérios de seleção definidos na seção 3.4 — não um universo absoluto de todos os indicadores imagináveis — e a pergunta é composicional.

3.7. Validação e Reprodutibilidade

A classificação ocorreu em duas rodadas independentes, com intervalo temporal, e as discrepâncias foram resolvidas por retorno à documentação oficial. Não houve segundo codificador independente. O protocolo garante, portanto, reprodutibilidade procedimental (regras públicas, aplicáveis por terceiros, com todos os casos ambíguos individualmente reportados), mas não oferece medida de confiabilidade intercodificadores no sentido estatístico. Três medidas mitigam essa limitação: publicidade integral das regras, justificativa individual dos casos de fronteira (seções 3.5, 4.6 e 4.7) e uma análise de sensibilidade classificatória (seção 4.7) que mostra que o achado central permanece válido sob as reclassificações plausíveis ali examinadas. Essa análise de sensibilidade, porém, opera dentro do mesmo protocolo interpretativo do codificador único: ela testa a robustez do achado a reclassificações alternativas dos casos de fronteira já identificados por esse protocolo, não a interpretações divergentes que um codificador externo pudesse aplicar às portarias do MEC/Inep ou aos acórdãos do TCU desde a primeira leitura. Uma eventual auditoria independente do protocolo deve ser tratada como teste de robustez da classificação, não como validação do mecanismo.

3.8. Rastreamento de Processo: Lógica Probatória

A explicação desenvolvida na seção 5 segue a lógica do rastreamento de processo: a reconstrução de uma sequência de eventos documentados como evidência de um mecanismo, e não a inferência por variância entre casos (George; Bennett, 2005; Mahoney, 2000). A escolha decorre da natureza do objeto, já que há uma única trajetória de sedimentação normativa a explicar, não

uma amostra de casos independentes. O padrão probatório mobiliza cinco tipos de evidência, cada qual vinculado a uma etapa do mecanismo: sequência temporal (ordem datada dos atos normativos), autorização institucional (competência legal de cada produtor de indicador), não revogação (ausência documentada de revogação entre camadas), exigência de agregação (norma que demanda indicador síntese) e persistência (reprodução do padrão em camadas posteriores). A seção 5.2 apresenta, em tabela, a correspondência entre cada etapa, sua expectativa observável, a evidência encontrada e a explicação rival que ela qualifica. O critério de validação não é significância estatística, e sim a consistência entre mecanismo e evidência documental, o maior poder explicativo relativo diante das alternativas (seção 5.4) e a capacidade de gerar previsões falseáveis (seção 5.5).

3.9. Data de Referência

O corpus reflete a arquitetura normativa considerada vigente no corpus até dezembro de 2025. No âmbito do controle externo, o Acórdão nº 461/2022-TCU-Plenário (Processo TC 026.147/2020-3, Rel. Min. Walton Alencar Rodrigues), resultante de auditoria operacional da Secretaria de Controle Externo da Educação, da Cultura e do Desporto, constatou lacunas de informações e de indicadores no modelo então vigente e determinou à Sesu/MEC a apresentação de plano de ação para a efetiva implementação de indicadores de gestão e desempenho das Ifes. O acórdão é utilizado como atualização institucional e delimitação temporal do alcance dos achados, não como unidade do corpus histórico empregada na classificação da seção 4. Até a data de referência deste estudo, não foi identificada alteração formal do conjunto de nove indicadores

identificado neste estudo como exigido pela Decisão nº 408/2002 e pelos Acórdãos nº 1.043/2006 e nº 2.167/2006 no corpus examinado.

A título de informação adicional, posterior à data de referência do corpus e por isso não incorporada como unidade analítica: a página temática do TCU sobre o assunto (TCU, 2026), atualizada em 13 de julho de 2026, indica que a implementação dessa determinação permanece sob acompanhamento, listando dois processos de monitoramento em curso (TC 002.712/2024-5 e TC 004.273/2025-7); não foi localizado, nas bases e documentos oficiais consultados até o fechamento deste estudo, acórdão de mérito que encerre esse monitoramento ou institua formalmente um novo modelo de indicadores que substituísse o conjunto examinado. O marco avaliativo da educação superior é, adicionalmente, objeto de propostas de reformulação em debate; alterações posteriores não estão contempladas, e a seção 5.5 indica de que modo funcionariam como testes das proposições, não como refutações do desenho.

4. RESULTADOS

4.1. Composição Geral do Corpus

A Tabela 1 apresenta os 23 indicadores identificados segundo os critérios de inclusão do protocolo (seção 3.4), sua fonte, o nível de análise e a classificação atribuída.

Tabela 1. Corpus de indicadores públicos, nível de análise e classificação segundo o protocolo

Indicador	Fonte	Nível	Classificação
-----------	-------	-------	---------------

Matrículas em graduação	Censo/Inep	Instituição	Atividade
Ingressantes	Censo/Inep	Instituição	Atividade
Vagas ofertadas	Censo/Inep	Instituição	Atividade
Docentes em exercício	Censo/Inep	Instituição	Atividade (insumo)
Cursos de graduação	Censo/Inep	Instituição	Atividade
Inscritos / relação candidato-vaga	Censo/Inep	Instituição	Híbrido/ambíguo
Concluintes	Censo/Inep	Instituição	Produto
Custo corrente / aluno equivalente	TCU 408/2002	Instituição	Atividade (insumo)
Aluno tempo integral / professor equivalente	TCU 408/2002	Instituição	Atividade
Aluno tempo integral / funcionário equivalente	TCU 408/2002	Instituição	Atividade
Funcionário equivalente / professor equivalente	TCU 408/2002	Instituição	Atividade (insumo)
Grau de Participação Estudantil (GPE)	TCU 408/2002	Instituição	Atividade
Grau de Envolvimento com Pós-Graduação (GEPG)	TCU 408/2002	Instituição	Atividade
Índice de Qualificação do Corpo Docente (IQCD)	TCU 408/2002	Instituição	Atividade (insumo)
Conceito Capes/MEC	TCU 408/2002	Programa	Híbrido/ambíguo
Taxa de Sucesso na Graduação (TSG)	TCU 408/2002	Instituição	Produto

Conceito Enade	Sinaes/Inep	Curso	Resultado
IDD	Sinaes/Inep	Curso	Resultado
CPC	Sinaes/Inep	Curso	Híbrido/ambíguo (composto)
IGC	Sinaes/Inep	Instituição	Híbrido/ambíguo (composto)
Taxa de Permanência (TAP)	Fluxo/Inep	Coorte/curs o	Híbrido/ambíguo
Taxa de Desistência Acumulada (TDA)	Fluxo/Inep	Coorte/curs o	Resultado
Taxa de Conclusão Acumulada (TCA)	Fluxo/Inep	Coorte/curs o	Produto

Fonte: elaboração própria, com base em Inep (2017, 2020, 2023), TCU (2002) e Brasil (2004a, 2004b).

Na taxonomia da seção 2.4, o corpus contém vinte indicadores de fórmula elementar, dois dos quais categorialmente ambíguos (relação candidato-vaga e TAP), dois compostos construídos no próprio sistema (CPC, IGC) e um composto importado de outro sistema avaliativo (Conceito Capes).

4.2. Anatomia das Definições Operacionais

As fórmulas que sustentam cada classificação são abertas a seguir, de modo que o leitor possa auditar o protocolo indicador a indicador.

Os nove indicadores da Decisão TCU nº 408/2002 são definidos por fórmula explícita no documento orientador conjunto (Brasil, 2004a). O Custo Corrente por Aluno Equivalente, as três razões de escala

(Aluno Tempo Integral por Professor Equivalente, por Funcionário Equivalente, e Funcionário Equivalente por Professor Equivalente), o Grau de Participação Estudantil e o Grau de Envolvimento com a Pós-Graduação medem volume e intensidade de uso de insumos e recursos: nenhum componente mensura entrega ou efeito. O Índice de Qualificação do Corpo Docente, $IQCD = (5D + 3M + 2E + G) / (D + M + E + G)$, é indicador puro de insumo. O Conceito Capes/MEC, média dos conceitos da avaliação quadrienal da Capes, é híbrido (seção 4.6). A Taxa de Sucesso na Graduação, razão entre diplomados e ingressantes fixada pela duração padrão dos cursos, computa diplomas normalizados pelo fluxo de entrada: apesar do nome, é produto (regra 3). Os Acórdãos nº 1.043/2006 e nº 2.167/2006 do Plenário do TCU desdobraram três desses componentes (Custo Corrente, Aluno Tempo Integral/Funcionário Equivalente e Funcionário Equivalente/Professor Equivalente) em versões com e sem as atividades dos hospitais universitários, sem alterar o número de indicadores nem sua natureza: as variantes hospitalares permanecem razões insumo/atividade (TCU, 2006a, 2006b).

O bloco do TCU contém, portanto, sete indicadores cuja fórmula relaciona exclusivamente insumos e operações entre si, um indicador de produto e um híbrido importado. Essa homogeneidade decorre do mandato constitucional do controle externo já discutido na seção 2.1.

Nos indicadores do Sinaes, o Conceito Enade deriva do desempenho dos concluintes no Exame Nacional de Desempenho dos Estudantes, aplicado por curso em ciclos trienais, com nota padronizada convertida em conceito de 1 a 5 (Brasil, 2004b; Inep, 2020). A fórmula computa aprendizagem demonstrada, mudança de estado no público-alvo, classificada como resultado. O IDD estima

a diferença entre o desempenho observado dos concluintes no Enade e o desempenho esperado dado o perfil de entrada, aferido pelas notas no Enem (Inep, 2020). É a medida oficial mais próxima do conceito de valor agregado: isola, ainda que imperfeitamente, a contribuição própria do curso, e também se classifica como resultado.

O CPC, instituído em 2008 (Brasil, 2008a), é o caso decisivo do argumento, e sua estrutura ponderal merece exposição integral. Conforme a metodologia oficial (Inep, 2020), o CPC combina oito componentes em três blocos, com os pesos da Tabela 2.

Tabela 2. Decomposição ponderal do Conceito Preliminar de Curso (CPC)

Bloco e componente	Peso	Categoria do componente
Desempenho dos estudantes: nota dos concluintes no Enade	20,0%	Resultado
Desempenho dos estudantes: nota do IDD	35,0%	Resultado
Corpo docente: proporção de doutores	15,0%	Insumo (atividade)
Corpo docente: proporção de mestres	7,5%	Insumo (atividade)
Corpo docente: regime de trabalho	7,5%	Insumo (atividade)
Percepção discente: organização didático-pedagógica	7,5%	Processo (atividade)
Percepção discente: infraestrutura e instalações	5,0%	Processo (atividade)

Percepção discente: oportunidades de ampliação da formação	2,5%	Processo (atividade)
--	------	----------------------

Fonte: elaboração própria, com base em Inep (2020).

O enunciado preciso é o seguinte: 55% da ponderação formal do índice corresponde a componentes classificados como resultado (Enade e IDD). Pesos formais não equivalem, por si, a importância substantiva na variância observada do índice, mas o que a decomposição estabelece é estrutural: os 45% restantes, correspondentes a insumo docente e percepção discente sobre processo, podem variar, por atuação sobre titulação, regime de trabalho e percepção de infraestrutura, sem qualquer alteração nos componentes de resultado. Pela regra 5, o CPC é híbrido: nenhum componente domina inequivocamente a fórmula, e a variação do índice é categorialmente ambígua por construção.

O IGC, também de 2008 (Brasil, 2008b), agrega em um único número institucional a média dos CPC do último triênio, ponderada por matrículas, e os conceitos Capes da pós-graduação (Inep, 2020). O resultado não constitui dimensão autônoma do IGC: sua contribuição é mediada pela parcela de graduação incorporada via CPC, no máximo 55% dessa parcela, e diminui conforme cresce o peso da pós-graduação, já que o Conceito Capes é, ele próprio, um composto de outra natureza. A presença dos conceitos Capes no IGC não implica ausência de dimensões substantivamente relacionadas a resultados na pós-graduação — a avaliação quadrienal pondera também produção intelectual e inserção social dos egressos, componentes com conteúdo de resultado; o ponto analítico é que tais dimensões não aparecem no IGC como uma categoria institucional autônoma e isolável, e sim já fundidas, dentro do

próprio Conceito Capes, a insumos e processos de outra natureza. No âmbito dos indicadores oficiais de qualidade examinados neste estudo, o IGC é o indicador calculado no nível institucional, no qual o resultado não é isolável.

Os indicadores de fluxo, segundo a metodologia do Inep (2017), acompanham coortes de ingressantes ano a ano, classificando cada estudante em permanência, conclusão ou desistência definitiva, donde derivam TAP, TCA e TDA por coorte de curso em horizonte plurianual. A TCA computa entregas concluídas (produto); a TDA computa desfecho definitivo da trajetória (resultado); a TAP mede estado intermediário, híbrido (seção 4.6). É a camada mais recente da arquitetura e a primeira a mensurar desfechos longitudinais, mas seus três indicadores são calculados e divulgados apenas no nível da coorte de curso, sem versão institucional oficial.

4.3. Distribuição Entre Atividade, Produto e Resultado

A Tabela 3 sintetiza a composição categorial do corpus.

Tabela 3. Composição da arquitetura de mensuração por categoria

Categoria	Nº de indicadores	Proporção aproximada
Atividade (incl. insumos)	12	52%
Produto	3	13%
Resultado	3	13%
Híbrido/ambíguo	5	22%
Total	23	100%

Fonte: elaboração própria. Percentuais arredondados.

Mais da metade dos indicadores mede atividade. A arquitetura descreve com riqueza o que as universidades fazem e com quais recursos, mas é rarefeita quanto ao que entregam e, sobretudo, quanto ao que muda em razão de sua ação. Produtos e resultados, somados, respondem por 6 dos 23 indicadores. Os produtos concentram-se na diplomação; os resultados puros restringem-se a duas medidas de aprendizagem (Enade, IDD) e uma de desfecho adverso de trajetória (TDA). Cinco indicadores são híbridos, e é neles que reside parte substancial da informação sobre resultados (seção 4.6). O achado composicional relevante, contudo, não é a proporção de cada categoria, e sim sua distribuição por nível, apresentada a seguir.

4.4. Distribuição por Nível de Análise

A distribuição categorial não é uniforme entre os níveis, como mostra a Tabela 4.

Tabela 4. Categorias por nível de análise do indicador

Nível de análise	Atividade	Produto	Resultado	Híbrido
Curso / coorte	0	1	3	2
Instituição	12	2	0	2
Programa (pós-graduação)	0	0	0	1

Fonte: elaboração própria.

O contraste entre três e zero constitui o teste descritivo central do artigo. Todos os indicadores classificados como resultado puro (Conceito Enade, IDD, TDA) são calculados no nível do curso ou da

coorte, enquanto no nível institucional, o da gestão universitária, da alocação orçamentária e da prestação de contas ao TCU, existem zero indicadores oficiais de resultado puro: dos 16 indicadores institucionais, 12 medem atividade, 2 medem produto e 2 são híbridos. "Zero" designa a inexistência, nesse nível, de indicador oficial que isole a categoria resultado, não a ausência absoluta de informação sobre resultados no sistema, que alcança a instituição apenas de forma encapsulada, como componente diluído do IGC.

Nenhum diagnóstico de falha de produção explica um zero que coexiste com um três; as seções seguintes referem-se a esse contraste sem rerepresentá-lo.

4.5. Evolução Temporal: As Camadas da Arquitetura

A composição atual não resultou de um desenho deliberado, e sim da sobreposição histórica de instrumentos criados em momentos distintos, com finalidades distintas. A Tabela 5 organiza essa estratificação.

Tabela 5. Camadas temporais da arquitetura de mensuração

Camada	Marco / período	Categoria predominante
1. Estatística censitária	LDB/1996; Censo consolidado nos anos 1990	Atividade
2. Controle de eficiência	Decisão TCU 408/2002	Atividade
3. Avaliação da qualidade	Sinaes/2004; CPC e IGC em 2008	Resultado, encapsulado em compostos

4. Trajetória estudantil	Indicadores de fluxo/Inep, 2017	Produto e resultado de trajetória
--------------------------	---------------------------------	-----------------------------------

Fonte: elaboração própria, com base em Brasil (1996, 2004b, 2008a, 2008b), TCU (2002) e Inep (2017).

Duas propriedades desse processo serão retomadas como evidência do mecanismo na seção 5. A primeira é que as camadas se acumulam sem se substituir. A Lei nº 10.861/2004 não revoga nem altera a Decisão TCU nº 408/2002, e a introdução dos indicadores de fluxo em 2017 não reformulou os instrumentos anteriores, já que nenhum dos quatro detém competência legal para redesenhar os demais. O contraste com a trajetória interna da avaliação é instrutivo: o Sinaes pôde substituir o Provão porque ambos pertenciam à mesma esfera de competência, e nenhuma substituição equivalente ocorreu entre esferas distintas. A segunda propriedade é um movimento direcional incompleto: as camadas recentes aproximam-se progressivamente do resultado, mas as camadas antigas, dominadas pela atividade, continuam sendo a base formal da prestação de contas institucional. Uma ressalva se impõe: parte das mudanças observáveis nas séries decorre de alterações de metodologia, e não de mudanças no fenômeno, razão pela qual esta análise se limita à composição dos instrumentos, sem inferir tendências de desempenho.

4.6. Os Híbridos e o Encapsulamento dos Resultados

Os cinco híbridos constituem a principal evidência empírica compatível com o mecanismo proposto, e não um resíduo classificatório.

O CPC combina, em um único número, resultados genuínos (55% da ponderação), insumos docentes (30%) e percepções de processo (15%), conforme a Tabela 2. O IGC agrega os CPC e os conceitos Capes, diluindo ainda mais a fração de resultado. O Conceito Capes, por sua vez, resulta de avaliação quadrienal externa ao corpus do TCU, que já pondera insumos, produção e inserção antes de entrar na fórmula do indicador de gestão nº 7 da Decisão 408/2002. Nos três casos, a informação sobre resultado existe, mas está fundida a componentes de outras categorias e lógicas avaliativas, de modo que a variação do índice é categorialmente ilegível: um IGC que sobe pode refletir melhora de aprendizagem, contratação de doutores ou melhora de infraestrutura percebida, e o número oficial, sozinho, não permite distinguir.

Esses fatos satisfazem as três condições da definição da seção 2.5. O resultado é oficialmente mensurado no nível do curso (condição 1), é incorporado a compostos heterogêneos (condição 2), e esses compostos são o veículo oficial pelo qual a categoria alcança o nível institucional, sem indicador autônomo equivalente nesse nível (condição 3). O fenômeno difere da simples ausência de mensuração, já que o Estado efetivamente mede resultados; ao agregá-los sob as condições que a seção 5 especifica, torna-os indisponíveis como categoria autônoma no nível da accountability. Tampouco se trata de indisponibilidade de dados, pois os microdados que permitiriam decompor os índices são públicos, mas essa decomposição não integra nenhum instrumento oficial de prestação de contas institucional. Em suma, é a arquitetura oficial, não a falta de dados, que não lê resultado no nível da instituição.

A Taxa de Permanência é híbrida por razão distinta: mede um estado intermediário da trajetória, que não é entrega concluída nem

desfecho, e sim a suspensão de ambos. A relação candidato-vaga mensura demanda e pode ser lida como atividade (dimensionamento da operação seletiva) ou como resultado reputacional difuso, sem que a definição oficial permita arbitrar.

4.7. Sensibilidade Classificatória

Dado que a classificação envolve julgamento, cabe perguntar se o achado central sobreviveria a decisões diferentes nos casos contestáveis. A resposta é afirmativa. Diferentemente de uma proporção, que poderia oscilar com pequenas mudanças de critério, o achado central é um zero localizado, que decorre de dois elementos documentais verificáveis no corpus selecionado segundo os critérios definidos: a composição do IGC, que funde categorias por fórmula oficial (Inep, 2020), e a inexistência de qualquer outro indicador institucional oficial que isole aprendizagem, valor agregado ou desfecho de trajetória.

Nas reclassificações plausíveis, o achado resiste. Tratar a TDA como produto deixa dois resultados puros, ambos de nível curso, com o zero institucional intacto. Rebaixar Conceito Enade e IDD por dúvidas psicométricas deixaria ainda menos resultado no sistema, agravando o diagnóstico. Classificar a relação candidato-vaga como atividade apenas aumenta a predominância dessa categoria. Forçar a TAP para resultado a mantém no nível coorte/curso. Mesmo reclassificar a TSG como resultado, contra a evidência literal de sua fórmula, embora eliminasse formalmente o zero institucional, colapsaria a distinção entre travessia e transformação que a literatura estabelece com precisão (Hatry, 2006; Boyne, 2002), sem que qualquer indicador institucional passasse a isolar aprendizagem, valor agregado ou desfecho. Esses testes mostram

que o achado não depende das escolhas de fronteira do protocolo, que apenas descreve e localiza o padrão de encapsulamento produzido pela arquitetura oficial.

No conjunto, a arquitetura é densa em atividade, rarefeita em resultado, e os resultados que produz aparecem majoritariamente encapsulados em índices compostos no nível em que a prestação de contas se exerce.

5. DISCUSSÃO

5.1. O Padrão a Explicar e a Tensão Que Ele Instala

O caso revela uma possibilidade ausente da cadeia convencional da gestão do desempenho: informação pode sobreviver à produção e falhar na transição categorial para o nível de accountability. A resposta empírica à pergunta de pesquisa é direta. A arquitetura distingue apenas parcialmente atividades, produtos e resultados, privilegia a primeira categoria e encapsula a terceira. Convém reiterar, antes de prosseguir, a natureza exata dessa terceira categoria: "encapsular" não equivale a "não produzir" nem a "não possuir o dado" — o resultado é produzido e está matematicamente recuperável dentro do IGC (seção 2.5) —, mas a uma perda de legibilidade categorial: o resultado deixa de existir como unidade autônoma de interpretação no nível em que a accountability se exerce, ainda que continue presente, diluído, como componente ponderado do composto. O que mais exige explicação é a coexistência documentada na Tabela 4: os dois diagnósticos consolidados para a escassez de informação de resultado (falha de produção e falha de uso) pressupõem que produzir informação equivale a disponibilizá-la, e o caso brasileiro viola esse pressuposto.

A contribuição teórica está em explicar essa violação de forma testável e generalizável.

5.2. O Mecanismo: Sedimentação Normativa e o Paradoxo da Agregação

A composição observada não decorre de escolha deliberada por medir atividade. A evidência é mais consistente com um processo que a literatura institucionalista denomina layering: mudança por sobreposição de novas regras às antigas, sem revogação, quando os atores que introduzem a mudança carecem de autoridade para revisar o arranjo preexistente (Streeck; Thelen, 2005; Mahoney; Thelen, 2010). Os conceitos mobilizados operam em níveis distintos e sucessivos da mesma cadeia: o layering é a condição institucional; a demanda por número síntese é a condição de ativação; o paradoxo da agregação, formulado adiante, é a transformação intermediária; o encapsulamento é o estado estrutural resultante; os retornos crescentes de trajetória (Pierson, 2000) são o mecanismo de persistência. O mecanismo é, ademais, emergente: não requer intenção coordenada de obscurecimento e resulta da interação de decisões locais tomadas sob competências institucionais distintas, cada qual racional dentro de seu mandato.

O mecanismo decompõe-se em cinco etapas, cada uma vinculada a expectativa observável e evidência documental datada, conforme o padrão probatório da seção 3.8. A Tabela 6 confronta, etapa por etapa, o que o mecanismo prevê com o que a documentação permite observar e com aquilo que as explicações rivais deveriam igualmente explicar.

Tabela 6. Etapas do mecanismo, expectativas observáveis e evidência documental

Etapas do mecanismo	Expectativa observável	Evidência documental	Rival qualificada
1. Base sedimentada de atividade: as camadas fundadoras respondem a mandatos de controle de insumos	Fórmulas fundadoras relacionando apenas insumos e operações	7 das 9 fórmulas da Decisão TCU 408/2002; variáveis do Censo (Brasil, 2004a; Inep, 2023); mandato constitucional da economicidade (Brasil, 1988)	Custo técnico (explica a base, não o que se segue)
2. Introdução de categoria nova por titular sem autoridade sobre a base	Novo instrumento de resultado criado sem revogação dos anteriores	Lei 10.861/2004 cria o Sinaes; nenhum dispositivo revoga ou altera a Decisão 408/2002	Isomorfismo (não explica a coexistência sem substituição)
3. Não revogação entre esferas de competência distintas	Substituição dentro da mesma esfera; acumulação entre esferas	O Sinaes substitui o Provão (mesma esfera, MEC); nada substitui os indicadores do TCU (esferas distintas)	Todas as rivais (nenhuma prevê essa assimetria)
4. Demanda regulatória por número síntese único	Norma exigindo indicador único por curso e instituição, anterior aos compostos	Portaria Normativa MEC 40/2007 vincula indicadores ao fluxo regulatório; PN 4/2008 usa o CPC para dispensar visita <i>in loco</i> (CPC $\geq$ 3); CPC (ago/2008) e IGC (set/2008) seguem-se à exigência	Ambiguidade estratégica (a demanda é regulatória e datada, não oportunista)

5. Persistência e reprodução do padrão	Camadas novas coexistem sem integração arquitetural plena ou substituição dos instrumentos anteriores; fórmulas estáveis	Indicadores de fluxo (2017) não reformulam nem são integrados aos anteriores; estrutura de blocos do CPC/IGC estável desde 2008	Ambiguidade estratégica (estabilidade estrutural é incompatível com manipulação contínua)
--	--	---	---

Fonte: elaboração própria.

O elo central é a conjunção das etapas 3 e 4. A expressão "não havia autoridade" comporta três sentidos: competência normativa (poder legal de editar ou revogar a norma alheia), autoridade de coordenação (capacidade de articular titulares distintos em torno de um redesenho conjunto) e capacidade política de revisão (força para sustentar o redesenho contra resistências). O Inep e o MEC não detinham a primeira sobre a Decisão do TCU, não havia instância com a segunda, e não há evidência documental de mobilização institucional para uma revisão conjunta do sistema que testasse a terceira. Nessas condições, diante da exigência regulatória de um número único, a via institucionalmente disponível para que a informação de qualidade alcançasse o nível institucional foi agregá-la a índices capazes de coexistir com os instrumentos sedimentados.

Uma alternativa institucionalmente concebível seria publicar no nível institucional, como indicadores oficiais autônomos, as agregações de Enade e IDD, como o Inep já publica médias por curso. Essa alternativa não atendia à demanda regulatória por número único nem dispunha de patrocinador com autoridade para reordenar o conjunto, e não foi adotada; nasceram em seu lugar o CPC e o IGC (Brasil, 2008a, 2008b).

O índice composto foi o veículo oficial efetivamente adotado para levar o resultado até o nível institucional, e é precisamente sua composição que dissolve a identidade categorial daquilo que transporta. Esse é o paradoxo da agregação. Convém precisar o estatuto causal dessa escolha: o CPC e o IGC não emergiram de um acidente de acumulação passiva, e sim de uma decisão regulatória deliberada — simplificar a comunicação e a supervisão por meio de um número único (etapa 4, Tabela 6).

A sedimentação normativa não explica essa decisão em si, tomada de forma intencional pelo Inep/MEC; explica por que a via escolhida para operacionalizá-la foi a composição heterogênea, em vez da publicação autônoma das médias de Enade e IDD já calculadas pelo próprio Inep: a primeira via coexistia sem atrito com os instrumentos sedimentados do TCU, e não há evidência documental de que algum ator dispusesse de autoridade para adotar a segunda. O trade-off entre legibilidade sintética e especificidade de resultado foi, portanto, deliberado quanto ao fim (um índice único), mas condicionado pela sedimentação quanto ao meio (um composto, e não um indicador autônomo). A Tabela 2 quantifica a perda de decomponibilidade: a informação de resultado entra no CPC com 55% da ponderação no nível do curso e, ao ser incorporada ao IGC junto aos conceitos Capes, deixa de constituir categoria autônoma no nível institucional. O padrão tende a persistir pela lógica de retornos crescentes (Pierson, 2000): cada nova camada eleva o custo de reversão das anteriores, o que ajuda a explicar por que a camada de 2017 reproduziu a coexistência sem integração arquitetural plena, em vez de ensejar reforma do conjunto.

5.3. Condições de Contorno e o Contraste Within-Case

O mecanismo não opera de modo uniforme; sua ativação depende de duas condições especificáveis a partir do próprio corpus.

A primeira é a mistura de lógicas avaliativas. Dentro da própria Decisão TCU 408/2002, oito dos nove indicadores são razões entre componentes homogêneos e não geram hibridismo; apenas o Conceito Capes é híbrido, precisamente por importar avaliação externa que já mistura dimensões em sua origem. O encapsulamento não é propriedade automática de qualquer coexistência entre camadas: surge quando uma camada introduz ou importa uma dimensão avaliativa ausente do conjunto preexistente.

A segunda é a demanda externa por número síntese único no nível agregado. O CPC e o IGC foram criados sob exigência regulatória de indicador único por curso e por instituição para fins de supervisão e credenciamento (Brasil, 2007/2010, 2008a), demanda que não tolera múltiplos números paralelos. Por contraste, quando a nova categoria pode permanecer disponível como indicador autônomo, ou quando existe autoridade efetiva para reconfigurar os instrumentos herdados, o mecanismo não prevê encapsulamento como consequência necessária.

Por que 2008 produziu encapsulamento e 2017 não? Os indicadores de fluxo, introduzidos em 2017, carecem de demanda regulatória equivalente; por isso permanecem simplesmente ausentes do nível institucional, e não encapsulados. Essa bifurcação tem valor probatório, pois as duas camadas surgiram após 2004, sob o mesmo ambiente institucional, e o contraste decorre da aplicação uniforme dos critérios de seleção do corpus, não de escolha retrospectiva por conveniência. Dela deriva uma previsão testável: se surgir exigência regulatória de um índice único de trajetória por instituição, o

mecanismo prevê compressão dos indicadores de fluxo em novo composto, reproduzindo o padrão CPC/IGC, e não sua publicação paralela e decomposta.

Uma limitação decorre diretamente dessa explicação: a bifurcação 2008/2017 repousa, no desenho aqui empregado, substancialmente sobre uma única condição diferenciadora identificada no corpus — a existência ou ausência de demanda regulatória por número síntese —, o que reduz o poder discriminante do mecanismo caso essa condição não se sustente em outros contextos. O caso único não permite estimar a robustez estatística dessa dependência; o que o desenho oferece, em vez disso, é a previsão testável formulada no parágrafo anterior e formalizada na Proposição 2 (seção 5.5), cuja eventual não confirmação — compressão dos indicadores de fluxo sem que surja nova demanda regulatória, ou ausência de compressão mesmo sob demanda regulatória equivalente — fragilizaria diretamente a explicação. A agenda de pesquisa (seção 6.3) propõe exatamente esse teste comparativo em outros setores sedimentados por camadas.

5.4. Explicações Rivais

Três explicações alternativas poderiam dar conta, ao menos parcialmente, do padrão. A Tabela 7 as compara ao mecanismo proposto.

Tabela 7. Explicações rivais e poder explicativo comparado

Explicação	O que explica bem	O que não explica
------------	-------------------	-------------------

Custo técnico de mensurar resultado (Hatry, 2006; Van Dooren; Bouckaert; Halligan, 2015)	Escassez geral de indicadores de resultado	Por que resultados produzidos foram fundidos a insumos, em vez de publicados à parte
Ambiguidade estratégica (Bevan; Hood, 2006; Smith, 1995)	Conveniência política de índices que obscurecem dimensões desfavoráveis	Estabilidade estrutural dos blocos do CPC/IGC desde 2008; origem regulatória datada da demanda por número único
Isomorfismo institucional (DiMaggio; Powell, 1983; Hazelkorn, 2015)	Adoção do formato de índice composto, à semelhança de rankings internacionais	O momento exato (2008) e a inclusão específica de componentes herdados dentro dos compostos

Fonte: elaboração própria.

As três rivais captam, cada uma, um aspecto real do fenômeno, e nenhuma deve ser descartada como falsa. A ambiguidade estratégica, em particular, funciona como rival parcialmente compatível: é plausível que a opacidade dos compostos tenha sido politicamente conveniente para alguns atores em alguns momentos.

O custo técnico também não é integralmente afastado: o argumento deste artigo assume que, por o dado primário já existir no nível do curso, o custo marginal de consolidá-lo no nível institucional seria baixo, mas essa consolidação envolve esforço metodológico real de padronização e ponderação entre cursos heterogêneos — o próprio desenho do CPC, com oito componentes e pesos diferenciados (Tabela 2), é evidência desse esforço. Some-se a isso que o ciclo trienal do Enade distribui a cobertura avaliativa ao

longo de três anos por área, de modo que uma eventual média institucional apurada num único ano não representaria necessariamente a instituição inteira, exigindo tratamento metodológico de janela temporal, comparabilidade entre ciclos e cursos não avaliados no período — complicações adicionais que o argumento de custo baixo, tal como formulado, não elimina. O custo técnico de agregar, portanto, não é nulo, ainda que seja categoricamente distinto do custo técnico de produzir a informação de resultado em si (mensurar aprendizagem, valor agregado); é esse segundo custo, mais alto e mais bem documentado pela literatura (Hatry, 2006), que a rival de custo técnico explica satisfatoriamente, enquanto o primeiro permanece uma alternativa parcialmente compatível para o padrão de agregação observado. O argumento deste artigo não é de exclusão, e sim de maior poder explicativo relativo dentro deste desenho documental, por acomodar mais evidências observáveis com menos pressupostos.

O teste discriminante é a variação interna ao caso descrita na seção 5.3. As três rivais, na forma operacionalizada neste estudo, tenderiam a prever desfechos homogêneos para as camadas pós-2004; a evidência mostra desfechos distintos, com o Sinaes encapsulado e o fluxo deixado ausente. A conjunção entre sedimentação e demanda por número único, presente em 2008 e ausente em 2017, é a explicação examinada que oferece maior poder explicativo relativo para essa bifurcação, além de acomodar a sequência temporal completa (criação do Sinaes em 2004, exigência regulatória em 2007, compressão em índice em 2008), a não revogação entre esferas e a persistência do padrão. A ambiguidade estratégica seria mais bem sustentada por evidência de deliberação política em torno das fórmulas, teste que este desenho documental não realiza; a

documentação disponível apenas mostra que essa motivação não é necessária para reproduzir a sequência observada.

5.5. Falseabilidade e Proposições

Convém separar a falseabilidade lógica do mecanismo dos testes empíricos que dependem de pesquisa futura. No plano lógico, o mecanismo seria refutado: (a) se se encontrasse, no corpus ou em setor análogo, caso de introdução de categoria ausente das camadas anteriores, sob demanda por número único, que não resultasse em compressão; (b) se as fórmulas do CPC ou do IGC houvessem sido reformuladas oportunisticamente em resposta a resultados desfavoráveis de instituições específicas, o que fortaleceria a rival da ambiguidade estratégica; (c) se a decomposição pública oficial dos componentes do IGC, caso ocorra sem revogação do índice, não eliminasse o encapsulamento informacional para as instituições afetadas, indicando que a ilegibilidade decorre de outra causa. Nenhum desses testes foi realizado neste estudo; as propostas de reformulação do marco avaliativo mencionadas na seção 3.9 funcionam como experimentos naturais em curso, cujo desfecho informará as proposições abaixo.

Proposição 1. Em arquiteturas estatísticas setoriais construídas por sedimentação normativa, isto é, por adição sucessiva de camadas sem revogação das anteriores, as camadas fundadoras orientadas ao controle de insumos e processos tenderão a permanecer sobrerrepresentadas na composição categorial após a adição de novas categorias.

Proposição 2. Quando uma nova categoria de mensuração é introduzida em uma arquitetura sem autoridade transversal de

revisão e sob demanda externa por um número síntese no nível de accountability, sua incorporação tenderá a ocorrer por composição heterogênea, produzindo encapsulamento; na ausência dessa demanda, a nova categoria tenderá a permanecer não agregada nesse nível.

Proposição 3. A publicação oficial autônoma da categoria no nível de accountability elimina o encapsulamento, independentemente da permanência do índice composto. A proposição refere-se à disponibilidade da categoria na arquitetura oficial; os efeitos comportamentais do desencapsulamento, isto é, se a informação decomposta será compreendida e utilizada, constituem questão empírica distinta.

5.6. O Erro Categorical: Atividade e Produto Não São Resultado

A predominância de indicadores de atividade tem implicação interpretativa direta. Relatórios de gestão, painéis públicos e prestações de contas tendem a narrar o desempenho universitário na gramática do volume, respondendo a perguntas legítimas de eficiência e transparência, mas deixando sem resposta a pergunta sobre os objetivos formativos e os efeitos substantivos da instituição. Quando indicadores de atividade ocupam o lugar discursivo dos resultados, produz-se a visão de túnel descrita pela literatura: a organização otimiza o que é medido, e o que é medido é a operação (Smith, 1995; Bevan; Hood, 2006). Power (1997) anteviu as raízes do fenômeno, já que a atividade é a dimensão mais auditável da vida organizacional, e sistemas orientados pela auditabilidade gravitam para ela. Nada disso implica juízo contrário à mensuração da atividade, indispensável ao planejamento; o risco é sua promoção semântica a evidência de efeito.

O erro simétrico ocorre com o produto. A TSG e a TCA informam quantos ingressantes se tornam concluintes, informação valiosa sobre capacidade de entrega, mas a diplomação certifica a travessia, não a transformação, e dois cursos com idênticas taxas de conclusão podem agregar valores formativos radicalmente distintos, como o contraste entre Conceito Enade e IDD evidencia (Verhine; Dantas; Soares, 2006; Inep, 2020). A tentação de tratar diplomas como resultado é reforçada por serem a métrica de entrega mais visível e cobrável, inclusive pela Meta 12 do PNE (Brasil, 2014), e a que sistemas internacionais de financiamento por desempenho historicamente privilegiaram, por sua contabilidade simples (Jongbloed; Vossensteyn, 2001). Diante de cada número, a pergunta corretiva é única: sua fórmula computa travessias ou transformações?

5.7. Consequências do Encapsulamento: O Que Está Demonstrado e o Que é Implicação

Dois planos devem ser distinguidos. No plano informacional, o efeito já está demonstrado pela evidência documental: a variação do índice composto é categorialmente ilegível por construção, o que degrada sua utilidade para diagnóstico e aprendizagem (Jannuzzi, 2005). No plano comportamental, o que se pode formular é implicação teórica plausível: índices compostos criam incentivos compatíveis com a possibilidade de melhora posicional por atuação sobre os componentes de insumo, mais baratos e controláveis, sem alteração correspondente nos componentes de resultado. A decomposição da Tabela 2 dimensiona esse espaço formal, correspondente aos 45% da ponderação do CPC que não integram os componentes de resultado. Trata-se de espaço formal de ponderação, não de incentivo efetivamente observado. A dinâmica

esperada é análoga à reatividade documentada na literatura de rankings (Espeland; Sauder, 2007; Dill; Soo, 2005; Hazelkorn, 2015), mas sua verificação no caso brasileiro constitui hipótese derivada, integrante da agenda, não resultado deste estudo.

O diagnóstico vai além do custo analítico já conhecido dos compostos (Jannuzzi, 2005; Dill; Soo, 2005) em dois pontos. Primeiro, desloca a explicação do sintoma para o mecanismo que o produz: a sedimentação sem competência de revisão mútua, associada à demanda regulatória por número único. Segundo, desloca a discussão do plano técnico para o teórico, ao identificar na cadeia mensurar, incorporar e usar um modo de falha que ela não prevê. O problema não está na existência de índices sintéticos, que cumprem função comunicativa legítima, e sim na ausência de publicação oficial paralela de seus componentes de resultado no nível institucional, exatamente a lacuna que a Proposição 3 identifica como reversível.

5.8. Implicações para Gestão e Accountability

Decorre logicamente do mecanismo que intervenções de produção (medir mais) não corrigem o déficit de legibilidade, enquanto intervenções de publicação (decompor o que já se mede) o corrigem no plano formal. Para gestores universitários, a leitura categorial qualifica o planejamento: metas de atividade, produto e resultado exigem estratégias e prazos distintos (McLaughlin; Jordan, 1999; Moynihan, 2008), e prometer "elevar o IGC" é prometer mover um composto cujos determinantes formais majoritários no curto prazo são insumos.

Para órgãos de controle, o mecanismo sugere uma intervenção testável pela Proposição 3: publicar como indicadores institucionais autônomos as agregações de Enade e IDD, ao lado dos nove indicadores da Decisão 408/2002, e a versão institucional decomposta dos indicadores de fluxo, medidas coerentes com a função da informação avaliativa na responsabilização democrática (Ceneviva; Farah, 2012). A sedimentação prevê que essas medidas enfrentarão resistência não por custo técnico, mas pela ausência de ator com incentivo para reordenar o conjunto. Para a sociedade, a distinção oferece uma chave de leitura crítica: números divulgados como "desempenho" devem ser interrogados quanto à categoria que efetivamente medem, como no debate orçamentário sobre o custo das universidades federais, que opõe custo a volume sem que a arquitetura oficial responda a que transformação esse custo se refere.

5.9. Alcance: Generalização Analítica Condicionada

A generalização proposta é analítica, não empírica: recai sobre o mecanismo e suas condições de contorno, não sobre a frequência do fenômeno em outros contextos. Setores construídos por sedimentação de camadas normativas com produtores distintos de informação são domínios de teste, não evidência confirmatória: a saúde (informação assistencial coexistindo com avaliações de desfecho), o Judiciário (movimentação processual coexistindo com medidas de efetividade) e a rede federal de educação profissional. A Proposição 1 prevê neles composições dominadas por atividade, e a Proposição 2 discrimina as condições de encapsulamento. Sua verificação é o teste mais direto do alcance da teoria, e o protocolo da seção 3.5, por operar sobre definições oficiais, é aplicável a esses contextos sem adaptação substancial.

6. CONCLUSÃO

6.1. Síntese dos Achados

Dos 23 indicadores públicos que acompanham as universidades federais brasileiras, 12 medem atividade, 3 medem produto, 3 medem resultado e 5 são híbridos. Nenhum indicador oficial de resultado puro é publicado no nível institucional, e o resultado alcança a instituição apenas encapsulado em índices compostos. A evidência documental é mais consistente com a conjunção entre sedimentação normativa e demanda regulatória por número único do que com falha de produção, ambiguidade estratégica ou isomorfismo. O déficit não está na produção da informação, que existe e é publicada, mas na etapa seguinte, quando ela passa a integrar índices compostos e deixa de ser reconhecível como categoria própria no nível em que a prestação de contas ocorre.

6.2. Implicações Teóricas e Práticas

O achado tensiona um pressuposto da literatura internacional de gestão do desempenho: a cadeia mensurar, incorporar e usar (Van Dooren; Bouckaert; Halligan, 2015) assume que a informação produzida permanece categorialmente íntegra até o uso. O estudo mostra que essa integridade pode falhar de modo independente e estrutural, e o mecanismo de sedimentação normativa, reconstruído por rastreamento de processo e confrontado com rivais, especifica as condições sob as quais esse padrão pode ocorrer. As condições de contorno e as proposições o tornam falseável.

Para a literatura brasileira, o artigo oferece um diagnóstico categorial integrado da arquitetura de mensuração das universidades federais e um protocolo replicável, ancorado em definições operacionais

oficiais, aplicável a outros setores. Para a prática, desloca a prescrição: onde o déficit é de legibilidade, a recomendação padrão de medir mais resultados é insuficiente; a intervenção eficaz é desencapsular o que já se mede.

6.3. Limitações e Agenda

As limitações organizam-se em três blocos, dos quais deriva a agenda. Quanto à validade da classificação: o corpus baseia-se exclusivamente em documentos públicos; a classificação, embora protocolada e testada por sensibilidade, foi feita por um único pesquisador, o que garante reprodutibilidade procedimental sem confiabilidade intercodificadores; e a classificação de Enade e IDD é ontológico-operacional, sem endosso da validade psicométrica dessas medidas. Quanto ao desenho causal: o mecanismo foi reconstruído sobre uma única trajetória, adequado ao objetivo processual, mas que pede teste comparativo, e a rival da ambiguidade estratégica permanece parcialmente compatível, por falta de evidência sobre deliberação política interna às reformas. Quanto ao escopo: o estudo restringe-se às universidades federais e à arquitetura vigente até a data de referência; em particular, o Acórdão nº 461/2022-TCU-Plenário (Processo TC 026.147/2020-3), que já apontava lacunas de informações e de indicadores no modelo então vigente e determinou à Sesu/MEC plano de ação para novos indicadores, é tratado como parte do próprio fenômeno documentado pelo corpus, evidência adicional de que a arquitetura mede predominantemente atividade e carece de resultado legível, não como refutação do achado central, mas delimita o alcance temporal das proposições até que o eventual novo modelo seja formalmente instituído.

A agenda decorre das três proposições: replicar o protocolo, com codificadores independentes, em setores construídos por camadas (institutos federais, hospitais universitários, Judiciário), para a Proposição 1; investigar, por análise de atas e pareceres regulatórios, se a compressão em índices envolveu deliberação estratégica, para a Proposição 2; e acompanhar decisões que decomponham os componentes do IGC ou reformulem o marco avaliativo - incluindo o desdobramento do Acórdão nº 461/2022 -, verificando as implicações comportamentais da seção 5.7, para a Proposição 3.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRUCIO, F. L. Trajetória recente da gestão pública brasileira: um balanço crítico e a renovação da agenda de reformas. **Revista de Administração Pública**, v. 41, n. spe, p. 67-86, 2007. DOI: <https://doi.org/10.1590/S0034-76122007000700005>

BARBOSA, G. C.; FREIRE, F. S.; CRISÓSTOMO, V. L. Análise dos indicadores de gestão das IFES e o desempenho discente no ENADE. **Avaliação: Revista da Avaliação da Educação Superior**, v. 16, n. 2, p. 317-344, 2011. DOI: <https://doi.org/10.1590/S1414-40772011000200005>

BEHN, R. D. Why measure performance? Different purposes require different measures. **Public Administration Review**, v. 63, n. 5, p. 586-606, 2003. DOI: <https://doi.org/10.1111/1540-6210.00322>

BEVAN, G.; HOOD, C. What's measured is what matters: targets and gaming in the English public health care system. **Public Administration**, v. 84, n. 3, p. 517-538, 2006. DOI: <https://doi.org/10.1111/j.1467-9299.2006.00600.x>

BOUCKAERT, G.; HALLIGAN, J. **Managing performance: international comparisons**. London: Routledge, 2008. DOI: <https://doi.org/10.4324/9780203935958>

BOYNE, G. A. Concepts and indicators of local authority performance: an evaluation of the statutory frameworks in England and Wales. **Public Money & Management**, v. 22, n. 2, p. 17-24, 2002. DOI: <https://doi.org/10.1111/1467-9302.00303>

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 15 dez. 2025.

BRASIL. Lei nº 10.861, de 14 de abril de 2004. Institui o Sistema Nacional de Avaliação da Educação Superior, SINAES. Brasília, DF: Presidência da República, 2004b. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/l10.861.htm. Acesso em: 20 dez. 2025.

BRASIL. Lei nº 13.005, de 25 de junho de 2014. Aprova o Plano Nacional de Educação, PNE. Brasília, DF: Presidência da República, 2014. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l13005.htm. Acesso em: 20 dez. 2025.

BRASIL. Lei nº 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. Brasília, DF: Presidência da República, 1996. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9394.htm. Acesso em: 15 dez. 2025.

BRASIL. Ministério da Educação. Portaria Normativa nº 12, de 5 de setembro de 2008. Institui o Índice Geral de Cursos da Instituição de Educação Superior (IGC). Brasília, DF: MEC, 2008b. Disponível em: https://download.inep.gov.br/download/condicoes_ensino/2008/PORTARIA_NORMATIVA_12.pdf. Acesso em: 9 jun. 2026.

BRASIL. Ministério da Educação. Portaria Normativa nº 4, de 5 de agosto de 2008. Regulamenta a aplicação do conceito preliminar de cursos superiores. Brasília, DF: MEC, 2008a. Disponível em: https://download.inep.gov.br/download/superior/condicoesdeensino/Portaria_N_4_de_5_de_agosto_2008.pdf. Acesso em: 22 abr. 2026.

BRASIL. Ministério da Educação. Portaria Normativa nº 40, de 12 de dezembro de 2007 (republicada em 29 de dezembro de 2010). Institui o e-MEC e consolida disposições sobre indicadores de qualidade. Brasília, DF: MEC, 2007/2010. Disponível em: https://download.inep.gov.br/educacao_superior/censo_superior/legislacao/2007/portaria_40_12122007.pdf. Acesso em: 3 mar. 2026.

BRASIL. Tribunal de Contas da União. **Orientações para o cálculo dos indicadores de gestão: Decisão TCU nº 408/2002, Plenário.** Brasília, DF: TCU/Sesu/SFC, 2004a.

BRESSER-PEREIRA, L. C. Da administração pública burocrática à gerencial. **Revista do Serviço Público**, v. 47, n. 1, p. 7-40, 1996. DOI: <https://doi.org/10.21874/rsp.v47i1.702>

BRUIJN, H. de. **Managing performance in the public sector.** London: Routledge, 2002.

CAVE, M. *et al.* **The use of performance indicators in higher education: the challenge of the quality movement.** 3. ed. London:

Jessica Kingsley, 1997.

CENEVIVA, R.; FARAH, M. F. S. Avaliação, informação e responsabilização no setor público. **Revista de Administração Pública**, v. 46, n. 4, p. 993-1016, 2012. DOI: <https://doi.org/10.1590/S0034-76122012000400005>

DILL, D. D.; SOO, M. Academic quality, league tables, and public policy: a cross-national analysis of university ranking systems. **Higher Education**, v. 49, n. 4, p. 495-533, 2005. DOI: <https://doi.org/10.1007/s10734-004-1746-8>

DIMAGGIO, P. J.; POWELL, W. W. The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. **American Sociological Review**, v. 48, n. 2, p. 147-160, 1983. DOI: <https://doi.org/10.2307/2095101>

ESPELAND, W. N.; SAUDER, M. Rankings and reactivity: how public measures recreate social worlds. **American Journal of Sociology**, v. 113, n. 1, p. 1-40, 2007. DOI: <https://doi.org/10.1086/517897>

GEORGE, A. L.; BENNETT, A. **Case studies and theory development in the social sciences**. Cambridge, MA: MIT Press, 2005.

GERRISH, E. The impact of performance management on performance in public organizations: a meta-analysis. **Public Administration Review**, v. 76, n. 1, p. 48-66, 2016. DOI: <https://doi.org/10.1111/puar.12433>

HATRY, H. P. **Performance measurement: getting results**. 2. ed. Washington, DC: Urban Institute Press, 2006.

HAZELKORN, E. **Rankings and the reshaping of higher education: the battle for world-class excellence.** 2. ed. London: Palgrave Macmillan, 2015.

HOOD, C. A public management for all seasons? **Public Administration**, v. 69, n. 1, p. 3-19, 1991. DOI: <https://doi.org/10.1111/j.1467-9299.1991.tb00779.x>

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Metodologia de cálculo dos indicadores de fluxo da educação superior.** Brasília, DF: Inep, 2017.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Nota Técnica nº 58/2020/CGCQES/DAES: cálculo do Conceito Preliminar de Curso referente a 2019.** Brasília, DF: Inep, 2020.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Sinopse estatística da educação superior.** Brasília, DF: Inep, 2023.

JANNUZZI, P. M. Indicadores para diagnóstico, monitoramento e avaliação de programas sociais no Brasil. **Revista do Serviço Público**, v. 56, n. 2, p. 137-160, 2005. DOI: <https://doi.org/10.21874/rsp.v56i2.222>

JONGBLOED, B.; VOSSENSTEYN, H. Keeping up performances: an international survey of performance-based funding in higher education. **Journal of Higher Education Policy and Management**, v. 23, n. 2, p. 127-145, 2001. DOI: <https://doi.org/10.1080/13600800120088625>

KROLL, A. Drivers of performance information use: systematic literature review and directions for future research. **Public Performance & Management Review**, v. 38, n. 3, p. 459-486, 2015. DOI: <https://doi.org/10.1080/15309576.2015.1006469>

MAHONEY, J. Path dependence in historical sociology. **Theory and Society**, v. 29, n. 4, p. 507-548, 2000. DOI: <https://doi.org/10.1023/A:1007113830879>

MAHONEY, J.; THELEN, K. (Eds.). **Explaining institutional change: ambiguity, agency, and power**. Cambridge: Cambridge University Press, 2010.

MCLAUGHLIN, J. A.; JORDAN, G. B. Logic models: a tool for telling your program's performance story. **Evaluation and Program Planning**, v. 22, n. 1, p. 65-72, 1999. DOI: [https://doi.org/10.1016/S0149-7189\(98\)00042-1](https://doi.org/10.1016/S0149-7189(98)00042-1)

MOYNIHAN, D. P. **The dynamics of performance management: constructing information and reform**. Washington, DC: Georgetown University Press, 2008.

MOYNIHAN, D. P.; PANDEY, S. K. The big question for performance management: why do managers use performance information? **Journal of Public Administration Research and Theory**, v. 20, n. 4, p. 849-866, 2010. DOI: <https://doi.org/10.1093/jopart/muq004>

OSBORNE, D.; GAEBLER, T. **Reinventing government: how the entrepreneurial spirit is transforming the public sector**. Reading, MA: Addison-Wesley, 1992.

PIERSON, P. Increasing returns, path dependence, and the study of politics. **American Political Science Review**, v. 94, n. 2, p. 251-267, 2000. DOI: <https://doi.org/10.2307/2586011>

POLIDORI, M. M.; MARINHO-ARAUJO, C. M.; BARREYRO, G. B. SINAES: perspectivas e desafios na avaliação da educação superior brasileira. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 14, n. 53, p. 425-436, 2006. DOI: <https://doi.org/10.1590/S0104-40362006000400002>

POLLITT, C. The logics of performance management. **Evaluation**, v. 19, n. 4, p. 346-363, 2013. DOI: <https://doi.org/10.1177/1356389013505040>

POLLITT, C.; BOUCKAERT, G. **Public management reform: a comparative analysis**. 3. ed. Oxford: Oxford University Press, 2011.

POWER, M. **The audit society: rituals of verification**. Oxford: Oxford University Press, 1997.

PROPPER, C.; WILSON, D. The use and usefulness of performance measures in the public sector. **Oxford Review of Economic Policy**, v. 19, n. 2, p. 250-267, 2003. DOI: <https://doi.org/10.1093/oxrep/19.2.250>

RADIN, B. A. **Challenging the performance movement: accountability, complexity, and democratic values**. Washington, DC: Georgetown University Press, 2006.

SCOTT, J. C. **Seeing like a state: how certain schemes to improve the human condition have failed**. New Haven, CT: Yale University Press, 1998. DOI: <https://doi.org/10.2307/j.ctvxkn7ds>

SECCHI, L. Modelos organizacionais e reformas da administração pública. **Revista de Administração Pública**, v. 43, n. 2, p. 347-369, 2009. DOI: <https://doi.org/10.1590/S0034-76122009000200004>

SMITH, P. On the unintended consequences of publishing performance data in the public sector. **International Journal of Public Administration**, v. 18, n. 2-3, p. 277-310, 1995. DOI: <https://doi.org/10.1080/01900699508525011>

STREECK, W.; THELEN, K. (Eds.). **Beyond continuity: institutional change in advanced political economies**. Oxford: Oxford University Press, 2005.

TALBOT, C. **Theories of performance: organizational and service improvement in the public domain**. Oxford: Oxford University Press, 2010.

TREVISAN, A. P.; VAN BELLEN, H. M. Avaliação de políticas públicas: uma revisão teórica de um campo em construção. **Revista de Administração Pública**, v. 42, n. 3, p. 529-550, 2008. DOI: <https://doi.org/10.1590/S0034-76122008000300005>

TRIBUNAL DE CONTAS DA UNIÃO. **Acórdão nº 1.043/2006, Plenário**. Brasília, DF: TCU, 2006a. Texto reproduzido em: TRIBUNAL DE CONTAS DA UNIÃO. Orientações para o cálculo dos indicadores de gestão: Decisão nº 408/2002 e Acórdãos nº 1.043/2006 e nº 2.167/2006, Plenário. Brasília, DF: TCU/Sesu/SFC, [s.d.]. Disponível em: https://www.ufc.br/images/_files/a_universidade/avaliacao_institucional/orientacoes_indicadores_gestao_tcu.pdf. Acesso em: 14 fev. 2026.

TRIBUNAL DE CONTAS DA UNIÃO. **Acórdão nº 2.167/2006, Plenário**. Brasília, DF: TCU, 2006b. Texto reproduzido em: TRIBUNAL DE

CONTAS DA UNIÃO. Orientações para o cálculo dos indicadores de gestão: Decisão nº 408/2002 e Acórdãos nº 1.043/2006 e nº 2.167/2006, Plenário. Brasília, DF: TCU/Sesu/SFC, [s.d.]. Disponível em: https://www.ufc.br/images/_files/a_universidade/avaliacao_institucional/orientacoes_indicadores_gestao_tcu.pdf. Acesso em: 14 fev. 2026.

TRIBUNAL DE CONTAS DA UNIÃO. **Acórdão nº 461/2022, Plenário.** Processo TC 026.147/2020-3. Rel. Min. Walton Alencar Rodrigues. Brasília, DF: TCU, 2022. Disponível em: <https://pesquisa.apps.tcu.gov.br/redireciona/acordao-completo/ACORDAO-COMPLETO-2470776>. Acesso em: 20 mai. 2026.

TRIBUNAL DE CONTAS DA UNIÃO. **Decisão nº 408/2002, Plenário.** Brasília, DF: TCU, 2002. Texto reproduzido em: TRIBUNAL DE CONTAS DA UNIÃO. Orientações para o cálculo dos indicadores de gestão: Decisão nº 408/2002 e Acórdãos nº 1.043/2006 e nº 2.167/2006, Plenário. Brasília, DF: TCU/Sesu/SFC, [s.d.]. Disponível em: https://www.ufc.br/images/_files/a_universidade/avaliacao_institucional/orientacoes_indicadores_gestao_tcu.pdf. Acesso em: 10 jan. 2026.

TRIBUNAL DE CONTAS DA UNIÃO. **Indicadores das Universidades Federais [página temática].** Brasília, DF: TCU, 2026. Disponível em: <https://portal.tcu.gov.br/publicacoes-institucionais/Temas%20Fiscalizados/indicadores-de-gestao-e-desempenho-das-universidades-federais>. Acesso em: 20 ago. 2026.

VAN DOOREN, W.; BOUCKAERT, G.; HALLIGAN, J. **Performance management in the public sector.** 2. ed. London: Routledge, 2015. DOI: <https://doi.org/10.4324/9781315817590>

VERHINE, R. E.; DANTAS, L. M. V.; SOARES, J. F. Do Provão ao ENADE: uma análise comparativa dos exames nacionais utilizados no ensino

superior brasileiro. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 14, n. 52, p. 291-310, 2006. DOI: <https://doi.org/10.1590/S0104-40362006000300002>

DECLARAÇÕES

Financiamento: Esta pesquisa não recebeu financiamento de agências dos setores público, privado ou sem fins lucrativos.

Conflito de interesses: O autor declara não haver conflito de interesses de natureza financeira, institucional ou pessoal relacionado a este manuscrito.

Disponibilidade de dados: Todos os dados utilizados neste estudo são públicos e estão disponíveis nas fontes oficiais citadas ao longo do texto e na lista de referências (Censo da Educação Superior, Decisão TCU nº 408/2002 e atos posteriores, Sinaes e indicadores de fluxo do Inep).

Comitê de ética: Não aplicável. A pesquisa tem natureza documental, baseada exclusivamente em fontes públicas oficiais, sem coleta de dados de seres humanos ou animais.

¹ Mestre em Economia e Gestão Empresarial pela Universidade Cândido Mendes (UCAM). Graduado em Ciências Contábeis pela Universidade Federal Fluminense (UFF). Especialista em Gestão Pública, Gestão Estratégica e Qualidade. Contador da UFF. Endereço postal: Rua Miguel de Frias, nº 9, Icaraí, Niterói, RJ, CEP 24220-900. E-mail: [acesse o artigo original para visualizar o e-mail](#).